

## **Artificial intelligence for public health: Priorities for successful use by public health organizations**

Stacey Fisher, PhD<sup>1,2,3,4</sup> and Laura C Rosella, PhD<sup>1,3,4,5</sup>

1. Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
2. Public Health Ontario, Toronto, ON, Canada
3. Vector Institute for Artificial Intelligence, Toronto, ON, Canada
4. ICES, Toronto, ON, Canada
5. Institute for Better Health, Trillium Health Partners, Mississauga, ON, Canada

Corresponding Author:

Dr. Stacey Fisher

Health Sciences Building 6<sup>th</sup> Floor

155 College Street Toronto, ON M5T 3M7

Phone: (416) 978-2201

Email: [stacey.fisher@utoronto.ca](mailto:stacey.fisher@utoronto.ca)

## **SUMMARY**

Artificial intelligence (AI) has the potential to improve public health surveillance, health promotion, and population health management through improved targeting of interventions and policy to populations that are most in need, known as precision public health. To successfully realize this potential and use AI for public health functions it is important for public health organizations to thoughtfully develop strategies for AI implementation. Five key priorities for successful use of AI technologies by public health organizations are identified in this commentary: 1) Contemporary data governance; 2) Investment in modernized data and analytic infrastructure and procedures; 3) Addressing the skills gap; 4) Development of strategic collaborative partnerships, and; 5) Use of AI best practices including explicit consideration of equity.

Keywords: artificial intelligence; public health; health policy

## **INTRODUCTION**

Rapid growth in the creation and accessibility of health-related data, together with advances in data storage, computational power, and analytic capacity, has brought about opportunities for artificial intelligence (AI) and its subfields to improve public health. Just as AI has been proposed as the cornerstone of precision medicine, it also has the potential to facilitate improved targeting of population health interventions and policy to populations that are most in need. Known as ‘precision public health’, AI can be used to inform how the right intervention can be targeted to the right population at the right time.[1] Public health organizations responsible for the delivery and oversight of public health functions are making significant efforts to plan how to incorporate AI into their core functions. Notably, the considerations for public health are distinctly different from those for clinical or health care systems. In this commentary, we identify opportunities and challenges associated with the use of AI for public health functions and discuss key priorities for successful implementation by public health organizations.

### **Using AI to improve public health**

The creation and availability of novel streams of data relevant for health has increased dramatically over the last 20 years, providing detailed insight into social, behavioral, and environmental determinants of health never before possible. For example, data from social media, web search engines and forums, news media, mobile devices and apps provide information about the social determinants of health that is more nuanced than that from traditional sources.[2-4] Wearable devices allow for the collection of detailed information about personal movements and physiological measurements.[5] Environmental sensors collect spatial data about air pollution, water quality, environmental noise, weather conditions, and green

space.[6] Much of this data is being produced continuously and can be analyzed in real-time using powerful and increasingly available application programming interfaces. Furthermore, linkage of these novel data sources to traditional public health data, including that from administrative records, electronic health records, census and health survey data greatly expands potential use. This increase in the type, size and complexity of health-related data has presented new opportunities for AI methods to improve public health through modernized disease and risk factor surveillance and improved targeting of health interventions.[1-3, 7-9]

Surveillance is one of the core activities of public health[10] and is traditionally performed using population health surveys, clinical data, and public health reporting systems. Access to new data sources and AI methods provides opportunities to identify emerging health threats and develop a more detailed understanding of population disease and risk factor distributions, often with improved geographic resolution more quickly. AI has been used to detect and track infectious disease outbreaks using commercial flight itineraries, climate data and animal and insect population data, including early detection of COVID-19 in Wuhan, China[11] and prediction of international Zika transmission.[12] The Chicago Department of Public Health has used an ML algorithm to identify children at high risk of lead poisoning and prioritize homes for lead inspections using historical blood lead level tests and home lead inspection data, child characteristics, property value assessments and census data.[13]

Traditional public health surveillance systems are often hindered by the lag time between event occurrence and reporting. AI-powered approaches can provide information that is more up to date as data are often collected, processed, and analyzed in real-time. Furthermore, they can be

used to summarize information from unstructured sources. For example, natural language processing analysis of free-text information in death certificates has been used to identify potential drug overdose deaths months prior to traditional coding and data release.[14] Public health surveillance dashboards powered by web-accessible news and social media data have been developed to display health events both geographically and temporally.[15, 16] SENTINEL, for example, is a syndromic surveillance tool that processes over 1.8 million tweets a day to predict disease occurrence and identify potential outbreaks in real-time.[16] News articles are collected to provide context, and an intuitive user interface displays event predictions geographically and over time compared to weekly counts from the United States Centers for Disease Control. Other examples of AI use for public health surveillance include forecasting of disease incidence for population health planning,[17] prediction of E.coli concentration in beach water,[18] and investigation of foodborne disease outbreaks using grocery store loyalty card[19] and web search engine data.[20]

Health promotion activities can also be better targeted and made more efficient using AI technologies.[21] For example, individuals with mixed opinions about hookah tobacco smoking have been identified using sentiment analysis of Twitter data, facilitating public health campaigns targeting those who may be most receptive.[22] Similar methods and data have also been used to identify individuals at risk of suicide.[23, 24] Interactive online tools or apps powered by AI technologies can also provide highly accessible individualized risk assessment and risk reduction recommendations that are more engaging and motivating than traditional approaches, for example, for chronic disease prevention[25] and management[26] or to encourage behavioral change.[27] The World Health Organization recently released Florence, a

‘digital health care assistant’, as part of their AI for Quitting Tobacco initiative.[28] Using computer-generated imagery, animation and AI, Florence is designed to help people quit tobacco and combat misinformation about COVID-19 and tobacco, through online, face-to-face conversation.

AI also has the potential to improve population health management by better identifying population subgroups most in need,[29] estimating the potential effects of policy change,[30] and for causal inference.[2] There is also potential for natural language processing to be used to de-identify personal health information,[31] automatically translate scientific literature,[32] and assist in the production of systematic reviews of public health interventions.[33]

### **Challenges associated with using AI for public health**

There are many challenges, risks, and limitations to using AI for public health, including ethical and privacy concerns, poor model interpretability and structural challenges including data sharing, insufficient infrastructure, and lack of public health workforce training in AI.

