Artificial Intelligence in Health Care

Promises and Challenges

IDB

We thank Marcelo D'Agostino, Myrna Marti, Cristina Pombo, Natalia Gonzalez, and Roberto Sanchez for their valuable comments.

Author: Christina Silcox
Design: www.souvenirme.com

Artificial Intelligence in Health Care

Promises and Challenges

Table of Contents

Executive Summary

This report is for those who want to learn more about how artificial intelligence (AI) is used in health care, and particularly AI's use in clinical care. The report describes the potential for AI to improve individual and public health through improved access, better clinical decision-making, and newly created efficiencies in clinical care and research. It then describes the differences between more traditional medical devices and products that use AI, discusses how countries are considering regulating these products, and examines specifics risks in what are likely to be the most popular health care uses of AI.

Key Takeaways

- AI is a broad category of software that performs tasks that have conventionally been done by humans. There are two main AI strands: symbolic AI uses knowledge-based software that relies on expert-derived rules to produce outputs, and subsymbolic AI, which includes machine-learning software, relies on algorithms to extract patterns from data. People commonly use "AI" as a shorthand for machine learning.

- Software is used in many aspects of health care, and many of these products use AI or are beginning to add functionality based on AI.

- Countries need to consider what risks exist with AI-enabled software products by assessing the potential impact on patient care and what regulations may be needed. Higher-risk products may need to meet higher standards and therefore be regulated as medical products, but other types of products with less risk may not.

- AI depends heavily on digitized health data. The availability, coding (terminology), and reliability of health data are highly variable, which can slow the development of effective AI and cause performance issues when the AI is widely implemented. Accordingly, countries whose digital health data infrastructure is more standardized will have an advantage when working with AI.

- AI has made health data simultaneously more valuable and easier to abuse. Countries need to consider how to balance privacy rights while promoting innovation.

- Data can be biased, so AI-enabled health care products should be rigorously tested to understand their performance across populations with medical and demographic differences.

- AI's fast update cycles and its frequently "black box" nature (the software reaches conclusions that are so mathematically complex they cannot always be understood by humans) mean that traditional regulatory regimes for medical products may not be sufficiently risk-averse or timely. Governments should provide clarity on applicable laws and regulations and consider new methods of regulations that take these challenges into account.

- Administrative and procedural systems also need to be established to monitor these AI systems during and after their implementation to determine if the software's performance in a particular real-world population meets expectations. The performance also needs to be examined periodically to ensure it does not degrade over time.

- Broad education on AI's potential and limitations is essential to safely deploying these products in health care systems.

Introduction

AI has enormous potential to improve population health as well as increase efficiencies around health care delivery, and it is poised to be an important tool in the continued transformation of health care over the next decade. Funding for AI in health care reached $3 billion in 2018, $4.09 billion in 2019—with record-setting third-quarter funding of $1.6 billion—and $2.1 billion in the first half of 2020 (CB Insights, 2020). Companies are using AI throughout the health care sector, including to triage patients, screen and diagnose diseases, recommend treatments based on precision medicine, improve administrative workflows, find areas of fraud and waste in the system, and develop and test new drugs.

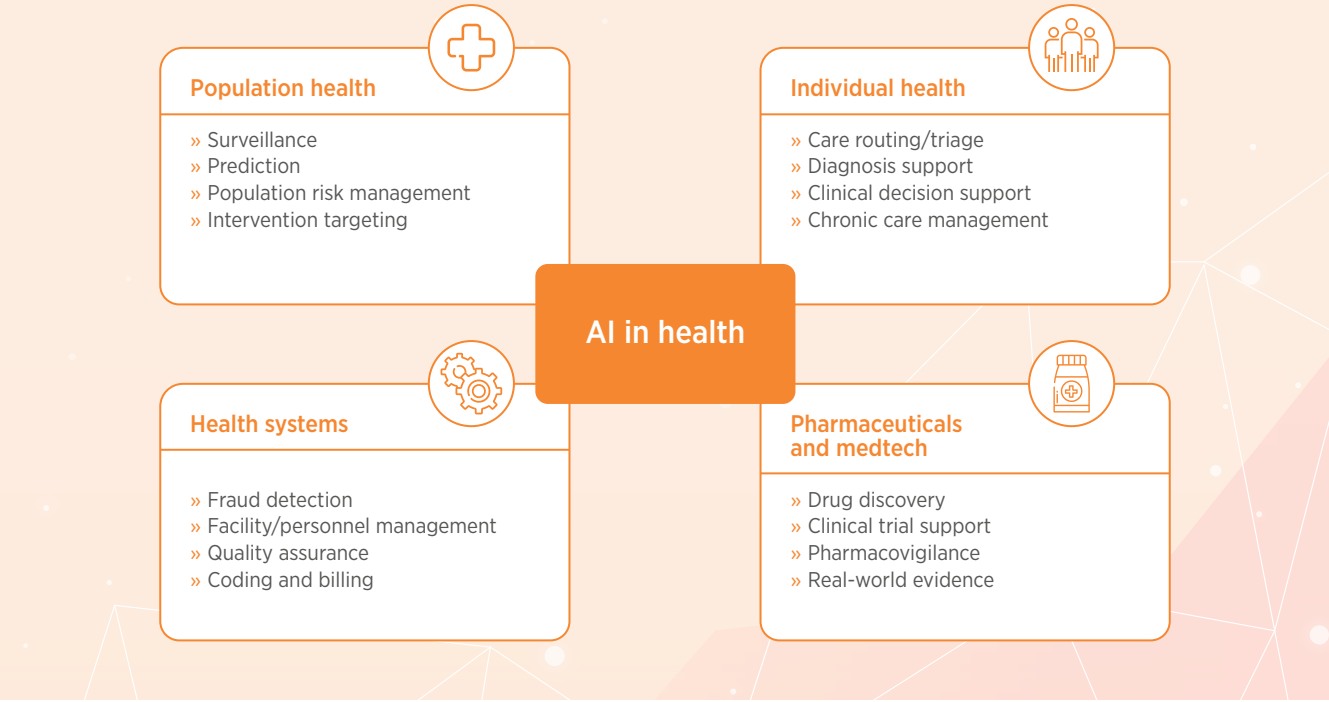
While there is no universal definition of AI, it can generally be considered a broad category of software that is able to perform tasks normally done by humans. This report focuses on the two large strands of AI known as symbolic and subsymbolic AI. Symbolic AI software is also called knowledge-based software. Symbolic AI software uses logical rules and axioms programmed into the software by people to make deductions, thereby reaching optimal decisions through predefined rules within a specific domain (Pombo et al., 2020). Subsymbolic AI[1] software systems learn automatically by extracting patterns and inferences. As such, subsymbolic AI, such as machine learning systems, may reach complex conclusions that cannot be obtained or explained by humans who are using the same inputs and parameters (Pombo et al., 2020).

Applications from both strands of AI have been common in health care systems for quite some time. Such systems include familiar products such as hospital staffing software and electronic alerts about potential medication interactions. However, as more health data are digitized and computer hardware becomes more sophisticated, much more powerful systems can be developed, and this potential is driving the renewed excitement around AI. In particular, machine learning systems, which can pull in more information about patients and be developed through algorithmic analysis of thousands of similar cases, have great potential to produce much more targeted and personalized health care recommendations.

[1] Subsymbolic AI is also called connectionist AI.

Figure 1 lists potential use cases in different areas of the health care ecosystem where AI is likely to have an impact in the near and medium terms, as will be discussed later in more detail.

Figure 1: AI Uses in Health Care



Source: USAID and the Rockefeller Foundation (n.d.).

