Current State and Near-Term Priorities for AI-Enabled Diagnostic Support Software in Health Care

Duke
MARGOLIS CENTER
for **Health Policy**

Working Group

Kathleen Blake
Vice President, Healthcare Quality,
American Medical Association

Richard Frank
Chief Medical Officer,
Siemens Healthineers

James Hickman
Software Developer, Machine Learning,
Epic

Erich Huang
Co-Director,
Duke Forge

Kamal Jethwani
Senior Director,
Partners Connected Health Innovation

Shantanu Nundy*
Professorial Lecturer,
George Washington University

W. Nicholson Price
Assistant Professor of Law,
University of Michigan Law School

Arti Rai
Elvin R. Latty Professor and Co-Director,
The Center for Innovation Policy,
Duke Law

Cynthia Rudin
Associate Professor, Computer Science,
Electrical and Computer Engineering,
Duke University

Sylvia Trujillo
Senior Washington Counsel,
American Medical Association

Daniel Yang
Program Fellow,
Gordon and Betty Moore Foundation

Authors

Gregory Daniel
Deputy Director, Duke-Robert J. Margolis, MD,
Center for Health Policy
Clinical Professor, Fuqua School of Business,
Duke University

Isha Sharma
Senior Research Assistant,
Duke-Robert J. Margolis, MD, Center for Health
Policy

Christina Silcox
Managing Associate,
Duke-Robert J. Margolis, MD, Center for Health
Policy

M. Bennett Wright
Margolis Scholar,
Duke Law School

* Dr. Nundy was Director of the nonprofit side of Human Diagnosis Project through December 2018.

Acknowledgements

About the Duke-Margolis Center for Health Policy

The Robert J. Margolis, MD, Center for Health Policy at Duke University is directed by Mark McClellan, MD, PhD, and brings together expertise from the policy community in Washington, DC, Duke University, and Duke Health to address the most pressing issues in health policy.

The mission of Duke University's Robert J. Margolis, MD, Center for Health Policy is to improve health and the value of health care through practical innovative, and evidence-based policy solutions.

Duke-Margolis catalyzes Duke University's leading capabilities including interdisciplinary academic research and capacity for education and engagement, to inform policymaking and implementation for better health and health care.

For more information, visit healthpolicy.duke.edu.

Table of Contents

Executive Summary

Artificial intelligence (AI) refers to the ability of a machine to perform a task that is normally done by humans, including problem-solving and learning. AI systems and applications are ubiquitous in human life. They recommend movies, give directions on the fastest routes to travel, detect credit card fraud, fulfill online purchases, and translate webpages into different languages. In the near future, it is likely that autonomous cars using certain AI methods and systems will become a significant form of transportation.

Given the proven effectiveness of AI-enabled software in other domains, AI has been introduced into many areas of medicine. One particular application is clinical decision support (CDS) software. This project examines the potential benefits and challenges when AI is incorporated into CDS software, particularly software that supports improved clinical diagnosis, as well as barriers that may be preventing development and adoption of this software.

Improved CDS could be useful in reducing diagnostic errors, which account for almost 60 percent of all medical errors and an estimated 40,000 to 80,000 deaths each year.[1] The National Academy of Medicine estimates that "nearly every American will experience a diagnostic error in their lifetime, sometimes with devastating consequences."[2] AI-enabled diagnostic support software (DxSS)—a subset of CDS software—has the potential to augment clinicians' intelligence, support their decision-making processes, help them arrive at the correct diagnosis faster, reduce unnecessary testing and treatments otherwise resulting from misdiagnosis, and reduce pain and suffering by starting treatments earlier.

CDS software, including AI-enabled software, also can assist clinicians with knowledge management. The volume of information and knowledge needed to practice medicine effectively is increasing enormously every year.[3] The volume of new information generated daily about wellness, disease, treatments, and prevention, greatly exceeds the ability of clinicians to absorb and effectively process it. In a world where clinicians are constantly being pushed to see more patients in shorter amounts of time, CDS software holds significant promise to create relevant and meaningful information, and deliver that information to the right person at the right point within the clinical workflow, at the right location, in a format that supports shared decision-making by providers and patients.[4]

To foster innovation and incentivize adoption of safe and effective AI-enabled DxSS, the Duke-Margolis Center for Health Policy, with a working group of experts throughout the health care and artificial intelligence ecosystem, has developed an overview of AI-enabled CDS software and the regulatory and policy environment surrounding use of CDS.