Although AI has the potential to decrease health inequities through their identification and subsequent targeting of resources, it also has the potential to create, sustain or exacerbate these inequities. This can occur throughout the development pipeline through factors that affect research question selection, representativeness of the data, choice of outcome definition, optimization decisions made during algorithm development and post-deployment decisions.[34] For example, an algorithm used in hospitals to allocate health care services to patients in the United States was found to be biased against Black patients.[35] Among Black and White

patients who were equally sick, Black patients were assigned lower risk scores and were therefore less likely to receive additional services. This bias arose because the algorithm was designed to predict health care costs rather than illness; as Black patients tend to have poorer access to care, they also tend to cost the health care system less. More examples of how inequities can occur throughout the AI development pipeline can be found in Chen et al., 2021.[34]

The use of AI and big data in health also raises privacy concerns. Linkage of multiple anonymized data sources is often performed to increase the richness of data prior to analysis with AI techniques; however, this also increases the risk of re-identification of individuals or stigmatization of small groups.[9] Use of data from social media sites, blogs and forums have been associated with risks to individual privacy and autonomy and with potential for stigmatization.[36] Research in differential privacy aims to mitigate these challenges but remains largely unable to maintain statistical properties that are important for successful AI use.[37] For example, a proposed application of differential privacy to the 2020 United States census was found to radically change population counts for racial/ethnic minorities and lead to biased mortality rate estimates.[38]

Another commonly discussed limitation of many AI methods is poor explainability, or interpretability, of a model and its outputs. Many AI algorithms are described as a ‘black boxes’ as models can contain many variables modelled in nonlinear and complex ways, making it difficult or impossible for a human to understand how the output was arrived at. This lack of interpretability can cause skepticism and be detrimental to user trust, especially in the health

context. Some machine learning models are more interpretable than others, including penalized regression methods and single decision trees, however, interpretability often comes at the cost of model performance.[39]

Most traditionally trained public health professionals have not had the training to develop, evaluate or implement AI-based technologies. As such, AI in public health is currently limited and does not take full advantage of the capabilities of the methods or the richness of available data. For example, a recent scoping review of machine learning for population health prediction found that few studies utilized big data, with a median feature size of only 17, and few models used non-traditional sources of health data.[40] It is therefore unsurprising that studies comparing the use of machine learning methods to AI methods for clinical prediction based on relatively limited data sources have found little advantage to using machine learning methods over traditional statistical methods.[41] There are many good resources that can serve as a starting point for public health professionals interested in enhancing AI, machine learning and big data skills.[2, 9, 42-44] Workforce training is an important step in facilitating AI adoption.[45, 46]

Other limitations include difficulties in accessing personal health data, sharing data across jurisdictions, poor data integration, outdated analytic infrastructure, and lack of software development to facilitate deployment of AI applications into the health system.[47]

**How should health organizations incorporate AI into public health activities?**



Many health organizations have begun to strategize how to best incorporate AI into their core functions and have developed AI or data-specific strategies, reports, and guidance documents (**Table 1**; see Appendix for search strategy). Review of these documents reveals many common priorities and approaches. Informed by this review, we have identified five key priorities needed for successful use of AI technologies by public health organizations (**Table 2**):

- 1) Contemporary data governance
- 2) Investment in modernized data and analytic infrastructure and procedures
- 3) Addressing the skills gap
- 4) Development of strategic collaborative partnerships
- 5) Use of AI best practices including explicit consideration of equity

### ***Contemporary data and analytic governance***

Every public health organization exists within a larger governance context. Comprehensive understanding of relevant legislation, policies and procedures that govern use of AI for health is therefore integrally important to the safe and successful use of AI for public health activities.

This governance exists at different levels, from the international and federal level to organization-specific governance frameworks designed to guide the strategic and efficient management of data and AI technologies. All the organizational documents we reviewed discussed organizational governance and the associated challenges. However, we argue that public health organizations need to focus on understanding and operationalizing higher-level governance rather than reinterpreting into organization-level governance frameworks.

Importantly, this should include the intimate involvement of subject-matter experts in AI, data management and information technology to help ensure that higher-level governance is being

**Table 1.** Data strategies and artificial intelligence guidance reports for health and public health

Country	Organization	Document Title	Link
<b>Organizational Data Strategies</b>			
Canada	Statistics Canada	Delivering insight through data for a better Canada (2019)[48]	<a href="https://www.statcan.gc.ca/eng/about/datastrategy">https://www.statcan.gc.ca/eng/about/datastrategy</a>
Canada	Health Canada	Data strategy (2019)[49]	Not publicly available
Canada	PHAC	Data strategy (2019)[50]	Not publicly available
Canada	ICES	Envisioning a data science strategy for ICES (2017)[51]	<a href="https://www.ices.on.ca/Publications/Atlases-and-Reports/2017/Data-science-strategy">https://www.ices.on.ca/Publications/Atlases-and-Reports/2017/Data-science-strategy</a>
Canada	Public Health Ontario	Informatics strategy (in development)[52]	Not publicly available
United States	National Institutes of Health	Strategic plan for data science (2018)[53]	<a href="https://datascience.nih.gov/nih-strategic-plan-data-science">https://datascience.nih.gov/nih-strategic-plan-data-science</a>
United States	Department of Health and Human Services	HHS Data Strategy: Enhancing the HHS evidence-based portfolio (2018)[54]	<a href="https://aspe.hhs.gov/system/files/pdf/261591/2018HHSDataStrategy.pdf">https://aspe.hhs.gov/system/files/pdf/261591/2018HHSDataStrategy.pdf</a>
United Kingdom	NHS Digital	Data and information strategy (2016)[55]	<a href="https://digital.nhs.uk/data-and-information/nhs-digital-data-and-information-strategy#">https://digital.nhs.uk/data-and-information/nhs-digital-data-and-information-strategy#</a>
<b>AI Reports and Guidance Documents</b>			
International	Pan American Health Organization	Public health in the information society (2017)[56]	<a href="https://www.paho.org/salud-en-las-americanas-2017/?tag=artificial-intelligence">https://www.paho.org/salud-en-las-americanas-2017/?tag=artificial-intelligence</a>
Canada	CIFAR	Building a learning health system for Canadians: Report of the Artificial Intelligence for Health Task Force (2020)[57]	<a href="https://cifar.ca/ai/national-program-of-activities/ai4health-task-force/">https://cifar.ca/ai/national-program-of-activities/ai4health-task-force/</a>
Canada	CIHR and CIFAR	AI for public health equity workshop report (2019)[58]	<a href="https://cihrirsc.gc.ca/e/documents/ai_public_health_equity-en.pdf">https://cihrirsc.gc.ca/e/documents/ai_public_health_equity-en.pdf</a>
Canada	CIHR and CIFAR	Application of artificial intelligence approaches to tackle public health challenges (2018)[59]	<a href="https://cihr-irsc.gc.ca/e/documents/artificial_intelligence_approaches-en.pdf">https://cihr-irsc.gc.ca/e/documents/artificial_intelligence_approaches-en.pdf</a>
United Kingdom	NHSX	Artificial intelligence: How to get it right (2019)[60]	<a href="https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf">https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf</a>
United Kingdom	Academy of Medical Royal Colleges	Artificial intelligence in health care (2019)[61]	<a href="https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf">https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf</a>
United States	Center for Open Data Enterprise for the Department of Health and Human Services (HHS)	Sharing and utilizing health data for AI applications: Roundtable report (2019)[62]	<a href="https://www.hhs.gov/sites/default/files/sharing-and-utilizing-health-data-for-ai-applications.pdf">https://www.hhs.gov/sites/default/files/sharing-and-utilizing-health-data-for-ai-applications.pdf</a>
United States	National Academy of Medicine	Artificial intelligence in health care: The hope, the hype, the promise, the peril (2019)[63]	<a href="https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf">https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf</a>
United States	JASON for the United States Department of Health and Human Services	Artificial intelligence for health and health care (2017)[64]	<a href="https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf">https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf</a>