While the excitement about the potential for AI to improve population and individual health is justified, more robust incorporation of AI into health care must be done carefully to protect patients from poorly performing products and avoid potentially worsening existing health disparities. The IDB is committed to bringing the public and private sectors together to ensure the responsible use of AI in Latin America and the Caribbean and is leading a platform called fAIr LAC to help guide standard development, pilot use cases, and convene discussions on inclusive, person-centered uses of AI (IDB, 2020). This platform strives to support AI development that is inclusive of vulnerable and excluded populations while respecting data privacy and people's fundamental rights (including the right to informed consent, which means that people know what data will be used and how) and building trust in AI systems. This report looks at the unique differences between AI-enabled software and traditional tools and medical devices, and it looks at the various categories of uses of AI in health care and considers what policies, regulations, and protocols may need to be developed to ensure the ethical development and adoption of AI in health care.

Given the broad scope of use cases for AI in health care, a risk-based framework should be used to understand which systems will require more evaluation and testing before and after implementation. Low-risk administrative software that does not directly affect patient care can generally be evaluated from a purely business and labor perspective, although there may be regulatory considerations that need to be built in specifically for health (for example, compliance with legally required minimum staff-to-patient ratios for staffing software used in hospitals). Where there is the potential to affect patient care, such as when selecting regions or individuals for supplemental health support, more evidence may be required on the software's performance, and adopters may need more information about how these systems were developed to understand whether biases may be present. For example, software products used to improve population health should be examined and tested carefully to ensure that they will not create or perpetuate health disparities through inadvertent bias, as discussed in the "Disparities and Bias in Health Data" section.

"Software as a Medical Device (SaMD)' is software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device."

International Medical Device Regulators Forum (2014, p. 5)

The highest-risk software is that which is meant to be used for one or more medical purposes on individual patients. National governments will need to clarify when a product would be categorized and regulated as a medical device, in which case it may be called an SaMD (software as a medical device).[2] Depending on the potential impact on patient care and the overall risk of the software, different regulations may be required for the product to enter the market and make advertising claims. It is helpful and encouraged to provide software manufacturers with clear guidelines about which regulations apply to which products.

2 "Software is a medical device if it is 'intended by the manufacturer to be used, alone or in combination, for human beings, for one or more...medical purpose(s)' such as 'diagnosis, prevention, monitoring, treatment or alleviation' of a disease or injury." (GHTF, 2005)

Using AI-Enabled Products in Health Care

Software—and particularly AI software—in health care differs in some key ways from traditional medical devices. This may affect how countries and providers think about how to best regulate, adopt, and use AI-enabled software products in health care. A recent paper from Duke University lists the three the major differences: **(1)** AI-enabled health software is built from, tested on, and used with digital health data, which are not standardized or consistent and can be biased and change over time; **(2)** AI-enabled software often lacks an explanation for how decisions are reached; and **(3)** software has a much more rapid development cycle as compared to traditional medical devices, making traditional regulation processes difficult (Silcox, Sharma and Rai, 2020).

>> How AI Uses Health Data

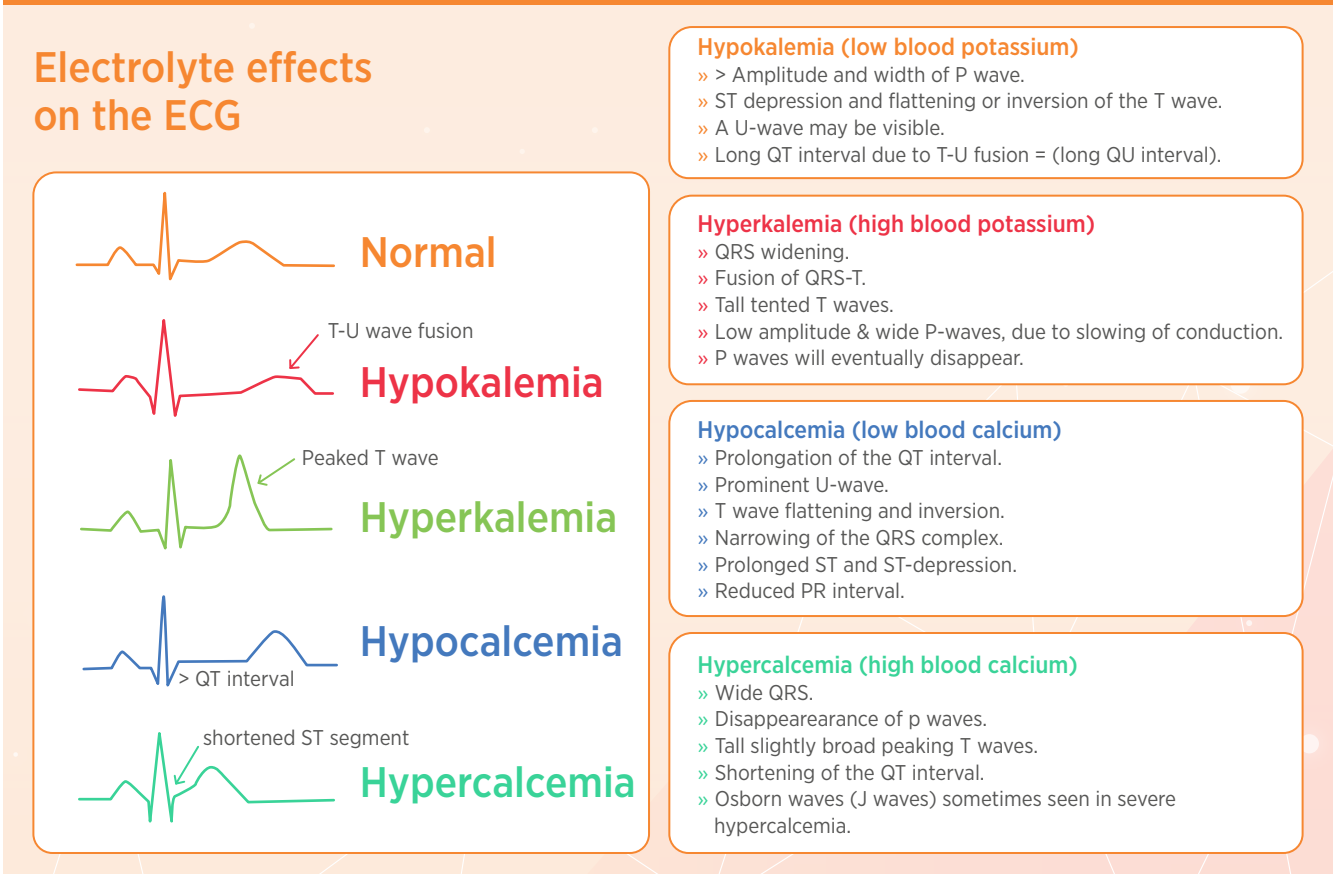
Health Software Acts on Health Data

One of the critical differences between traditional medical devices and SaMD and other AI-enabled health care software is that software is applied to patients' digital or digitized data rather than to the patients themselves. Countries with infrastructures that support digitized and interoperable health databases will, therefore, be best positioned to take advantage of AI's potential to transform health care.

Knowledge-based software generally works on limited, structured input data that are often chosen specifically because those data elements are more standardized across institutions, such as age, weight, and specific lab values. In health care, knowledge-based systems generally rely on consensus-based data such as clinical guidelines or information approved by regulatory authorities like product labeling. In contrast, machine learning software often uses a much larger dataset as input, including free text from electronic health records (EHRs) and/or imaging, device, or lab data. Using larger amounts of data can be beneficial, as it allows novel associations to be discovered. For example, as detailed in *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human*

Again, a company named AliveCor used a deep neural network (a type of subsymbolic AI) to analyze electrocardiograms (ECGs) and detected high levels of potassium (hyperkalemia) in the blood. This discovery meant that a condition that normally requires a blood test using an electrolyte panel can be continuously monitored noninvasively. Previous attempts to find a bloodless blood test used a single aspect of the ECG waveform rather than all the information in the waveform. The earlier attempts did not create clinically meaningful predictive values, but the inclusion of all information in the waveform produced a much more effective algorithm.