This paper discusses how AI may improve CDS software, and it shows current examples across a variety of conditions. There are diverse types of CDS software with AI, and those types (rules-based vs data-based, continuously learning vs locked) affect how the applications may be used and regulated. The paper further outlines the legal and regulatory landscape for AI-enabled CDS software, which is largely affected by the U.S. Food and Drug Administration's (FDA's) medical device regulations. Due to recent legislation, some types of CDS are not considered medical devices (and not subject to FDA regulation), although other types of AI-enabled CDS software remain subject to regulation by the agency. The most impactful regulatory update affecting AI-enabled CDS software is the FDA's proposed precertification program for software that are regulated as medical devices.

The Center also explored potential benefits and considerations for using AI in clinical decision-making and diagnostic support and identify the key issues that may be delaying innovation and adoption of AI-enabled DxSS.

This work surfaced priority concerns around the development, regulation, and adoption of safe and effective DxSS that stakeholders will need to address, including:

- **Evidentiary needs for increased adoption of these technologies**. This evidence will include the effect of the software on patient outcomes, care quality, total costs of care, and workflow; the usability of the software and its effectiveness at delivering the right information in a way that clinicians find useful and trustworthy; and the potential for reimbursement for use of these products by payers
- **Effective patient risk assessment of these products**. The degree to which a software product comes with information that explains how it works and the types of populations used to train the software will have significant impact on regulators' and clinicians' assessment of the risk to patients when clinicians use this software. Product labeling may need to be reconsidered and the risks and benefits of continuous learning versus locked models must be discussed.
- **Ensuring AI systems are ethically trained and flexible**. Best practices to mitigate bias that may be introduced by the training data used to develop software are critical to ensuring that software developed with data-driven AI methods do not perpetuate or exacerbate existing clinical biases. In addition, developers will need to assess how the data inputs required by their software may affect scalability of their products to settings that are different from the original setting that provided the data used to train the algorithms. Finally, best practices and, potentially, new paradigms are needed for how to best protect patient privacy.

We intend for this report to serve as a resource for developers, regulators, clinicians, policy makers, and other stakeholders as they strive to effectively, ethically, and safely incorporate AI as a fundamental component in diagnostic error prevention and other types of CDS.

Section I: Introduction to Artificial Intelligence, Clinical Decision Support Software, & Diagnostic Support Software

Clinical Decision Support Software & Diagnostic Support Software

According to the Office of the National Coordinator (ONC) for Health Information Technology, clinical decision support (CDS) software should provide "clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care."[5] CDS software tools support health care providers in diagnosis, treatment decisions, and population health management.[*]

At its best, CDS software supports clinicians in providing better, more personalized care with improved value. Optimally, use of CDS software results in more timely and accurate diagnosis; identification of and/or personalized presentation of evidence-based treatment choices; or efficient performance of low-value routine tasks that free clinicians to focus on more complex activities. Conceived in this way, CDS software provides cognitive support for physicians and other clinicians instead of attempting to replace them. It may also enable primary care physicians to be able to diagnose and/or treat patients in situations that may have otherwise required referral to specialists.[6,7]

Software that provides support and/or recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition by displaying, analyzing, or printing medical information about a patient or other medical information

Software that supports a clinician by analyzing medical and/or other data to make recommendations of potential diagnoses

Figure 1. How Artificial Intelligence (AI)-Enabled Diagnostic Support Software (DxSS) Fits into the CDS Ecosystem.

---

[*] CDS software also includes administrative support software (e.g., software that predicts length of stay to assist provider systems when scheduling surgeries). However, since the 21st Century Cures Act removed administrative software from the definition of a "medical device" regardless of the explainability of its recommendations, we will consider that type of software outside of the scope of this paper. For the same reason, we will also not consider software that simply displays input data in innovative ways without making recommendations, even if the purpose is to facilitate diagnosis. See Appendix B for more information on what types of software are no longer classified as a medical device due to the 21st Century Cures Act.