Abbreviations: AI, artificial intelligence; CIFAR, Canadian Institute for Advanced Research; CIHR, Canadian Institutes of Health Research; HHS, Health and Human Services; IBC, International Bioethics Committee; NHS, National Health Service; NIH, National Institutes of Health; PHO, Public Health Ontario; UNESCO, United Nations Educational, Scientific and Cultural Organization

**Table 2.** General recommendations to support five strategic priorities for successful use of artificial intelligence by public health organizations

Strategic Priority	General Recommendations
<b>Governance</b>	<ul style="list-style-type: none"> <li>- Clarify data leadership roles and responsibilities</li> <li>- Review current organizational governance</li> <li>- Understand and operationalize higher-level governance               <ul style="list-style-type: none"> <li>o Involve subject-matter experts in AI, data management and information technology</li> <li>o Mechanism for community and public engagement</li> </ul> </li> <li>- Establish transparent oversight and accountability</li> </ul>
<b>Infrastructure</b>	<ul style="list-style-type: none"> <li>- Assess infrastructural and analytic needs</li> <li>- Increase data access</li> <li>- Improve data interoperability</li> <li>- Increase availability of advanced analytic infrastructure and tools               <ul style="list-style-type: none"> <li>o Consider investment in distributed data platforms and cloud computing</li> </ul> </li> </ul>
<b>Upskilled Workforce</b>	<ul style="list-style-type: none"> <li>- Identify and forecast desired skills and competencies, and review existing skills and capacity</li> <li>- Upskill existing staff               <ul style="list-style-type: none"> <li>o Increase data literacy across the organization, with a focus on bias and equity considerations</li> </ul> </li> <li>- Recruit new staff with desired skills</li> <li>- Engage with trainees; consider development of trainee fellowship programs</li> <li>- Foster multidisciplinary collaboration and diversity</li> </ul>
<b>Partnerships</b>	<ul style="list-style-type: none"> <li>- Identify areas where partnerships may be helpful (e.g., gain expertise, gain or share access to data or infrastructure, engage a wider variety of perspectives)</li> <li>- Consider partnerships with               <ul style="list-style-type: none"> <li>o Local, provincial/state, federal government</li> <li>o Educational institutions</li> <li>o Private sector</li> </ul> </li> </ul>
<b>Best Practices</b>	<ul style="list-style-type: none"> <li>- Consider use of an existing ethical AI framework</li> <li>- Default to transparent data and analytic processes and following open science principles whenever possible</li> <li>- Engage with the public</li> <li>- Ensure access to practical guidelines for AI development, evaluation, and implementation</li> </ul>

interpreted appropriately, operationalized realistically, benefits and risks are both fully understood, and that unnecessary restrictions are not being implemented. It is important to also recognize that the higher-level governance context can change, and that organizational governance must be able to easily adapt. The European Union General Data Protection Regulation represented a massive shift in data protection and privacy and has prompted review of privacy regulation around the world.[65] Canada's Privacy Act, for example, is currently under review.[66]

Organization-level governance should focus on the development and maintenance of effective and efficient data and information technology (IT) systems within the constraints of higher-level regulation. This should include an emphasis on data procurement, linkage and access, privacy, data and IT interoperability, investment in and maintenance of IT infrastructure, prioritization of AI projects, and workforce management of AI, data, and IT personnel. Common governance priorities identified in the documents reviewed include transparent and clear definition of roles and responsibilities and strict oversight and accountability.[48-50, 54, 56] Several organizations have established new roles to lead data governance activities, including a Chief Data Officer at the Public Health Agency of Canada (PHAC)[50] and Health Canada[49], and a Chief Data Strategist at the United States National Institutes of Health.[53] Individuals in these roles are tasked with leading data strategy implementation in collaboration with relevant organizational data councils. Other organizations have prioritized increased communication and coordination between relevant individuals and councils responsible for data governance activities.[54]

For ethical AI use, the public should be engaged and informed about how their data is used, how AI applications may influence their lives, and be given space to voice their preferences and concerns.[57, 63, 67, 68] Community governance, which involves participation and engagement of the public in decision-making about one's community, has become recognized as particularly important when considering First Nations' and Indigenous data and information. The First Nations Principles of OCAP (Ownership, Control, Access, and Possession) were developed to protect Canadian First Nations' data and information and ensure that it is used and shared in a way that brings benefit to the community while minimizing harm ([www.fnigc.ca](http://www.fnigc.ca)). Similarly, the CARE (Collective Benefit, Authority to Control, Responsibility and Ethics) Principles for Indigenous Data Governance are global principles for governance of Indigenous data ([www.gida-global.org/care](http://www.gida-global.org/care)).