Figure 2: Using AI to Improve Prediction of Hyperkalemia Using ECGs



Source: USAID y Rockefeller Foundation (s.f.).
Notes: The image depicts how taller T-waves often accompany hyperkalemia in electrocardiograms. Though commonly observed, this feature was not predictive enough to be clinically meaningful until machine learning techniques applied to the full cardiac waveform significantly improved accuracy. A company called AliveCor then ran a machine learning analysis that used a deep neural network (subsymbolic AI) to produce a bloodless blood test. However, because of the complexity of the neural network, there is no way for humans to know exactly which features of the ECGs are being used in the test to make the hyperkalemia predictions.

Health data, however, can change substantially depending on where and when (and on whom and by whom and for what purpose) it was collected. Both structured and unstructured EHR data can be substantially different depending on national and regional standards, local workflow norms, insurance requirements, EHR template structures, and other factors. Imaging, device, and lab data can differ substantially depending on the test, make, and model of the equipment used and the skill of the operator. Therefore, the data can be quite heterogeneous, and at times, substantial changes must be made to a software program to accommodate how data is defined, formatted, and stored in different locations.

As such, a deep understanding of the variability of input data availability, definitions, and reliability should be considered when developers are determining what information will be used as inputs and when adopters are considering whether a system will work for their specific context. Manufacturers and adopters of software products that combine data collected from disparate sources will also need to carefully consider this issue. However, in cases where recommendations or outputs are not needed in real-time, it may be possible to check, clean, and curate the data before use.

In addition to the variability discussed above, health data changes over time as new clinical practices and medical products come into use. For example, software products that use medication lists as an input will need to account for new medicines coming onto the market. Even improvements in sensor and imaging technology (for example, newer models of CT [computed tomography] and MRI [magnetic resonance imaging] machines that create sharper contrasts) may cause software performance to significantly degrade if those improvements cause the data being entered to be significantly different than the data used to build and test the software.

Disparities and Bias in Health Data

Health data can also be biased for multiple reasons. Cultural differences may make some patients less likely to report certain symptoms that may be used by the software as input factors. Prejudices may result in the person entering information into the EHR to discount some of the symptoms described by the patient. Language challenges may limit the information offered or cause incorrect information to be entered. Economic and accessibility factors may mean certain tests or treatments were not a possibility for a particular patient—even though the software may rely on those results or use them as proxy information for disease states (particularly in machine learning). For example, a sudden change in blood pressure readings or the ordering of a certain blood test may be highly weighted by an algorithm that predicts patient

decomposition in a hospital ICU. For health care systems with fewer resources for equipment and staff, these blood pressure readings may be taken less frequently, and that lab test may not be available. Therefore, the absence of that data in the EHR could affect the accuracy of the software prediction for that patient.

For AI-enabled population health tools, in particular, more information than just health data may be required for the systems to perform adequately. For example, weather, population density, and local water system information may be critical to accurately predicting infectious disease outbreaks. This information may be more readily available and accurate in certain portions of a country.

Machine Learning Is Built with Health Data

For software developed with machine learning, the challenges around the availability, quality, and reliability of digital health data are even more critical. Knowledge-based systems use expert-derived rules and consensus-based data such as clinical guidelines or information approved by regulatory authorities (e.g., product labeling) to come to decisions, so data are not required to develop these products. Software based on supervised machine learning is built by feeding large numbers of *input-output pairs*, consisting of the patient health data that will be used as inputs in the final software and labeled with the desired output or recommendation. For example, hundreds or thousands of images of the back of patients' eyes will be labeled as showing evidence of diabetic retinopathy or not (see Text Box 3). A learning algorithm then finds the best mathematical equation to fit the data. This allows the AI-enabled algorithm to predict what the recommendation should be when new input data are entered into the software. However, the algorithm can learn from only the data it was shown. If the data used to train the algorithm (the *training data*) do not contain sufficient numbers of patients with diverse demographic and medical characteristics, the resulting software may not work as well on those populations. One of the reasons that AI has not yet had a large impact in prediction, diagnosis, and treatment of COVID-19 is that, as a novel disease, there is not yet a base of data collected to use to train systems. What data has been collected, especially in the early months was not high quality or standardized, and may also be incomplete and biased. **Researchers that have looked at AI-enabled systems that have been built report that the methodology around the training is often poorly explained** but that the amount of data used to train the systems is generally not enough to build accurate models.

Having sufficient representation of diverse populations in the dataset is only the first step to ensuring that the software will work well on all populations. As described above, the information in the recorded data may also be biased, often in ways that can be hard to predict or detect. When collecting training data, the

selection of a label that is biased can have serious consequences. As such, developers should carefully consider and disclose the methods used to select labels for the training data. For example, in fall 2019, an algorithm used in the United States to predict high-risk patients with serious health needs was found to be racially biased. To train the system, developers used the costs of each patient to the health care system as a proxy for serious health needs (Obermeyer, Powers and Vogeli, 2019). Since Black patients historically use health care services at a lower rate than other Americans, particularly White patients, Black patients had to be much more ill than White patients before they were flagged as eligible for supportive services. This particular algorithm was not an isolated case either—the researchers noted that this approach to labeling is extremely common. Another example of how training data labels can be biased is in diagnosis. It is known that heart attacks are frequently initially misdiagnosed in women because they are less common and often present differently than in men (Heim and Brunsell, 2000). Therefore, training data labeled with an initial clinical diagnosis of heart attacks will be biased if the initial diagnosis is wrong more often for women, and a software product trained with that data could end up exacerbating that bias. However, a software product trained on the same input data but more rigorously labeled could be used to combat that known bias by suggesting that diagnosis to a health care worker who might not otherwise have considered it.

Regulators and adopters of AI-enabled software should incentivize careful development and rigorous testing by requiring details about how the software was developed, what patient characteristics were used in the training data, and how that data was defined, formatted, and labeled. This information will help to ensure that the software will continue to perform as expected when implemented in new systems and used on a wider population. These systems must also be monitored over time as data collection methods change. Frequent updates may be needed to keep these products performing at the expected level. As these updates are put into place, careful attention must be made to ensure that the systems are retested and continue to work on diverse patient populations and health systems.

Policymakers can support AI innovation in health care by incentivizing **digital data capture of health information**, with careful attention to standardizing definitions and monitoring accuracy. A more universal data infrastructure will make developing and scaling software more efficient. At the same time, countries must examine how they protect the privacy of their citizen's health data. This includes both data that have traditionally been considered health data as well as new types of data such as personal activity trackers or geolocation data. Although the latter is not currently considered health data, AI is becoming increasingly effective at using it to predict future health outcomes. Countries will need to determine what sort of consent should be required to

collect this data for initial product development. Once a product is implemented into a health care system, ongoing performance tracking may be required to ensure that the software continues to work as expected in the changing data environment. Countries will need to consider who will be responsible for this performance tracking and when patient consent may be needed to access the data to allow these evaluations. Consideration will also be needed about whether developers can use the data to update the product or even create new products and how systems that collect health data can create appropriate data use agreements with developers to protect patients while also facilitating innovation. And policymakers will need to reflect on and develop regulating bodies and processes for answering these and other questions.