Historically, the rationale for clinical recommendations from CDS software came from U.S. Food and Drug Administration (FDA) labeling, clinical guidelines, and other citable authorities. Unidirectional informational websites such as the original version of UpToDate™ or athenahealth's Epocrates® provide evidence-based CDS at the point-of-care using thousands of clinician, editor, and peer-reviewed responses to queries supported by extensive citations.[8,9] The types of CDS software that cause medication alerts within electronic health records (EHRs) are often based on a cross comparison of a patient's medication list with known adverse interactions listed in medication labels. More recently however, CDS software has begun to incorporate more complex algorithms, including artificial intelligence (AI), to increase the accuracy and relevance of the results.

Diagnostic support software (DxSS) is a subset of CDS software that is specifically designed to support a clinician in diagnosis. Most DxSS on the market today is used in conjunction with medical imaging. Examples include classifying tumors on x-ray and quantifying coronary stenosis from coronary angiograms.[10,11] Figure 1 shows how DxSS is a subset of CDS software, and AI-enabled DxSS is a subset of DxSS. Examples of AI-enabled CDS software and DxSS are detailed below.

Examples of AI-Enabled CDS Software

The following are several examples of therapeutic areas where AI-enabled CDS software is already in use, either as FDA-cleared commercial products or systems being developed within provider systems. These examples include both CDS software and DxSS, and use input data from a range of sources, including medical imaging, medical sensors, data from EHRs, etc.

Diabetic Retinopathy

If untreated, diabetic retinopathy can lead to visual impairment or blindness, affecting about 2.6 million people in the world.[12] Timely referral to a specialist can reduce the chance of complications; however early screening and diagnosis can be difficult due to the lack of early symptoms before vision is affected. IDx-DR screens for diabetic retinopathy without the need for clinician interpretation (although on-label use requires the test to be performed by health care providers on specialized equipment).[13,14] The IDx-DR software analyzes images of the eye taken by a retinal camera and provides clinicians with a recommendation that signs of diabetic retinopathy were detected and the patient should be referred to a specialist, a recommendation that no signs of diabetic retinopathy were detected and the patient should be rescreened in 12 months, or an indication that the input image was not of high enough quality to produce a recommendation.[15]

The company submitted clinical data to FDA that consisted of retinal images obtained from 900 patients with diabetes at 10 primary care sites. These data showed that IDx-DR was able to correctly identify the presence of more than mild diabetic retinopathy with 87 percent sensitivity and was able to correctly identify those that did not have more than mild diabetic retinopathy with 90 percent specificity.[16]

Stroke

In the United States alone, someone has a stroke every 40 seconds and dies every 4 minutes.[17] Treatment is time-dependent; therefore early detection of stroke and referral to a specialist can be critical.[18] Two companies have recently received FDA clearance for AI-enabled CDS software in this space. The Viz.ai ContaCT is a standalone CDS tool developed with AI methods and designed to analyze computed tomography (CT) results to aid providers in identifying the most appropriate treatment plan for a patient.[19] The MaxQ-AI platform, Accipiolx, is an AI-enabled software workflow tool that aids physicians with identifying acute intracranial hemorrhage and prioritizing the treatment of strokes or head trauma.[20]

The Viz.ai ContaCT application uses AI to analyze CT images of the brain and send a notification to a "qualified specialist" if it detects a high likelihood that a patient had a stroke.[21] As part of its application to FDA, Viz.ai submitted the results of a retrospective study involving 300 CT scans comparing the performance of the Viz.ai ContaCT application against that of a specialist. The results showed that the software was able to triage patients faster than the clinician by accurately identifying images which contained findings indicative of stroke in more than 95 percent of cases, with a sensitivity and specificity of about 88 and 90 percent, respectively.[22,23] The notification function, which alerts "qualified specialists" to triage patients to expedited care, saved an average of 52 minutes per case.[24] In April 2018, FDA cleared a second product from Viz.ai called Viz CTP.[25] Viz CTP uses AI-enabled image analysis to generate and analyze functional and dynamic CT perfusion images to aid with earlier identification and diagnosis of stroke to improve patient outcomes.