### ***Investment in modernized data and analytic infrastructure and procedures***

Modernization of organizational data infrastructure and procedures is widely recognized by health organizations as vital to moving forward with AI application and strategic use of data. A common priority of all organizational strategies we reviewed was to improve data access[48-55] by reducing administrative barriers, reviewing, and revising data use agreements, exploring new data de-identification techniques and establishing remote access to data and analytic tools. Investment in distributed data platforms and cloud computing infrastructure is widely discussed as a means of facilitating rapid and seamless data access in addition to improving data storage and increasing computational power for advanced analytics.[48, 50, 51, 53, 54, 57] These platforms may also reduce infrastructure and maintenance costs in the long-term, compared to local data centers.[53] Health Canada additionally provides access to data through application

programming interfaces, [49] which Statistics Canada are also looking to use to provide data access to Government of Canada departments.[48]

Many organizations are also seeking to improve data interoperability. The NHS is aiming to modernize data infrastructure and increase interoperability through development of a Data Services Platform that will serve as a single place for data collection, processing and management.[55] Similarly, the United States National Institutes of Health (NIH) has goals to connect their data systems and reduce data ‘silos’.[53] Interoperability is also a primary goal of the Statistics Canada Data Strategy, which they are seeking to improve through the development and use of open data standards.[48] Similarly, Health Canada is aiming to improve data standardization, consolidation and integration through use of open standards and sharing of expertise.[49] Easily accessible data documentation, essential for data interoperability, has also been prioritized in several of the organizational data strategies we reviewed. Examples of this include the Health Canada Information Reference Model,[49] the United States NIH Data Discovery Index[53] and a data holding inventory by PHAC.[50] Some organizations are also seeking to improve data interoperability through use of common data models, schema for data harmonization and standardization.[49, 51, 53] Use of existing commercial tools, technologies and services as opposed to internal development of project or organization-specific data infrastructure is also recognized as a means of improving system interoperability and data integration both within and outside of an organization. [48, 49, 59] Increased data linkage is also a common organizational priority.[51-55]

In addition to modern data infrastructure and procedures, successful use of AI also requires advanced analytic infrastructure and tools. Many organizational strategies outline plans to increase organizational capacity for advanced analytics by assessing organizational needs,[49] increasing computational power,[48, 50] facilitating access to new analytic tools,[48-53, 55] and through pilot projects using AI methods.[50, 51] It is important to establish what analytic tools are needed to enable AI use, as most traditional public health tools are incapable and/or are not familiar to those with AI or machine learning expertise. For example, Python (Python Software Foundation) and R (R Foundation for Statistical Computing) are programming languages commonly used to develop machine learning models and Git is a popular, free, and open-source version control system that tracks coding changes. TensorFlow, which is also free and open-source, is an end-to-end software library for machine learning that is especially effective at efficiently deploying machine learning algorithms.[69] As public health professionals do not traditionally use these tools currently, it is important for those with expertise in computer science, AI, and machine learning to determine the appropriate infrastructure, software and tools needed to perform advanced analytics. Several organizations also recognized the importance of flexibility in accessing new analytic tools to enable ‘nimble and agile data analytics’.[48-50]

### ***Addressing the skills gap***

Successful use of big data, advanced analytic methods and AI requires a workforce with strong data literacy and capacity in data management, statistics, computer science, software engineering, data privacy, bias, and ethics, among other skills. All organizational data strategies we reviewed recognized the importance of building a workforce that is educated in these skills and outlined plans to achieve it through training staff and leveraging existing skills, targeted

recruitment, and engagement with trainees and educational institutions. Most of the strategies also discussed the intention to increase organizational skills and capacity in AI specifically.[48-54]

Upskilling existing staff will generally be an important priority of all public health organizations interested in increasing use of AI. It should first involve identifying and forecasting desired data and analytic competencies and a review of existing organizational skills and capacity.[50, 52, 55] Data literacy, defined as the ability to collect, manage, evaluate, and critically apply data,[70] is widely recognized as a vital competency to be emphasized across health organizations interested in AI.[48-50, 52, 54] Statistics Canada has developed data literacy training products including the Framework for Responsible Machine Learning Processes at Statistics Canada[71] and introductory training videos on machine learning, data stewardship and data quality, among others.[72] The Government of Canada developed a Digital Academy in 2018 to “help federal public servants gain the knowledge, skills and mindsets they need in the digital age”, and includes training on data literacy and competencies, cloud computing, cyber security, AI and machine learning, among other topics.[73] The Digital Academy is being used by PHAC and Health Canada to train existing and new employees.[49, 50] PHAC outlined many additional training strategies, including use of third-party web-based tools, self-directed learning, trainings customized to specific audiences, development of a Data 101 onboarding package and specific training in innovation.[50] The United States Department of Health and Human Services (HHS) is looking to increase data science and statistical training opportunities and increase multidisciplinary collaboration across the organization, recognizing that informed data science decisions require a wide range of skills and expertise.[54] The NHS is seeking to leverage



existing skills through the creation of teams specializing in particular data skills and through external and internal staff rotation, in addition to the development of training programs.[55] ICES is looking to develop a data science staff education strategy, which will include data science workshops and increased exposure of analysts and methodologists to the R statistical programming language.[51] Statistics Canada is seeking to develop a culture of ‘continuous learning’.[48] Continuous learning can be facilitated in part by increased access to scientific publications, a priority of PHAC.[50]

Targeted recruitment of new employees is another means of developing a workforce educated in data science and AI and is an important component of the workforce development plan for many health organizations.[48-50, 52-55] As individuals with many of the desired skills have not traditionally worked in health, it is important to consider how to best attract and retain this talent. This begins with increasing data literacy across the organization and provision of appropriate infrastructure and tools, and is further facilitated by an organizational culture that is receptive to change and taking risks. A goal of Health Canada Data Strategy, for example, is to provide employees with “an agile collaborative space for learning and innovative uses of data” and is seeking to create an strong data culture and environment that values experimentation and learning from failure.[49] PHAC and PHO have similar innovation and risk-taking goals.[50, 52] The United States HHS strategy outlines four approaches to hiring data scientists, including participating in job fairs and industry events, creation of intern and fellowship programs, hiring of individuals with non-traditional backgrounds into senior positions and making use of existing specialized hiring programs.[54] The United States NIH is looking to develop a Data Fellows

program, in which individuals with desired skills are recruited from the private sector and academia for short-term national service sabbaticals.[53]