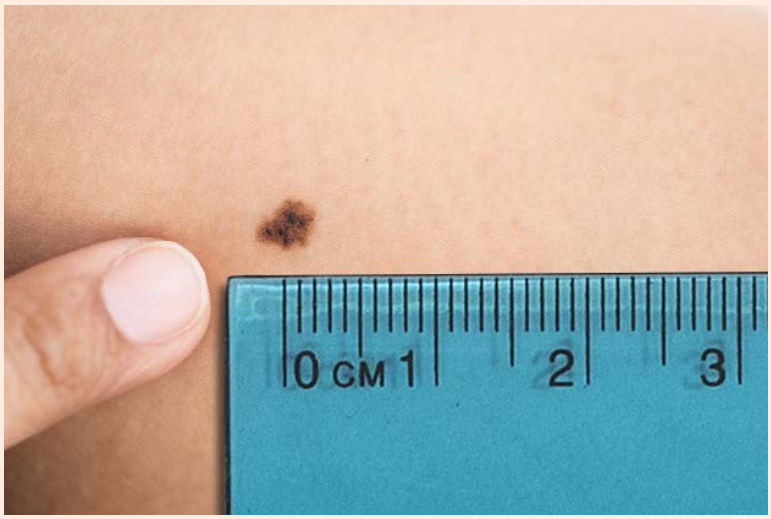
>> Explainability

As described in fAIr LAC's publication *Responsible and Widespread Adoption of Artificial Intelligence in Latin America and the Caribbean*, machine learning models may be so complex that it may not be possible for humans to understand how the models reach conclusions (Pombo et al., 2020). Software systems that produce outputs without explanation of how the output was reached are often referred to as black box models. The danger of a black box system is that the device may fail in unexpected ways or with certain populations, however this is not a situation that only occurs with AI-enabled software. There are medical devices and drugs whose mechanism or mode of action is unknown or was unknown when the product was initially approved, and rigorous performance evaluation can sometimes substitute for this lack of knowledge. Note that this "black box" challenge is quite different from a situation where the user of a device doesn't fully understand how the systems works, but the manufacturers and regulatory experts who do understand the mechanism have reviewed the device for known dangers and added appropriate warnings in the labeling.

With black box models, though, this expert understanding is lacking and the software may therefore fail in unusual and unpredictable ways. This sometimes occurs because the training data included information (clues) that the developers did not recognize as such, or because clinically relevant information was not included in the digital health data. There are several well-known examples of this happening, such as when an algorithm trained to predict whether moles were cancerous was heavily influenced by whether there was a ruler in the picture with the mole (see **Figure 3**) (Esteva et al., 2017). Another is an algorithm that analyzed chest x-rays for signs of pneumonia and was found to depend on whether the x-ray machine used was a portable model or not (Couzin-Frankel, 2019). Additionally, an algorithm that predicted whether patients who present to the hospital with pneumonia would develop complications was never put into clinical use

because it incorrectly assessed patients with a history of asthma as having a decreasedrisk of complications (Caruna et al., 2015). It is clinically well-known that asthma patients are at increased risk of pneumonia, and therefore, they receive higher-level care than nonasthmatic patients when they are admitted to a hospital. The higher level of care given to these patients prevents the poor outcomes that would have been expected from this flawed machine learning algorithm had it been implemented, but the algorithm simply did not have the information required to "learn" about the importance of advanced care for these patients.

Figure 3: How Predictive Software May Misinterpret Medical Imagery



Notes: Image for illustrative purposes only, does not represent an image used in a study.
Developers of early software to predict which moles may be cancerous used clinical photographs to train their algorithms. However, clinicians are more likely to include a ruler in these clinical images if, in their clinical judgment, the mole was more likely to be cancerous. It was later determined that software developed with these types of photographs heavily weighed the presence or absence of a ruler in the photo to make the prediction rather than using features related to the mole.

Often these types of mistakes are found when the software is tested with new data. Ideally, this sort of testing should occur with prospectively collected data in a situation where the AI is incorporated into a real-world workflow. When a medical product lacks an explanation for how it works, careful and extensive performance testing is the only way to discover its weaknesses. Manufacturers may also be able to give users information about how certain an algorithm's prediction is or what factors were most important to the decision, which may give a user insight into when a particular recommendation should not be trusted—but that limited amount of information is still not a substitute for an explanation of how the software works. People are also working on systems that attempt to come to statistically likely explanations of how an unexplainable algorithm is working, but that information should also not be considered the equivalent of a complete explanation (Rudin, 2019).

Because of the factors discussed above, black box algorithms may not be appropriate for every use, and they should always be considered higher risk than an understandable algorithm that performs the same function. Therefore, when choosing between black box software and its explainable equivalent software, it is optimal to choose the software whose functionality can be understood. Entities may also use AI-enabled software to select patients to receive certain services or allow additional benefits. If the results of such software prevent patients from gaining what they see as the services they require, the patients may understandably feel that they deserve an explanation for why the service was rejected. Concerns about bias and mistakes would be reasonable, and, depending on the applicable laws in that situation, the patients may have a right to appeal that decision—something that would have no meaning without an understanding of the standards or rules that determine why the decision went against someone. However, in some circumstances, where the decision is less critical or the carefully demonstrated performance of an unexplainable algorithm greatly exceeds that of other available options, these types of algorithms may still be a good choice.

Liability is an additional consideration for adopters of black box algorithms. An article from the *Journal of the American Medical Association* discussed this issue, and the authors note that there is essentially no case law on this yet. However, they also note that if physicians follow a recommendation from AI software that contradicts the typical standard of care, the physicians are opening themselves up to liability for any bad outcomes that may result from the recommendations (Price, Gerke and Cohen, 2019). If the software does not give a reason for why a physician should deviate from the standard of care, it might be difficult for the physician to justify doing without having already planned to make that recommendation. However, as AI improves and becomes incorporated into standards of care, it is possible to conceive of a doctor being held liable for not following the advice of the software.

Developers of AI-enabled software should keep this liability consideration in mind, as it will likely affect the willingness of clinicians to adopt and use such software. Countries may also want to reexamine their laws regarding liability, particularly for black box software and explainable software whose users are not intended to have the expertise to evaluate the software's predictions or recommendations.

>> Faster Software Development Cycles

Software is updated on a much more frequent schedule than traditional medical devices, especially since AI algorithms developed with machine learning have the ability to update continuously. This can be a benefit for systems that are designed to change to accommodate a user's preferences (for example, a behavioral therapy app that learns what types of prompts are more likely to affect change for a particular user) or applications in areas where knowledge changes quickly (for example, cancer treatment options). But there are also obvious implications for evaluation or regulation as well as for making sure the user is informed of any changes that may affect the safe and effective use of the product. In most cases, changes to a medical device require the manufacturer to go back to the relevant authorities and show that the changes have either not affected or have improved the performance of the device. Continuous updating, or even monthly updating, would not be practical within a traditional regulatory paradigm. However, one of the strengths of machine learning is its ability to learn from real-world feedback.

Also, as discussed below, performance can degrade over time without updates from manufacturers. The United States Food and Drug Administration (FDA) released a discussion paper that imagines a novel "total product lifecycle (TPLC) regulatory approach [for machine learning–based SaMD] that facilitates a rapid cycle of product improvement and allows these devices to continually improve while providing effective safeguards," and for the past few years, the FDA has been exploring a manufacturer-based total product lifecycle option for regulating SaMD (FDA, 2019a, 2019b). For software that is not a medical device, best practices and standards on how manufacturers should communicate changes and how users should consider these changes as they continue to use the software will need to be developed.

>> How to Evaluate Risk Regarding AI-Enabled Software Products?

In the discussion regarding explainability above, we suggest that unexplainable (black box) software should be considered higher risk than explainable software. Multiple groups are working on best practices and risk frameworks around AI-enabled health care software, but there are no general-consensus standards yet. The International Medical Device Regulators Forum (IMDRF) has documented other factors that should be considered regarding SaMD, and these may help with thinking about AI-enabled health care software more generally as well. A 2014 IMDRF framework on risk categorization (IMDRF, 2014) and a 2017 IMDRF document on the clinical evaluation of SaMD (IMDRF, 2017) suggest that two

major factors should be used to categorize an SaMD product and its intended use. The first is the significance of the information provided by the SaMD to the health care decision. The second is the state of the health care situation or condition.

The first factor, the significance of the information provided, considers the level of influence that the software has on the next clinical intervention, the dependence of the user on that software, and the ability of the user to detect erroneous outputs. SaMD that causes an "immediate or near-term action," through either direct connection to other medical devices or information provided to a user who does not have the expertise to evaluate that recommendation would be higher risk. For example, there are software products that allow health care technicians to photograph the eye of a diabetic patient to diagnose whether the patient has diabetic retinopathy, a diagnosis that normally must be done by a specialist (FDA, 2018). In contrast, SaMD that that acts to "guide next diagnostics or next treatment interventions" by providing enhanced support of medical products, analyzing relevant information to predict risk or aid in making a definitive diagnosis, or triaging patients would be of medium risk. If the software simply triaged patients for final diagnosis by a specialist rather than making the diagnosis directly, it would be considered lower risk[3], as would software that simply informs a user of available options or aggregates relevant information for clinicians.