The MaxQ-AI Accipiolx software uses AI to identify findings of acute intracranial hemorrhage from non-contrast CT scans.[26] The device is composed of two main components, the Accipiolx Agent and the MaxQ-AI Engine, which analyze head CT scans, identify suspicious findings, and relay them to a clinician's workstation labeled with recommendations regarding the presence or absence of a suspected hemorrhage.[27] As part of its submission to FDA, MaxQ-AI conducted a retrospective study of 360 CT cases collected from more than 30 sites within the United States to demonstrate a sensitivity of 92 percent and specificity of 86 percent, with an average per-case processing time of 4.1 minutes.[28]

Distal Radial Fractures

One of the most commonly injured parts of the body are the hand and wrist, with one out of six fractures presenting to emergency rooms as distal radial fractures.[29] Today, in cases where a distal radial fracture is suspected, standard clinical practice is to cast the wrist and perform a follow-up x-ray or MRI 10 to 14 days after to determine the final diagnosis.[30] Imagen's OsteoDetect is a computer-aided detection and diagnostic software that uses AI to analyze two-dimensional X-ray images at the original point-of-care for signs of a distal radius fracture in adult patients, marking the location of the fracture to aid the clinician, which potentially results in earlier diagnosis while reducing unnecessary treatment and follow-up imaging.[31]

A retrospective study of 1,000 images assessed the independent performance of the algorithm and the accuracy of the fracture localization of OsteoDetect against the performance of three board-certified orthopedic hand surgeons. A retrospective study of 24 clinicians who reviewed 200 patient cases was submitted as well, with both studies indicating that clinical diagnostic performance (sensitivity, specificity, positive predictive value, and negative predictive value) was improved while using OsteoDetect compared to unaided performance conducted according to standard clinical care. Specifically, the studies demonstrated a software-aided sensitivity of 80.3 percent and specificity of 91 percent, compared to the non-aided sensitivity of 74.7 percent and specificity of 89 percent.[32]

Sepsis[*]

Sepsis is one of the most expensive conditions treated in hospitals, accounting for approximately five percent of total hospitalization costs and an overall annual cost of $20.3 billion USD in the United States.[33,34] Early detection of sepsis via automated systems and more timely intervention may reduce treatment costs and overall resource use.[35,36]

---

[*] Adapted from Xavier Health's August 2018 white paper on "Perspectives and Good Practices for AI and Continuously Learning Systems in Healthcare."

Automated detection systems offer the possibility of monitoring patients in "real-time," and can alert the relevant physicians or nurses (e.g., by email or pager) to the need for timely clinical evaluation and potential need to initiate treatment. The system is able to process a vast amount of clinical data, including real-time information, in order to identify the onset of sepsis earlier (DxSS) or produce sepsis risk scores at time of admission (CDS software).[37] Duke and a number of other health systems are developing such AI-enabled systems.[38]

Appendicitis

Appendicitis is one of the most common causes of abdominal pain, though optimal diagnosis remains a challenge.[39] Relevant information is often entered into the EHR as free text notes by medical personnel, making it difficult for an accurate diagnosis to be made in the limited amount of time available to the clinician.

The National University Health System (NUHS) in Singapore has been working to bring AI into the clinical workflow. One of the projects undergoing clinical trials is a scalable, free-text based, automated tool to diagnose appendicitis.[40,41] CT utilization rates for the diagnosis of appendicitis have been increasing without a relative improvement in outcomes.[42] Therefore, the main objective of the appendicitis diagnosis system is to guide the use of CT scans only in uncertain cases of appendicitis. The model was trained with a dataset of over 200,000 cases of patients presenting with abdominal pain over 10 years, and data continue to be collected to periodically update the software.[43]
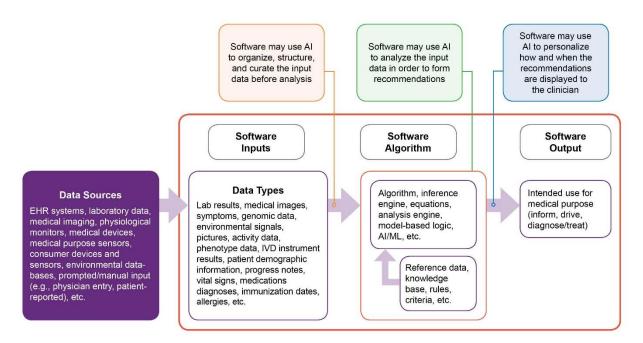
Primary Care Diagnostic Support

Early research in collaboration with Harvard Medical School suggests that when combining multiple physicians' opinions together, diagnostic accuracy increases to 86 percent as opposed to the diagnostic accuracy of an individual doctor at 63 percent.[44] An e-consult system could allow generalist physicians working with underserved patients the ability to efficiently collaborate with specialists electronically about a case.[45]