It has been suggested that trainees are “the glue that tie researchers together”, fostering interdisciplinary research and learning.[58] Engagement with trainees additionally increases awareness of organizational data science career possibilities and provides the organization with access to new and developing data science talent. Most of the organizational strategies we reviewed include engagement with trainees and educational institutions as part of their workforce development plans.[48, 50, 51, 53, 54] In Canada, the Health System Impact Fellowship[74], funded by the Canadian Institutes for Health Research, has an equitable AI stream in which PhD and postdoctoral fellows with skills in computer science, AI and data science are embedded within health system organizations to help solve critical health system challenges.[75] Both PHAC and PHO have hosted fellows through this program. Other organizations have plans for similar organization-specific fellowship programs, including Statistics Canada[48] and the United States NIH and HHS.[53, 54]

Lastly, it has been recognized that scientific teams greatly benefit from being diverse and multidisciplinary.[34, 58, 63] An AI report from the United States National Academy of Medicine recommends that AI teams are diverse in “gender, culture, race, age, ability, ethnicity, sexual orientation, socioeconomic status, privilege, etc.” to promote the development of impactful and equitable AI tools.[63] The United States NIH is looking to increase workforce diversity, in part through their Big Data to Knowledge Diversity Initiative.[53] The United States

HHS has goals to promote multidisciplinary data science teams and increase cross-program and interdepartmental collaboration.[54]

### ***Development of strategic collaborative partnerships***

Development of collaborative partnerships is an important component of strategic data use and successful AI implementation. Collaboration can come in many forms and be used to gain expertise, gain, or share access to data and infrastructure, and engage a wider variety of perspectives. The CIFAR AI for Public Health Equity report recommends collaboration of public health professionals and researchers with computer science and AI researchers, in addition to a wide range of other groups (e.g. sociologists, political scientists, engineers, civil society and citizen scientists, people with lived experience, policymakers) to help ensure health equity when using AI technologies.[58] ICES is looking to facilitate development and implementation of data and computational infrastructure through partnerships in addition to continued collaboration with external scientists and research institutions for data science and AI expertise.[51] Many governmental organizations plan to closely collaborate with other local, provincial/state or federal government organizations or departments, sharing infrastructure, data and expertise.[48-50, 53, 54] As mentioned previously, engagement with trainees can be greatly beneficial to workforce development and promotes collaboration with educational institutions. Collaboration with the private sector can additionally be advantageous. The United States NIH, for example, is seeking to leverage private sector infrastructure through strategic collaboration.[53]

### ***Use of AI best practices including explicit consideration of equity***

Health equity has been defined to mean that “all people can reach their full health potential and should not be disadvantaged from attaining it because of their race, ethnicity, gender, age, social class, language and minority status, socio-economic status, or other socially determined circumstance”.[76] Ethical considerations exist at all stages of the AI development and implementation pipeline, from problem selection and data collection to post-deployment.[34] As public health professionals are trained to think about bias, generalizability and equity, they are especially able to recognize and inform mitigation strategies for AI use in public health in collaboration with computer science and AI professionals. As mentioned previously, an organization’s workforce should also be diverse and educated in bias and equity issues.

Best practices have been established to guide the development, implementation, and evaluation of AI-powered technologies to ensure that they are not only useful, but additionally do not create, sustain, or exacerbate health inequities. Many principles and frameworks have been developed to guide the ethical use of AI.[67] The UK National Health Service (NHS)[55, 60] and United Kingdom (UK) National Academy of Medical Royal Colleges[61] recommends following the ‘Guide to Good Practice for Digital and Data-Driven Health Technologies’.[77] This document developed by the UK Government outlines ten principles to guide the development and implementation of data-driven health and care technologies. Many organizations[49, 53, 78] refer to the FAIR data principles: research data should be findable, accessible, interoperable and reusable.[79] A report from the United States National Academy of Medicine refers to several existing frameworks and principles including ‘Artificial Intelligence at Google: Our Principles’[80] and the ‘AI Now Report 2018’.[81] Common principles among many of these frameworks include transparency, non-maleficence, responsibility and

accountability, privacy, freedom and autonomy, beneficence, trust and justice, fairness and equity.[67]

Transparency is intended to foster trust and prevent harm and is one of the most common ethical AI principles.[67] Transparency in AI often refers to efforts to increase explainability and interpretability and generally involves detailed disclosure of how an AI model or technology was developed, how it performs, the data it uses, how it is deployed and used, discussion of limitations, and may involve sharing of source code and data.[67] Transparent AI promotes freedom and autonomy by increasing the public's knowledge of AI and promoting informed consent.[67] The EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network is an international initiative promoting transparent reporting of health research literature by encouraging wider use of robust reporting guidelines ([www.equator-network.org](http://www.equator-network.org)).

Particularly relevant to AI is the 'Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis' (TRIPOD) reporting guideline,[82] with an upcoming extension specifically for models developed using machine learning methods.[83] Health Canada is looking to improve data transparency and access to data as a means of increasing public confidence in decision-making.[49] The NHS has committed to greater transparency in data use and in algorithmic development and deployment.[77]

Closely related to the principle of transparency is open science. Open science is a movement to make scientific research transparent and accessible to all. Open science reduces research waste, facilitates reproducibility, and allows for AI to more easily benefit everyone (related to the ethical principle of beneficence). One of the main recommendations of the CIFAR Public Health

Equity report[58] and the NHSX[60] is that organizations should use and allow for sharing of datasets, data repositories, and resources. Open science efforts include providing the public with access to data and information, use of open data standards and open-source programs, open-source code, use of open data, and open access publication. A commitment to increased data sharing was stated in several organizational strategies[49, 50, 53, 54] Several organizations also prioritized increasing their use of open data standards to improve interoperability.[48, 50, 53] Statistics Canada is committed to increased transparency of data use and processes, including through publishing of code on the Open Data Portal.[48] The Pan American Health Organization also lists open science and open data as guiding principles.[56] Principles of transparency and open science, however, need to be carefully balanced with privacy and confidentiality through organizational governance. Protections must be in place to ensure data protection and security and prevent discrimination of individuals and small population sub-groups.[84] The Health Canada data strategy states that “getting privacy and ethics right will actually enable increased use and sharing of data, since data stewards will have knowledge of the data limits and have confidence that they can use and share data without harm.”[49]