The second factor, state of the health care situation or condition, considers the clinical context, including a patient's fragility and disease/condition progression. SaMD is higher risk in critical situations when "accurate and/or timely diagnosis or treatment action is vital to avoid death, long-term disability or other serious deterioration of health." Next is serious situations, where "accurate diagnosis or treatment is of vital importance to avoid unnecessary interventions (e.g., biopsy) or timely interventions are important to mitigate long-term irreversible consequences." Finally, non-serious situations occur when "accurate diagnosis and treatment is important but not critical for interventions to mitigate long-term irreversible consequences."

These two factors act in relative significance to each other, and the IMDRF formulated a table (see **Table 1**) to show how the factors interact to inform the impact and risk of a software device.

[3] However, regulators may want to examine how the software is likely to be used. Software that triages patients in a low-resource environment where access is an issue may result in images of patients categorized as "low-risk" by the software never making it to the "top of the pile" to be manually examined by over-burdened specialists.

Table 1: IMDRF Categorization of Health Care Software Risks			
State of Healthcare situation or condition	Significance of information provided by SaMD to healthcare decision		
	Treats or diagnoses	Drives clinical management	Informs clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

Source: IMDRF (2014, p. 14).
Notes: Category IV risks have the highest impact and category I have the lowest.

In addition to the considerations around explainability, clinical context, and how the information will be used, a final risk factor that should be considered is the software user and the information provided at the point of use, particularly when the information is meant to inform or drive decision-making. Is the information presented in a way that helps the user decide how to appropriately contextualize any recommendations? For example, for a product that makes diagnosis or treatment recommendations, such information could include the key input data that informed a recommendation, any missing data that might affect the recommendation, the certainty of the recommendation, or pointers to clinical guidelines or generally accepted literature that support such a recommendation. The usability of software products can have a large effect on the risk of user error.

While the IMDRF framework applies to medical devices, the same framework can be used to think about the risk of AI-enabled software that is not categorized as a medical device. In general, these are low-risk software that drive better general health or increase administrative efficiencies. For example, the United States has determined that general wellness software (sleep tracking, nutrition advice, etc.), EHRs, population health management tools, and simple clinical decision support tools that allow users to understand the basis of the recommendation (for example, by identifying the relevant published clinical guideline) are not medical devices (21st Century Cures Act, 2016). Because these software types do not meet the legal definition of a medical device, the FDA does not have authority over them. However, the United States Federal Trade Commission will still monitor claims made by the manufacturers when advertising products and methods the company uses to protects data privacy (FTC, 2016a).

At the same time, these products can have an enormous effect on population health if deployed widely. For example, administrative software could drive health care resources to certain areas that, if it turns out users were misdirected, could have a substantial health effect. This framework, which considers how high a factor the software recommendation is in the ultimate decision, the immediacy and irreversibility of that decision, and the impact of inaccurate recommendations on the overall public health, can also serve as a guide in those cases when determining the benefit-risk ratio of implementing a particular AI-enabled software solution.

To encourage safe and effective innovation of nonmedical-device software, where the manufacturers are often new to medical product development, regulatory clarity should be encouraged. For example, several United States agencies worked together to create an **online tool** to help mobile app developers understand which federal laws may apply to their products based on use cases and types of data being entered (FTC, 2016b).

Categories and Use Cases Involving AI-Enabled Software

Despite the challenges facing stakeholders in developing, regulating, and scaling AI in health care, AI has the potential to expand access to care, improve health outcomes, create more efficient work processes, and lower health costs. **Figura 4** shows a framework developed by the United States Agency for International Development in spring 2019 to highlight many of the common use cases for AI in health care in the areas of population health, individual health, health systems, and pharmaceuticals and medtech. Below we describe each category, talk about risks within each, and give examples of companies working on these issues in South America.

Figure 4: Framework for AI Use Cases in Health Care

POPULATION HEALTH	**1A** Surveilance and prediction	**1B** Population risk management		**1C** Intervention selection	**1D** Intervention targeting
INDIVIDUAL HEALTH	Care routing	**2A** Self-referral		**2B** Triage	**2B** Personalized outreach
	Care services	Prevention	Diagnosis	Acute treatment	Follow-up and chronic treatment
		3 Behavior change • Exercise • Diet • Wellness • Education	**4A** Data-driven diagnosis • symptom-based • lab-based **4B** Image-based diagnosis • Radiology • Pathology	**5A** Clinical decision support • Treatment guidance • Medication prescribing **5B Monitoring:** Inpatient monitoring, device monitoring **5C** AI-facilitated care: Self-care guidance, psych counseling **5D AI-facilitated care:** Robotic surgery, robotic PT	**5A** Monitoreo de cumplimiento • Cumplimiento de medicamentos • Cumplimiento de rehabilitación • Cumplimiento de dieta
HEALTH SYSTEMS	**7A** Medical records **7D** Fraud prevention	**7B** Capacity planning and personnel management **7E Quality assurance and training**			**7C** Claims processing **7F** Coding and biling
PHARMA & MEDTECH	**8A** Clinical trial support and recruitment **8D** Supply chain and planning optimization	**8B** Drug discovery **8E** Process optimization		**8C** Drug safety and pharmacovigilance **8F** Real world evidence and HEOR	

Source: USAID and the Rockefeller Foundation (n.d.).

>> Population Health

AI-enabled population health tools use AI to make predictions on the population as a whole to allow for resources to be transferred to appropriate areas or a subset of patients. Some products in this category can predict the emergence and spread of infectious disease outbreaks. For example, an algorithm developed at Boston Children's Hospital uses news sources, official reports, social media, and other sources to detect emerging infectious diseases—and it raised one of the first alarms about the novel coronavirus COVID-19 in December 2019 (Cho, 2020). Another AI-based health monitoring program called BlueDot predicted the outbreak early and predicted where the virus would move next using global airline ticketing data (Niiler, 2020).

Other products in the population health category include tools that can be used to predict which patients may shortly be at higher risk of needing costly interventions that may be prevented with earlier intervention. End-users of such products would, for example, be administrators redirecting workers or resources or clinicians receiving notes about their patients who require screenings or vaccinations.

Because these tools are working on a population level, these types of products are generally considered to be lower risk than products that affect decisions regarding individual patients. However, as resources are not limitless, mistaken predictions may divert attention from where it may be needed most, and users may be able to learn to game the systems to their benefit. The latter is easier to do in knowledge-based systems, where the basis of how decisions are made is clearly defined, although the exact process may be proprietary. However, machine learning methods can also be exploited, as demonstrated by industries being created to help people and companies optimize their position in search results and timelines on Google, Twitter, and YouTube.

If a computer system will manage populations across multiple health systems, the interoperability of the data and the potential for bias should be examined carefully before implementation. As discussed above, adopters should carefully examine how the product performs differently across regions and subpopulations and think through how the software recommendations will factor into final decisions. For example, software that makes recommendations about resource allocation can be used by administrators to work more efficiently, with the software recommendation considered as one factor in their decision-making. In this case, it is still important for the users to be trained on both well-understood and potential limitations of the software product, and continuing education should be encouraged as the product is updated.

However, another use case may involve software recommendations being implemented automatically without additional oversight. In that case, will the communities affected be informed that specific resources are being managed through AI-enabled software? Will they be given information regarding any differential performance of the software across regions and subpopulations and an explanation of how the software works (if known)? Will there be an opportunity for people to suggest alternative software products or appeal decisions that affect them?