The Human Diagnosis Project, or Human Dx, is building an AI platform that helps identify the relevant expertise and combine physician insights.[46] Rather than the AI directly providing recommendations on diagnosis, Human Dx uses collective intelligence, which combines the human intelligence of multiple physicians with machine learning. Researchers at the University of California, San Francisco have now implemented the platform in a primary care setting where physicians choose relevant cases where they would like additional diagnostic support and manually enter information into a case report form, which is then reviewed by five or more physicians who independently attempt to diagnose the case. These independent assessments are combined into a single output with the aid of machine learning, and are sent to the attending physician to incorporate into their diagnostic process. Preliminary evidence shows that this process is substantially increasing physician confidence in diagnosis, results of which are currently being validated in a clinical trial.[47]

AI in CDS Software & DxSS

Conceptually, there are two types of AI: narrow and general. Narrow AI refers to AI designed to address a specific application area (e.g., strategic games, autonomous vehicles).[48] In contrast, general AI is a theoretical system that would be able to fully mimic human reasoning or intelligence in any context and capable of conscious, abstract thought.

AI-enabled CDS software and DxSS are narrow AI applications designed to use various types of input and analyze it in order to make recommendations to support decision-making. Figure 2 shows the process in which software can be used to produce these recommendations, as well as the different points at which AI may be used to make software more effective. The first area AI may be used is automated curation of software inputs in order to structure datasets that can be analyzed more easily (Figure 2 includes examples of data sources and data types that might be used in CDS software). The analysis function of the software may also use AI, by using a continuously learning AI algorithm or a model that is "locked" (although, like all software, it may be updated periodically). The next section explains continuously learning vs locked algorithms in more detail. Finally, AI-enabled software may have the potential to individualize recommendations and provide the most useful information for a specific health care

Figure 2. Stages of AI-Enabled Software. The diagram starts with examples of data sources and data types that may be used as inputs for the software as well as examples of areas which may be utilizing either AI directly or functions developed with AI. The figure is modified from Figure 1 from FDA's June 2018 Pre-Cert Update. EHR, electronic health record; IVD, in vitro diagnostic medical device.

provider at a specific time. This ability could enhance integration of the software into the provider system's workflow, similarly to how a smartphone uses AI to personalize the predictive text suggestions over time by learning which language the user regularly selects and which words are less relevant.

Software may use AI for any or all of these stages. For example, in a study published in May 2018 and led by Rajkomar and Oren from Google, AI was used to transform raw EHR files including clinical notes into a structured and standardized format. AI deep learning methods were then applied to the structured data to develop models to predict risk of mortality, hospital readmission, prolonged hospital stay, and discharge diagnosis.[49]

Types of AI

AI can also be divided into two categories depending on how it is programmed.[50] Rules-based AI uses previously validated information (e.g., clinical guidelines, risk calculators, published studies) to set up a

flow chart of individual decisions that lead to a recommendation.[*] Data-based AI, often referred to as machine learning, applies methods to data to learn without being explicitly programmed. There are multiple types of machine learning algorithms, such as linear and logistic regression, decision trees, support vector machines, and deep learning (also known as neural networks). Different types of algorithms are more suitable for different types of problems, similar to how certain statistical methods are more appropriate for certain types of analyses. More information on AI algorithms can be found in Appendix D. In addition, the choice of machine learning algorithm affects how likely the resulting software is to be either interpretable or explainable, which we discuss in detail in Section III.

Rules-Based:
Uses clinically accepted rules to guide decision-making (using clinical guidelines, FDA labels, published literature, etc.)

Artificial Intelligence (AI)

Data-Based:
Uses data to learn without being explicitly programmed Also referred to as machine learning, and includes methods such as deep learning, logistical regression, random forest

Locked Model:
While functions within the software are developed with data-based AI techniques, the software no longer changes with each use

Continuously Learning Algorithms:
The software automatically changes with each use

Figure 3. Types of AI Algorithms. AI includes both rules-based and data-based algorithms. Data-based algorithms can be locked or continuously learning. FDA, U.S. Food and Drug Administration.

Continuously Learning Algorithms vs Locked Models

Once AI software that was developed through machine learning is in use, the model can continue to learn or be "fine-tuned." This additional "learning" requires additional labeled examples, meaning both the real-world input data it may be using to make recommendations but also information on whether

---

[*] While we have chosen to include rules-based algorithms in our definition of AI, some definitions only include data-based AI.

those recommendations were considered "correct."[*] These changes can be made automatically each time a new labeled example is received or the labeled examples can been stored for periodic updating of the software. How this type of ongoing "learning" will be regulated in regards to methodological standards, ongoing verification and validation, product labeling, etc. is an ongoing discussion, which will be explored more in Section III.