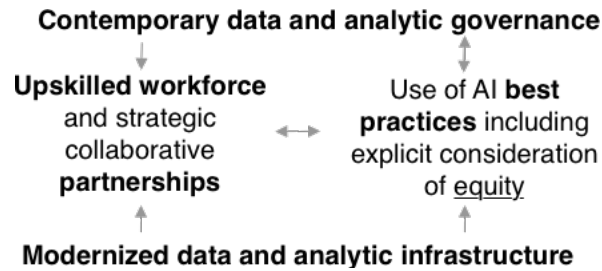
As mentioned previously, public engagement is also important for the ethical use of AI.[57, 63, 67, 68] The CIFAR Artificial Intelligence for Health (AI4H) task force report recommends that “members of the public and patients should be included as active partners in the development, governance and evaluation of AI4H policies and strategies”.[57] For public health specifically, it has been recommended that rural and remote communities and people with lived experiences be engaged in relevant AI research and implementation from project initiation.[58, 59] Another means of engaging with the public is through citizen science, in which members of the public

lead or participate in scientific research, recommended by two reports from the United States.[63, 64] The United States NIH is committed to facilitating citizen science in their Strategic Plan for Data Science, through public access to data, tools and education in addition to exploration of other community engagement models.[53]

In addition to ethical principles and guidelines there exist practical guidelines for developing and reporting prediction models. The detailed explanation and elaboration document for the previously mentioned TRIPOD reporting guideline lists many practical recommendations for developing well performing models, including predictor measurement and description, defining the outcome, handling of missing data and variable preprocessing.[85] We anticipate that the upcoming TRIPOD-ML guideline will be especially useful.[83] Other practical considerations for the use of AI in health include the importance of representative data, cross-validation and data leakage, overfitting, and rigorous model evaluation.[86, 87]

### **Moving towards an AI-enabled public health organization**

Among those that provided a timeline, organizations generally planned to take steps toward all identified priorities in parallel [48, 50, 55] although progress on governance issues and infrastructure are likely needed before significant progress can be made on other priorities **(Figure 1)**.



**Figure 1.** Relationship between five key priorities for successful use of artificial intelligence by public health organizations

Initial governance activities should include clarification of data and analytic leadership roles and responsibilities, and a review of current organizational governance to ensure alignment within the larger governance context. Existing data and analytic infrastructure should be evaluated in consultation with data management, data science and AI experts to identify priorities for modernization and identify places where a small early investment may have a large impact. An early focus on data standardization and documentation may have long-term benefits to data interoperability within the organization. Pilot projects evaluating use of new infrastructure and/or advanced analytic methods should be initiated in several application areas. For example, the United States NIH piloted use of a cloud computing environment with a small number of test datasets to establish the architecture, policies and processes for storage, sharing and analysis of data through the NIH Data Commons Pilot.[53]

To begin to address the skills gap and establish a workforce educated in data and analytic skills, desired skills and competencies should be identified and forecasted, and existing skills and capacity reviewed to inform the development of employee training and targeted hiring programs. Training of existing and new employees at all levels of seniority in data literacy, bias and equity



should be prioritized early, as organizational culture changes slowly. Areas where partnerships may be beneficial should be identified and relationship-building prioritized. Organizations should consider use of an existing ethical AI framework to guide AI activities and default to transparent data and analytic processes and following open science principles whenever possible. Access to practical guidelines for AI development, evaluation, and implementation should be ensured.

## **Conclusion**

To successfully realize the potential for AI to improve public health it is important for public health organizations to thoughtfully develop strategies for AI implementation. This should include review and modernization of exiting organizational data and analytic governance and infrastructure, addressing the AI and data science skills gap, development of strategic collaborative partnerships, and use of AI best practices including explicit consideration of equity.

## **CONTRIBUTIONS**

Both authors contributed equally to the conception and design of the work. SF performed the environmental scan and drafted the manuscript. LR critically reviewed and revised the intellectual content. Both authors have approved the final manuscript and agree to be accountable for all aspects of the work.

## **DECLARATION OF INTERESTS**

The authors declare no conflicts of interest.

## REFERENCES

1. Khoury, M.J., M.F. Iademarco, and W.T. Riley, *Precision Public Health for the Era of Precision Medicine*. Am J Prev Med, 2016. **50**(3): p. 398-401.
2. Mooney, S.J. and V. Pejaver, *Big Data in Public Health: Terminology, Machine Learning, and Privacy*. Annu Rev Public Health, 2018. **39**: p. 95-112.
3. Shaban-Nejad, A., M. Michalowski, and D.L. Buckeridge, *Health intelligence: how artificial intelligence transforms population and personalized health*. NPJ Digit Med, 2018. **1**: p. 53.
4. Gupta, A. and R. Katarya, *Social media based surveillance systems for healthcare using machine learning: A systematic review*. J Biomed Inform, 2020. **108**: p. 103500.
5. Graham, D.J. and J.A. Hipp, *Emerging technologies to promote and evaluate physical activity: cutting-edge research and future directions*. Front Public Health, 2014. **2**: p. 66.
6. Bublitz, F.M., et al., *Disruptive Technologies for Environment and Health Research: An Overview of Artificial Intelligence, Blockchain, and Internet of Things*. Int J Environ Res Public Health, 2019. **16**(20).
7. Panch, T., et al., *Artificial intelligence: Opportunities and risks for public health*. Lancet Digital Health, 2019. **1**(1): p. e13-e14.
8. Neill, D.B., *New directions in artificial intelligence for public health surveillance*. IEEE Intell Syst, 2012. **27**(1): p. 56-59.
9. Dolley, S., *Big Data's Role in Precision Public Health*. Frontiers in Public Health, 2018. **6**(68).
10. Centers for Disease Control and Prevention, *The 10 essential public health services*. 2020: USA.
11. Niiler, E. *An AI Epidemiologist Sent the First Warnings of the Wuhan Virus*. 2020 September 6, 2020]; Available from: <https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/>.
12. Bogoch, I., et al., *Anticipating the international spread of Zika virus from Brazil*. Lancet, 2016. **387**(10016): p. 335-336.
13. Potash, E., et al., *Validation of a Machine Learning Model to Predict Childhood Lead Poisoning*. JAMA Netw Open, 2020. **3**(9): p. e2012734.
14. Ward, P.J., et al., *Enhancing timeliness of drug overdose mortality surveillance: A machine learning approach*. PLoS One, 2019. **14**(10): p. e0223318.
15. Muller, M.M. and M. Salathe, *Crowdbreaks: Tracking Health Trends Using Public Social Media Data and Crowdsourcing*. Front Public Health, 2019. **7**: p. 81.
16. Serban, O., et al., *Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification*. Inf Process Manag, 2019. **56**(3): p. 1166-1184.
17. Chiolerio, A. and D. Buckeridge, *Glossary for public health surveillance in the age of data science*. J Epidemiol Community Health, 2020. **74**(7): p. 612-616.
18. Lucius, N., et al., *Predicting E. coli concentrations using limited qPCR deployments at Chicago beaches*. Water Res X, 2019. **2**: p. 100016.
19. Moller, F.T., K. Molbak, and S. Ethelberg, *Analysis of consumer food purchase data used for outbreak investigations, a review*. Euro Surveill, 2018. **23**(24).
20. Sadilek, A., et al., *Machine-learned epidemiology: real-time detection of foodborne illness at scale*. NPJ Digit Med, 2018. **1**: p. 36.