>> Individual Health

Tools that affect individual health are the most likely products to be considered as medical devices, but there are low-risk contexts where the products give general health and wellness information or supply scientifically accepted health information that may not be categorized as such. If a software is considered to be a medical device, different levels of regulation might be considered based on the risk of the device and depending on the laws within individual countries. As explained below, an aspect of risk consideration should focus on whether the device's intended user is a patient, a frontline health care worker, or a physician.

Patient as User

One common type of health software directed at patients is virtual health assistant "chatbots." These products can vary enormously in function, with some serving as information sources, wellness coaches, providers of chronic care management assistance, or even first-line contacts advising patients on when to seek professional assistance. Note that these systems are not meant to replace doctors or other health care workers but to help patients appropriately decide when to seek care and to augment the care provided by the health systems already in place. Such chatbots have been used extensively in the COVID pandemic to help patients understand what symptoms may indicate they may have the disease and what steps to take next (Miner, Laranjo and Kocaballi, 2020).

For these systems to be effective, it is very important to provide information at the health literacy level of the population to whom the product is marketed. In countries with more than one common language, these apps may need to offer multiple language options and potentially even certain dialect options to be broadly usable by the population. These apps can also be customized and targeted to certain underserved populations as a way of reducing health disparities. However, if these apps need to be online to work or regularly update themselves, regional internet access and cellphone network coverage may limit usefulness.

That said, these can be enormously valuable tools—particularly in countries or regions where access to doctors and other health care workers is especially limited—by triaging patients and freeing up health care workers' time for patients that truly need the help (see Box 1 for example). In this way, even software products that may not be accessible for all populations can benefit everyone. These software systems may also be bundled with telemedicine services.

The risk posed by these types of products can vary dramatically. Wellness-type apps that only provide guidance on healthy lifestyle behaviors are considered relatively low risk, while a software program that recommends whether a patient should see a doctor for a particular health concern is higher risk. Creating a product that can be trusted is critical given the amount of erroneous information available for free on the internet. We may begin to see "free" (i.e., ad-supported) AI decision-support tools, and schools and doctors may need to start to teach students and patients what type of information they should research about a tool before trusting the result. For tools that are not regulated as medical devices, countries may want to consider if other entities should have some regulatory authority regarding false claims and advertising to help protect their citizens. Trusted health or patient organizations may wish to consider becoming a trusted source of information about which software products work well.

Box 1. Brazil's Vitalk Chatbot

TNT Health in Brazil has created a virtual chatbot for patients called *Vitalk* that provides mental health screening and emotional therapy. Patients enter data through structured forms and then can ask the chatbot questions and tell it how they are feeling. Through a combination of knowledge-based and machine learning–based natural language processing, the chatbot is able to determine what the patient is saying and deliver evidence-based information on specific topics or guide the patient through simple behavioral therapy exercises. More serious concerns can be flagged so healthcare professionals are alerted or the chatbot advises the patient to see a doctor.

Other versions of this software have been used to help patients at pharmacies, municipalities, and health insurance companies to manage and triage health care questions.

More information can be found at www.vitalk.health.

Patients should also ask questions about data privacy and sharing before entering their health information into a software product. The collection of health-related data outside of clinical visits is a major potential benefit of these applications. Such data may be able to be used by a patient or their clinician to help manage

their health and wellness. However, the data are also valuable for research and development of medical products (both AI-related and not). Depending on a particular country or region's laws, data entered into a commercial website or app may not be protected the same way more traditionally collected health data is. (In the United States, for example, data collected by software programs unaffiliated with a patient's doctors or insurers is not covered by federal laws such as the Health Insurance Portability and Accountability Act of 1996 [HIPAA]). Patients should not expect the information they enter to be kept confidential or even de-identified before it is shared unless this stipulation is written in the app's terms of service. The terms of service of any trustworthy app should clearly state how the data collected will be used and by whom, exactly what data will be shared, and if and how it will be de-identified. Patients should also be given the ability to opt out.

Frontline Health Care Workers as Users

Another common use case of AI-enabled software is virtual assistant products for frontline health care workers. These tools are meant to assist with real-time triage and clinical decision-making to reduce unnecessary visits to health facilities whose staff may be overburdened and save patients from unnecessary travel. These software products can also be paired with simple hardware to allow diagnostic testing in the field. One such example is a machine learning algorithm that runs on a smartphone fixed to a microscope to quickly detect malaria parasites in a blood smear (Yang et al., 2019). These applications are meant to extend frontline health workers' ability to advise and treat patients, and the risk of using these devices depends on the functionality of the software and whether the software output is a recommendation meant to inform decision-making or make recommendations the frontline worker is not trained to do. The latter involves much higher risk. Some people may fear that this type of AI will take away jobs, but these types of systems may increase job opportunities by increasing the types of screening tests that frontline workers can perform while also increasing the likelihood that the patients referred to overburdened specialists will truly have a diagnosis that requires specialized treatment.

These software systems for frontline workers may be knowledge-based, developed with machine learning, or both. Challenges around interoperability and common data definitions are greatly lessened if frontline health care workers are trained on data entry and information is generally entered to receive AI recommendations, although bias can still be a concern. However, if the system is intended to pull in a patient's medical history, interoperability and data completeness in the patient population should be carefully considered before the system is adopted. Patient privacy should also be a concern. Patients should be told how the data being entered about them may be

used by the software company, either internally for product development and improvement or sold externally, and given the ability to opt out of such data use.

Furthermore, frontline health care workers will need to be carefully trained on the circumstances in which the software system may break down. For example, workers should be told what types of patients may not have been well represented in the training dataset or whether certain subgroups of patients showed lower predictive performance when the software was tested. If the software can diagnose multiple conditions, frontline workers should be reminded regularly of which diseases are not part of the system so they can keep those options in mind when performing differential diagnoses. Patients should also be informed when the frontline health care worker is relying on their training versus depending on AI-enabled software and, in the latter case, informed about the system's potential limitations so they can make informed decisions for themselves.

Physicians as Users

The final clinical use case for AI-enabled software is decision support tools for physicians. These tools are not meant to replace a physician but to automate simple tasks and augment decision-making. These clinical decision support tools may act to triage patients, highlight pertinent information in images or EHRs, create differential diagnosis lists, present potential treatment plans, or predict emerging health issues.

Box 2. Brazil's Portal Telemedicina System

Portal Telemedicina in Brazil has created software that directly integrates with medical devices such as ECGs and electroencephalograms (EEGs) as well as and imagers like X-rays and MRIs. This data is combined with medical reports from EHRs and manual information from field workers to a cloud-based telemedicine platform that uses AI to detect the type of information entering the system. The AI triages the patients according to their test results to prioritize emergencies in the queue and predict the diagnosis and medical finding to preformat buttons to minimize the need for typing.

This software increases access to physician and specialist care, as field health care workers can collect information and send it on to the physician. It also increases physicians' productivity by eliminating administrative tasks from the daily routine that contribute to backlogs, and less administrative work makes for more time with patients. A consistent absence of backlogs also means that additional physicians do not have be keep up with the department's workload. This system is in use in over 300 cities in Brazil and Angola across 500 clinics and hospitals, with a focus on rural and underserved areas.

More information can be found at http://portaltelemedicina.com.br/es.

These tools have the potential to make routine tasks more efficient, freeing up clinicians' time to see more patients and/or engage more deeply with patients during visits (see Box 2 for example). This software can also remind doctors of the results of new studies and changes in clinical guidelines, which can help busy doctors stay current. However, we are far from AI tools being able to replace physicians, as each of these tools is designed to perform only specific tasks rather than consider a patient holistically as clinicians do. As such, and just as with frontline workers, doctors will need to understand the intended use and limitations of the software. And again, decisions that depend on AI-enabled software should be discussed with patients so they can be informed, active participants in their own care.