As the software receives feedback, changes can be made automatically to the fitting function, updating the model in real time. This process is called continuous learning. There are both benefits and risks to this approach. DxSS that incorporates continuous learning could more precisely calibrate suggestions to specific demographic or geographic areas over time, taking into account that certain diagnoses are more common in that setting and/or adjusting for local norms in the input data formatting or presentation.[†] However, as software changes, the rate and distribution of false-positives and false-negatives may also change, potentially in ways that no longer have an acceptable benefit-risk profile. As such, there are serious concerns about the risks and ethics of deploying a continuously learning software system in the clinical setting.

Alternatively, software manufacturers can use AI learning methods on a discrete set of training data to develop a model. Once the model is developed, it can be "locked" and used without automatic updates.[‡] Software that uses locked models can still gather the same type of feedback used in continuous learning in parallel with use of the locked model, and then use those data to update the software on a regular (not real-time) basis while allowing for additional testing to understand the impact of any changes made. This is also referred to as discontinuous learning. However, if these regular updates are not done, locked models have the potential to degrade over time if inputs change significantly. All of the examples discussed earlier use locked systems that are periodically updated.

---

[*] Some types of AI do not require labeled examples to learn, but we have limited the scope of this paper to DxSS developed with "supervised learning" techniques, as these techniques are the most likely to be used by developers in the near-term.

[†] This is a particular concern with EHR data inputs, which can change substantially over time. The criteria for diagnoses can change, new treatments can be introduced, the precision or completeness of data may change due to automatic pulls of data from medical devices or standard of care changes – all of these changes may affect the performance of a locked algorithm, which depends on that data. For example, if an algorithm uses the presence of particular prescription drugs as a contributor to its diagnosis recommendation (possibly because of known side effects), a new drug with the same potential side effect would not be accounted for in the locked algorithm.

[‡] Note that algorithms are also generally "locked" when being evaluated with the testing sets that gauge their sensitivity and specificity.

Section II: Current Legal & Regulatory Environment for AI-Enabled CDS Software & DxSS

FDA Oversight of Software as a Medical Device

Since 1976, FDA has regulated medical devices, defined as instruments used in the diagnosis, cure, mitigation, treatment, or prevention of disease, or affect the structure or function of the body. This includes types of software, termed "Software as a Medical Device" or SaMD, that are "intended to be used for one or more medical purposes and to perform these purposes without being part of the hardware of the medical device."[51] SaMD can be used across a broad range of technology platforms such as medical device platforms, general computing platforms, virtual networks, and so forth.[52] It differs from Software in a Medical Device (SiMD), in which the software is integral to the hardware of a medical device.[53] Leveraging the International Medical Device Regulators Forum (IMDRF)-supported guidance, FDA has provided guidance for developers on the approaches FDA will use to evaluate the safety and efficacy of CDS software and DxSS if the software falls into the definition of a medical device.

When Is CDS Software a Medical Device?

In December 2016, Congress enacted the 21st Century Cures Act removing certain types of software from FDA authority, including certain types of CDS software. The relevant provisions of the 21st Century Cures Act state that CDS software is not considered a medical device (and is therefore not subject to FDA regulation) as long as the software does not "acquire, process, or analyze a medical image or signal from an in vitro diagnostic device, or pattern or signal from a signal acquisition system" and meets one of the following two conditions:

1. It simply displays, analyzes, or prints medical information about a patient or other medical information such as peer-reviewed clinical studies and clinical practice guidelines; or
2. It supports or provides "recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition" while "enabling such health care professional to independently review the basis for such recommendations that such software presents so that it is not the intent that such health care professional rely primarily on any of such recommendations to make a clinical diagnosis or treatment decision regarding an individual patient."[54]

Since 2017, FDA Commissioner Scott Gottlieb has publicly specified agency efforts intended to reduce regulatory ambiguity regarding SaMD and CDS software, releasing a series of guidance documents in December 2017. Notably, the draft guidance, "Clinical and Patient Decision Support Software," delineated the types of CDS software that would be excluded from the classification of a medical device based on the agency's reading of the 21st Century Cures Act.[55] Section 3060 of the 21st Century Cures Act gives FDA the authority to bring some of the excluded software back under its jurisdiction if it finds significant likelihood and severity of patient harm if the software does not perform as intended or if the software is likely to supplant vs support the clinical judgement of the user based on the opportunity to review the basis of the decision and the intended user and use environment.