21. Kreps, G.L. and L. Neuhauser, *Artificial intelligence and immediacy: designing health communication to personally engage consumers and providers*. Patient education and counseling, 2013. **92**(2): p. 205-210.
22. Chu, K.H., et al., *Identifying Key Target Audiences for Public Health Campaigns: Leveraging Machine Learning in the Case of Hookah Tobacco Smoking*. J Med Internet Res, 2019. **21**(7): p. e12443.
23. Shen, Y., et al., *Detecting risk of suicide attempts among Chinese medical college students using a machine learning algorithm*. J Affect Disord, 2020. **273**: p. 18-23.
24. Roy, A., et al., *A machine learning approach predicts future risk to suicidal ideation from social media data*. NPJ Digit Med, 2020. **3**: p. 78.
25. Stein, N. and K. Brooks, *A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults*. JMIR Diabetes, 2017. **2**(2): p. e28.
26. Neuhauser, L., et al., *Using design science and artificial intelligence to improve health communication: ChronologyMD case example*. Patient Educ Couns, 2013. **92**(2): p. 211-7.
27. Grolleman, J., et al. *Break the habit! designing an e-therapy intervention using a virtual coach in aid of smoking cessation*. in *International Conference on Persuasive Technology*. 2006. Springer.
28. World Health Organization. *Using AI to quit tobacco*. 2020; Available from: <https://www.who.int/news-room/spotlight/using-ai-to-quit-tobacco>.
29. Ravaut, M., et al., *Development and Validation of a Machine Learning Model Using Administrative Health Data to Predict Onset of Type 2 Diabetes*. JAMA network open, 2021. **4**(5): p. e2111315-e2111315.
30. Lu, X.H., et al., *Application of Machine Learning and Grocery Transaction Data to Forecast Effectiveness of Beverage Taxation*. Stud Health Technol Inform, 2019. **264**: p. 248-252.
31. Meystre, S.M., et al., *Automatic de-identification of textual documents in the electronic health record: a review of recent research*. BMC Med Res Methodol, 2010. **10**: p. 70.
32. Zulfiqar, S., et al., *Is Machine Translation a Reliable Tool for Reading German Scientific Databases and Research Articles?* J Chem Inf Model, 2018. **58**(11): p. 2214-2223.
33. O'Mara-Eves, A., et al., *Using text mining for study identification in systematic reviews: a systematic review of current approaches*. Syst Rev, 2015. **4**(1): p. 5.
34. Chen, I.Y., et al., *Ethical Machine Learning in Health*. ArXiv, 2020. **abs/2009.10576**.
35. Obermeyer, Z., et al., *Dissecting racial bias in an algorithm used to manage the health of populations*. Science, 2019. **366**(6464): p. 447-453.
36. Mamiya, H., A. Shaban-Nejad, and D.L. Buckeridge, *Online public health intelligence: ethical considerations at the big data era*, in *Public Health Intelligence and the Internet*. 2017, Springer. p. 129-148.
37. Dwork, C. and A. Roth, *The algorithmic foundations of differential privacy*. Foundations and Trends in Theoretical Computer Science, 2014. **9**(3-4): p. 211-407.
38. Santos-Lozada, A.R., J.T. Howard, and A.M. Verdery, *How differential privacy will affect our understanding of health disparities in the United States*. Proc Natl Acad Sci U S A, 2020. **117**(24): p. 13405-13412.
39. Linardatos, P., V. Papastefanopoulos, and S. Kotsiantis, *Explainable AI: A Review of Machine Learning Interpretability Methods*. Entropy (Basel), 2020. **23**(1).