All adopters of AI-enabled software systems are likely to require a lot of proof of value from these systems, but doctors may require more than most, especially from software products that perform tasks that they are used to performing and consider themselves experts in. These proofs need to include evidence that these tools will improve patient outcomes and effectively fit into the workflow of a physician's day. For example, Does the software tool allow physicians to review a diagnostic image or an EHR chart faster? Does the software give them actionable insights or just information that they need to somehow interpret themselves? In particular, physicians should honestly evaluate whether software that makes diagnostic or treatment recommendations will change their minds about their diagnoses because any software that they pay attention to only when they already agree with it does not add value. Does the software give enough information about the recommendation and/or demonstrate a high enough performance that the clinician is willing to take on the risk that they will be held liable for an incorrect recommendation? Physicians must also manage their expenditures wisely, so will a third party pay for the use of this software? If not, does the software create enough efficiencies that the doctor is willing to independently fund the use of the software? Developers will need to carefully consider these questions as they are developing and marketing their products.

Box 3. Chile's DART System

TeleDx in Chile has developed an AI-enabled software system called *DART* that uses AI to detect and prevent diabetic retinopathy, the fastest growing cause of blindness worldwide. Trained technicians take specialized digital images of the eye and input them into DART, which uses machine learning to screen for diabetic retinopathy. The software flags images with features that are potentially indicative of the disease and sends them to a remote monitoring site where specialist physicians review the images and make a diagnosis.

This system makes the screening process for diabetic retinopathy much more efficient, and it has created jobs for imaging technicians, as more patients can be screened because of the automated technology. The software has been deployed in Brazil, Chile, Colombia, Mexico, Peru, and Uruguay.

More information can be found at www.teledx.org.

>> Health Systems

Health systems, like other areas of health care, can benefit from efficiencies created through AI-enabled tools and insights. Examples of this type of software are tools used to predict likely bed utilization in hospitals and staffing needs and analyze data to detect fraud or abuse, such as flagging unusual prescribing practices. AI tools can be used to create efficiencies in administrative tasks such as using maintenance data to predict machine or equipment failures and order replacement parts, or to use natural language processing to extract data from reports to automatically fill in registry reports, potentially reducing labor costs (Hosny and Aerts, 2019).

Because these tools are working on a population level, the safety challenges with this sort of software sometimes mirror those described in the population health section above. As resources are not limitless, mistaken conclusions or flags may divert attention from where it may most be needed, and people may be able to learn to game the systems to their own benefit. Yet, these types of systems do not involve clinical decisions on the individual patient level, so the risk is lower. But because these software products can affect decision-making throughout the entire health system, careful attention to performance is still important. Adopters should carefully test the performance of a product with their system's data and in their patient population. Both the interoperability of the data being entered into the system and the potential for bias within that data need to be carefully examined and tested. Because these products are trained with historical data, historical inequities in resource allocation and health system needs can be perpetuated if developers do not involve people who understand the hospital systems and their data.

However, careful programming can do the opposite, allowing the introduction of more evidence-based decision-making.

>> Pharmaceuticals and Medtech

The final category of use cases for AI in health involves the use of AI in the research, development, testing, and surveillance of pharmaceuticals and medical devices.

AI is being used to search medical literature and clinical trial data to discover novel uses for existing drugs and quickly analyze enormous amounts of genetic and cellular data to create hypotheses on new compounds that may be able to combat diseases (Fleming, 2018). Some of the drugs in trials to treat COVID-19 were identified using machine learning techniques (Harrison, 2020). As these techniques become more common, students studying biology and chemistry will increasingly need training in these techniques to be able to use these tools and effectively understand the findings.

AI can also be used to help with more efficient clinical trials (Taylor et al., 2020). AI tools can improve patient recruitment by shuffling through health records to find trial sites that have significant numbers of patients with the correct digital phenotype (characteristics that indicate patients that will respond to the treatment). Some AI tools can then rank those trial sites as more or less likely to run high-quality trials, and others can be used to monitor patient adherence and progress through mHealth and data analytics. All tools need to be carefully validated to understand how they may affect clinical trials. **For example, many countries are concerned that clinical trials are not widely available outside of academic medical centers, and AI tools can be designed to help find sites that may not have previously been considered as a potential trial site, but tools that rely heavily on system performance in previous trials may increase this challenge rather than solve it.** Most importantly, patient selection characteristics need to be clearly understood to understand how to label a medical product that makes it through trials and is approved, but that can be a challenge if the AI tool is a black box with AI that cannot be explained.

Finally, AI can be used for medical-product surveillance. Surveillance includes more administrative tasks such as analyzing supply chain data to optimize supply chain movement (Hutchinson, 2020). But it also includes the potential to mine health records and social media to look for adverse reaction signals (Basile and Yahi, 2019). The latter is still in the early stages of development, so it is currently used only to identify potential concerns. These issues can then be examined with more traditional methods to understand if there is an actual risk to patients. These tools have the potential to identify issues with drugs faster, protecting patients by removing dangerous drugs from the market, or identifying

subpopulations more likely to have an adverse event. However, if these tools end up finding a lot of false signals, patients may be scared and stop taking their drugs unnecessarily, and regulators and manufacturers may be forced to waste time and resources showing that the signal was false. Regulators need to start considering how they will evaluate these tools and responsibly incorporate the most effective tools into their pharmacovigilance strategies.

Conclusion and Recommendations

AI-enabled health care software has the potential to significantly improve patient outcomes, relieve burdens for health care workers, and create efficiencies in health care administration and delivery. To capture this value while safeguarding patients, there are key opportunities and responsibilities for national and regional governments and organizations.

First, having a digital data infrastructure is critical. For truly efficient creation and use of AI-enabled software, data should be structured and standardized—including data definitions—and be accessible to innovators. Complete longitudinal histories can often add accuracy to these products, so national registries or other centralized databases can be valuable. At the same time, the privacy concerns around digitalizing and consolidating health information should be seriously considered, and data should be protected both legally and technically from cyberattacks. Special attention should be given to ensure that all patient populations have complete and unbiased health records. In the interim, countries should support research to understand how health data may be biased so stakeholders can take that information into account as they design, test, and evaluate AI products.

Another important requirement for safely implementing this type of innovative software is detailed clarity around applicable laws and regulations. In particular, authorities will need to clarify how they address risk when a product uses black box algorithms. **Clear frameworks of how manufacturers should categorize the risk of their products and the presence of applicable legal and regulatory authorities are crucial, especially since many product developers may not have extensive experience in health care product development.** This clarity will also help adopters of these technologies know what sort of certifications to expect and where their legal liability may be. This is particularly true for AI products that are automated rather than just informing decision-making.

Education on AI generally and on how to evaluate AI-enabled health care software and recommendations will be essential to deploying AI safely and using it to augment human intelligence.

And best practices need to be developed then systematically put into place so that a) systems are monitored both as they are implemented and over time to determine if performance in a particular real-world population meets expectations, and b) it is possible to ensure that performance does not degrade over time.

Countries around the world are grappling with these issues, and there are multiple approaches to these challenges. The European Commission has put out a series of papers on AI, emphasizing when and how AI decision-making should be trusted (European Commission, 2020). The National Academy of Medicine in the United States recently released a special report titled "Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril" (Matheny et al., 2020). China's Ministry of Science and Technology has been promoting a culture of AI innovation with an emphasis on the use of AI in health and has indicated a desire to be a leader in these technologies (Webster et al., 2017).

Every country will need to determine its own path to using AI in health care, consistent with its values, resources, and specific health care challenges. However, all countries should prioritize the development and implementation of AI tools that can be used to promote ethical, equitable, and more inclusive health care that respects and involves patients in their health and well-being. If basic research is freely shared and lessons learned along the way are broadly disseminated, advancement will happen more quickly and safety issues can be more effectively addressed. In this way, AI's potential for improving human health can be fully realized in transparent and trustworthy ways.

References

2016. "21st Century Cures Act." Pub. L. No. 114-255. https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf.

Basile, A., and A. Yahi. 2019. "Artificial Intelligence for Drug Toxicity and Safety." *Trends in Pharmacological Sciences* 40(9): 624–635. https://doi.org/10.1016/j.tips.2019.07.005. https://www.sciencedirect.com/science/article/pii/S0165614719301427.