For AI-enabled CDS software, the provision that the software must enable the "independent review of the basis for recommendations" by a health care professional is a critical consideration.[56] This language

suggests that 'black box'[*] algorithms are not exempt from FDA oversight under the 21st Century Cures Act, although FDA may still decide to exercise enforcement discretion in low-risk contexts. For AI-enabled software that does have some explanation, the situation is less clear. Section III explores this in more detail.

How Does FDA Currently Regulate Medical Devices?

If a specific piece of CDS software meets the definition of a medical device, it is currently regulated like any other medical device. The Medical Device Amendments of 1976 to the United States Federal Food, Drug, and Cosmetic Act (FD&C Act) established a classification system with three distinct classes of devices (Class I, Class II, and Class III) categorized by risk and associated regulatory controls (general and special) and requirements. Class I and Class II devices are "cleared" by FDA through the 510(k), also known as pre-market notification (PMN), pathway rather than "approved" through the more stringent pre-market approval (PMA) pathway required of high-risk Class III devices. Novel devices are either classified as Class III by default or go through the De Novo pathway described in the section below. Appendix C goes through each of these pathways in more detail. Alternatively, FDA also has the authority to use enforcement discretion to not regulate medical devices.[†]

Table 1. FDA Regulatory Pathways, Processes, and Classification System for Medical Devices.

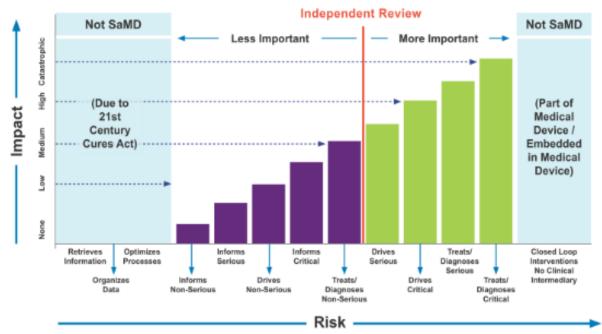| Regulatory Pathway/ Process	Class of Medical Device	Level of Risk	Regulatory Controls	Example
PMN/ 510(k)	**Class I**	Low	General or "Exempt" Status[‡]	Software that displays longitudinal readings from a continuous glucose monitor
	Class II	Moderate to High	General and Special	Software that processes and analyzes radiological imaging to identify the presence or lack of a disease/condition
De Novo	**Classified as Class I or Class II during process**	Low to Moderate	General and Special	Screening devices that analyze images and recommend or directly refer patients to specialists
PMA	**Class III**	High	General	While there are examples of Class III SiMDs, there are no examples of Class III SaMDs known to the authors
Unregulated	Class of Medical Device	Level of Risk	Regulatory Controls	Example
Enforcement Discretion	**N/A**	Low	None	Software that provides information to a patient about the use of a prescription drug that is consistent with FDA-required labeling

---

[*] The term 'black box' refers to software that does not explain how the input data are analyzed in order to come to a recommendation. This can be because the algorithm/model is too complex to be understood by humans (see Section III) or because the functionality is considered proprietary.

[†] For example, FDA stated that it "intends to adopt an enforcement discretion policy for [patient decision software] that generally parallels the CDS for health care professionals excluded from the device definition under section 520(o)(1)(E) of the FD&C Act."[57]

[‡] "Exempt" status means that there is no need for proof of safety or efficacy, or clinical trials. Therefore, these devices do not need to undergo the standard PMN/510(k) process.

Table 1 lists the different classes of devices with their associated risks and regulatory pathway as well as an example of SaMD in each category.