40. Morgenstern, J.D., et al., *Predicting population health with machine learning: A scoping review*. BMJ Open, 2020. **10**(10): p. e037860.
41. Christodoulou, E., et al., *A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models*. J Clin Epidemiol, 2019. **110**: p. 12-22.
42. Lavigne, M., et al., *A population health perspective on artificial intelligence*. Healthc Manage Forum, 2019. **32**(4): p. 173-177.
43. Bi, Q., et al., *What is Machine Learning? A Primer for the Epidemiologist*. Am J Epidemiol, 2019. **188**(12): p. 2222-2239.
44. Liu, Y., et al., *How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature*. Jama, 2019. **322**(18): p. 1806-1816.
45. Alsheibani, S., Y. Cheung, and C. Messom. *Artificial Intelligence Adoption: AI-readiness at Firm-Level*. in PACIS. 2018.
46. Baker, J., *The technology–organization–environment framework*. Information systems theory, 2012: p. 231-245.
47. Rosella, L.C., et al., *Building a data science platform for better health [White paper]*. 2020.
48. Statistics Canada. *Statistics Canada Data Strategy: Delivering insight through data for a better Canada*. 2019; Available from: <https://www.statcan.gc.ca/eng/about/datastrategy>.
49. Health Canada, *Health Canada Data Strategy*. 2019: Internal report: unpublished work.
50. Public Health Agency of Canada, *PHAC Data Strategy*. 2019.
51. Stukel, T.A., et al., *Envisioning a data science strategy for ICES*. 2017, Institute for Clinical Evaluative Sciences: Toronto, ON.
52. Public Health Ontario, *Informatics Strategy*. (In development): Toronto, Ontario.
53. National Institutes of Health, *NIH Strategic Plan for Data Science*. 2018.
54. U.S. Department of Health and Human Services Data Council, *2018 HHS Data Strategy: Enhancing the HHS Evidence-Based Portfolio*. 2018.
55. NHS Digital, *Data and information strategy*. 2016.
56. Pan American Health Organization. *Public health in the information society*. 2017; Available from: <https://www.paho.org/salud-en-las-americas-2017/?p=75>.
57. Task Force on Artificial Intelligence for Health (AI4Health). *Building a learning health system for Canadians: Report of the Artificial Intelligence for Health Task Force*. 2020; Available from: <https://cifar.ca/ai/national-program-of-activities/ai4health-task-force/>.
58. Canadian Institute for Advanced Research and Canadian Institutes of Health Research. *AI for public health equity. Workshop Report*. 2019; Available from: [https://cihr-irsc.gc.ca/e/documents/ai\\_public\\_health\\_equity-en.pdf](https://cihr-irsc.gc.ca/e/documents/ai_public_health_equity-en.pdf).
59. Canadian Institute for Advanced Research and Canadian Institutes of Health Research. *Applications of artificial intelligence approaches to tackle public health challenges. Workshop Report*. 2018; Available from: [https://cihr-irsc.gc.ca/e/documents/artificial\\_intelligence\\_approaches-en.pdf](https://cihr-irsc.gc.ca/e/documents/artificial_intelligence_approaches-en.pdf).
60. Joshi, I. and J. Morley, *Artificial intelligence: How to get it right. Putting policy into practice for safe data-driven innovation in health and care*. 2019, NHSX: London, United Kingdom.
61. Academy of Medical Royal Colleges, *Artificial intelligence in health care*. 2019, AoMRC London.

62. The Center for Open Data Enterprise. *Sharing and utilizing health data for AI applications: Roundtable report*. 2019; Available from: <https://www.hhs.gov/sites/default/files/sharing-and-utilizing-health-data-for-ai-applications.pdf>.
63. Matheny, M., et al., *Artificial intelligence in health care: The hope, the hype, the promise, the peril*. NAM Special Publication. Washington, DC: National Academy of Medicine, 2019: p. 154.
64. JASON, *Artificial Intelligence for Health and Health Care*. 2017: McLean, VA.
65. Forcier, M.B., et al., *Integrating artificial intelligence into health care through data access: can the GDPR act as a beacon for policymakers?* Journal of Law and the Biosciences, 2019. **6**(1): p. 317-335.
66. Government of Canada Department of Justice. *Modernizing Canada's Privacy Act*. March 8, 2021]; Available from: <https://www.justice.gc.ca/eng/csj-sjc/pa-lprp/modern.html>.
67. Jobin, A., M. Ienca, and E. Vayena, *The global landscape of AI ethics guidelines*. Nature Machine Intelligence, 2019. **1**(9): p. 389-399.
68. Smith, M.J., et al., *Four equity considerations for the use of artificial intelligence in public health*. Bulletin of the World Health Organization, 2020. **98**(4): p. 290.
69. Abadi, M., et al. *Tensorflow: A system for large-scale machine learning*. in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*. 2016.
70. Ridsdale, C., et al., *Strategies and best practices for data literacy education: Knowledge synthesis report*. 2015.
71. Statistics Canada, *Framework for responsible machine learning processes at Statistics Canada July 2020*. 2021.
72. Statistics Canada. *Data Literacy Training Products*. 2021 August 2, 2021]; Available from: <https://www150.statcan.gc.ca/n1/en/catalogue/89200006#wb-auto-2>.
73. Government of Canada. *Digital Academy*. 2020 March 28, 2021]; Available from: <https://www.cspc-efpc.gc.ca/digital-academy/index-eng.aspx>.
74. McMahan, M., et al., *Developing Competencies for Health System Impact: Early Lessons Learned from the Health System Impact Fellows*. Health Policy, 2019. **15**(Sp): p. 61-72.
75. Canadian Institutes of Health Research. *Health System Impact Fellowship - AI and Public Health stream*. 2019 March 10, 2021]; Available from: <https://cihr-irsc.gc.ca/e/51332.html>.
76. CIHR Institute of Population and Public Health, *Strategic Plan Refresh 2015-2018 Health Equity Matters*. 2015: Ottawa, ON, Canada.
77. Department of Health and Social Care. *A guide to good practice for digital and data-driven health technologies*. 2021; Available from: <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology>.
78. Government of Canada, *Report to the Clerk of the Privy Council: A Data Strategy Roadmap for the Federal Public Service*. 2018.
79. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific data, 2016. **3**(1): p. 1-9.
80. Google. *Artificial Intelligence at Google: Our principles*. 2018; Available from: <https://ai.google/principles>.

81. Whittaker, M., et al., *AI now report 2018*. 2018: AI Now Institute at New York University New York.
82. Collins, G.S., et al., *Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement*. *J Clin Epidemiol*, 2015. **68**(2): p. 134-43.
83. Collins, G.S. and K.G.M. Moons, *Reporting of artificial intelligence prediction models*. *Lancet*, 2019. **393**(10181): p. 1577-1579.
84. Murphy, K., et al., *Artificial intelligence for good health: a scoping review of the ethics literature*. *BMC Med Ethics*, 2021. **22**(1): p. 14.
85. Moons, K.G., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration*. *Annals of internal medicine*, 2015. **162**(1): p. W1-W73.
86. Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. 2009: Springer Science & Business Media.
87. Kuhn, M. and K. Johnson, *Applied predictive modeling*. Vol. 26. 2013: Springer.