Caruna, R., Y. Lou, J. Gehrke *et al*. 2015. "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission." En *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD '15*). Association for Computing Machinery, New York, NY, United States, 1721–1730. https://doi.org/10.1145/2783258.2788613. http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf.

CB Insights. 2020. "The State of Healthcare Report Q2'20." Page 32. New York, NY, United States: CB Insights. https://www.cbinsights.com/reports/CB-Insights_Healthcare-Report-Q2-2020.pdf.

Cho, A. 2020. "Artificial Intelligence Systems Aim to Sniff Out Signs of COVID-19 Outbreaks." *Science*, May 12. https://doi.org/10.1126/science.abc7698. https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks.

Couzin-Frankel, J. 2019. "Artificial Intelligence Could Revolutionize Medical Care. But Don't Trust It to Read Your X-Ray Just Yet." *Science*, June 19. https://doi.org/10.1126/science.aay4197. https://www.sciencemag.org/news/2019/06/artificial-intelligence-could-revolutionize-medical-care-don-t-trust-it-read-your-x-ray.

Esteva, A., B. Kuprel, R. Novoa *et al*. 2017. "Dermatologist-Level Classification of Skin Cancer with Deep Neuronal Networks." *Nature* 542:115–118. https://doi.org/10.1038/nature21056. https://www.nature.com/articles/nature21056#main.

European Commission. 2020. "Shaping Europe's Digital Future: Artificial Intelligence." Updated July 20, 2020. https://ec.europa.eu/digital-single-market/en/artificial-intelligence.

FDA (United States Food and Drug Administration). 2019a. *Developing Software Precertification Program: A Working Model*. Washington, DC, United States: FDA. https://www.fda.gov/media/119722/download.

———. 2019b. *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*. Washington, DC, United States: FDA. https://www.fda.gov/media/122535/download.

———. 2018. "FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems." FDA news release, April 11, https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye.

Fleming, N. 2018. "How Artificial Intelligence Is Changing Drug Discovery." *Nature* 557:S55–S57. https://doi.org/10.1038/d41586-018-05267-x.

FTC (United States Federal Trade Commission). 2016a. "Mobile Health App Developers: FTC Best Practices." April. https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-app-developers-ftc-best-practices.

———. 2016b. "Mobile Health Apps Interactive Tool." April. https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-apps-interactive-tool.

GHTF (Global Harmonization Task Force). 2005. *Information Document Concerning the Definition of the Term "Medical Device."* International Medical Device Regulators Forum. http://www.imdrf.org/docs/ghtf/final/sg1/technical-docs/ghtf-sg1-n29r16-2005-definition-medical-device-050520.pdf.

Harrison, C. 2020. "Drug Researchers Pursue New Lines of Attack Against COVID-19." *Nature Biotechnology* 38:659–662. https://doi.org/10.1038/d41587-020-00013-z.

Heim, L., and S. Brunsell. 2000. "Heart Disease in Women." *Primary Care* 27(vii): 741–766. https://doi.org/10.1016/S0095-4543(05)70172-X.

Hosny, A., and H. Aerts. 2019. "Artificial Intelligence for Global Health." *Science* 366:955–956. https://doi.org/10.1126/science.aay5189.

Hutchinson, G. 2020 "How Artificial Intelligence Is Improving the Supply Chain." *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2020/01/31/how-artificial-intelligence-is-improving-the-pharma-supply-chain/#60f60a303225.

IMDRF (International Medical Device Regulators Forum). 2017. *Software as a Medical Device (sAMD): Clinical Evaluation*. IMDRF. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf.

———. 2014. *Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations*. IMDRF. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf.

IDB (Inter-American Development Bank). 2020. fAIr LAC. [web page]. https://fairlac.iadb.org/es/home.

Jercich, K. 2020. "Why AI fell short in slowing the spread of COVID-19". Healthcare IT News. https://www.healthcareitnews.com/news/why-ai-failed-help-slow-spread-covid-19.

Matheny, M., S. Thadaney Israni, M. Ahmed M *et al*., eds. 2020. "Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril." Una publicación especial de la Academia Nacional de Medicina. Washington, DC, Estados Unidos: National Academy of Medicine. https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf.

Miner, A.S., L. Laranjo and A.B. Kocaballi. 2020. "Chatbots in the Fight against the COVID-19 Pandemic." *npj Digital Medicine* 3:65. https://doi.org/10.1038/s41746-020-0280-0.

Niiler, E. 2020. "An AI Epidemiologist Sent the First Warnings of the Wuhan Virus." *Wired*, January 25. https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/.

Obermeyer, Z., B. Powers, C. Vogeli *et al*. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366:447–453. https://doi.org/10.1126/science.aax2342. https://science.sciencemag.org/content/366/6464/447.

OECD (Organisation for Economic Co-operation and Development). 2019. *Artificial Intelligence in Society*. Paris, France: OECD Publishing.

Pombo, C., M. Cabrol, N. González Alarcón *et al*. 2020. "fAIr LAC: Responsible and Widespread Adoption of Artificial Intelligence in Latin America and the Caribbean." Washington, DC, United States: Inter-American Development Bank: https://publications.iadb.org/es/fair-lac-adopcion-etica-y-responsable-de-la-inteligencia-artificial-en-america-latina-y-el-caribe.

Price, W.N., S. Gerke and I.G. Cohen. 2019. "Potential Liability for Physicians Using Artificial Intelligence." *Journal of the American Medical Association* 322(18):1765–1766. https://doi.org/10.1001/jama.2019.15064. https://jamanetwork.com/journals/jama/fullarticle/2752750.

Rudin, C. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1:206–215. https://doi.org/10.1038/s42256-019-0048-x. https://www.nature.com/articles/s42256-019-0048-x.

Silcox, C., I. Sharma and A. Rai. 2020 "Trust, But Verify: Informational Challenges Surrounding AI-Enabled Clinical Decision Software," excerpted and summarized." Durham, North Carolina, United States: Duke-Margolis Center for Health Policy and The Center for Innovation Policy at Duke Law. https://healthpolicy.duke.edu/publications/trust-verify-informational-challenges-surrounding-ai-enabled-clinical-decision. Durham, NC, United States: Duke University.

Smith G and I Rustagi. 2020. "The Problem With COVID-19 Artificial Intelligence Solutions and How to Fix Them". Stanford Social Innovation Review. https://ssir.org/articles/entry/the_problem_with_covid_19_artificial_intelligence_solutions_and_how_to_fix_them#.

Taylor, K., F. Properzi, M.J. Cruz *et al*. 2020. "Intelligent Clinical Trials: Transforming though AI-Enabled Engagement." *Deloitte Insights.* London, United Kingdom: Deloitte Centre for Health Solutions. https://www2.deloitte.com/content/dam/insights/us/articles/22934_intelligent-clinical-trials/DI_Intelligent-clinical-trials.pdf.

USAID (United States Agency for International Development) and the Rockefeller Foundation. n.d. *Defining a Collective Path Forward: Artificial Intelligence in Global Health*. CII's Innovating for Impact Series. Washington, DC, United States: USAID. https://www.usaid.gov/cii/ai-in-global-health.

Webster, G., R. Creemer, P. Triolo *et al*. 2017. "Full Translation: China's 'New Generation Artificial Intelligence Development Plan.'" *Cybersecurity Initiative* (blog), New America. August 1, 2017. https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.

Whitten, A. 2020. "Is AI Ready to Help Diagnose COVID-19?". Discover Magazine. https://www.discovermagazine.com/technology/is-ai-ready-to-help-diagnose-covid-19.

Yang, F., M. Poostchi, Y. Hang *et al*. 2020. "Deep Learning for Smartphone-Based Malaria Parasite Detection in Thick Blood Smears." *IEEE Journal of Biomedical and Health Informatics* 24(5): 1427–1438. https://doi.org/10.1109/JBHI.2019.2939121.


IDB