Characterizing risk with regards to SaMD poses unique challenges for regulators. The IMDRF framework for SaMD Risk Categorization, which was adopted by FDA in a 2017 guidance, identified rapid development cycles, frequent changes to software, and updates delivered by mass distribution that are

Figure 4. Risk-Based Approach to Importance of Independent Review. Illustrates where independent review is more or less important. Based on Figure 13 from FDA's December 2017 guidance on SaMD: Clinical Evaluation.

often left to the user to install as specific lifecycle-related challenges that could contribute to the assessment of the risk of a SaMD product.[58]

The risk categorization framework recommends regulators use two main factors to characterize SaMD:

1. The significance of information provided by the SaMD to the health care decision
   a. Does the information provided by the SaMD cause immediate or near-term action? For diagnosis-related software, does the SaMD diagnose/screen/detect a condition?
   b. Does the information provided by the SaMD drive clinical management? For diagnosis-related software, does the SaMD analyze relevant information to aid in making a definitive diagnosis?
   c. Does the information provided by the SaMD inform clinical management? For diagnosis-related software, does the SaMD inform of options for diagnosing a disease/condition and/or aggregate relevant information?
2. The state of the health care situation or condition
   a. Critical (i.e., accurate and/or timely diagnosis is vital to avoid death, long-term disability or serious deterioration of health)
   b. Serious (i.e., accurate diagnosis is vitally important to avoid unnecessary interventions or timely interventions are important to mitigate long-term irreversible consequences to health)

c. Non-serious (i.e., accurate diagnosis is important but not critical to mitigate long-term irreversible consequences)

As Figure 4 shows, the categorization of risk occurs across a continuum where the lowest-risk software is more descriptive and used for general population health interventions (e.g., wellness apps), and the highest-risk software is prescriptive with the SaMD/SiMD itself as the sole service/intervention.

For DxSS, which we have defined as software that augments the provider's best clinical judgement by making recommendations after an analysis of input data, risk is likely to be in the low-to-mid-range because the provider maintains control over the final diagnosis made.[*] However, depending on the health care situation or condition, FDA may judge the risk higher or lower. Risk assessments of AI-enabled software may also depend on whether the software uses a locked model or a continuously learning algorithm, as well as how understandable the basis of the software's recommendations is to the user. This is explored more in Section III.

Table 2 lists recent FDA clearances of AI-enabled SaMD. It is notable that all have been classified as Class II devices either though the 510(k) or De Novo pathway,[†] and all utilize locked models that will be updated periodically.

Table 2. Recent FDA Decisions on AI-Enabled Decision Support Software.

Device Name	Decision Date	Regulation Name	Regulatory Pathway	Regulatory Class	Product Code
Arterys Cardio DL[59]	January 2017	Picture archiving and communications system	510(k)	Class II with General Controls	LLZ
QuantX[60]	July 2017	Radiological computer-assisted diagnostic (CADx) software for lesions suspicious for cancer	De Novo	Class II with Special Controls	POK
Viz.ai ContaCT[61]	February 2018	Radiological computer-aided triage and notification system	De Novo	Class II with Special Controls	QAS
Viz CTP[62]	April 2018	Picture archiving and communications system	510(k)	Class II with Special Controls	LLZ
IDx-DR[63]	April 2018	Retinal diagnostic software device	De Novo	Class II with Special Controls	PIB
OsteoDetect[64]	May 2018	Radiological computer-assisted detection and diagnosis software	De Novo	Class II with Special Controls	QBS
ECG App[65]	September 2018	Electrocardiograph software for over-the-counter use[‡]	De Novo	Class II with Special Controls	QDA
Accipiolx[66]	October 2018	Radiological computer-aided triage and notification software	510(k)	Class II with General Controls	QAS

---

[*] Even for a software such as IDx-DR which does not require a clinician's interpretation (and is therefore not technically DxSS by this paper's definition), the result is a referral to a specialist for final diagnosis and treatment decision-making, although the primary care clinician may base care decisions from the software results.

[†] Appendix C gives more details on predicate devices and the 510(k) and De Novo pathways.

[‡] Since this is for over-the-counter use, it is patient decision support (PDS). PDS software is similar to CDS software, but is designed to support the layperson in the decision about whether to seek medical advice and/or in following their wellness or treatment regimens.

After a medical device enters the market, companies are responsible for any post-market studies or surveillance required by FDA. In addition, FDA requires manufacturers to submit reports of adverse events in order to monitor compliance to regulatory standards and track whether a malfunction in an approved device is the cause of a suboptimal health outcome.

The Software Precertification Program: How FDA Is Considering Regulation of SaMD

In July of 2017, FDA released its Digital Health Innovation Action Plan, announcing that it was reimagining the approach to digital health medical devices.[67] This was done out of concern that the traditional device approval/clearance pathway is not suited to the faster, more iterative design and development cycles of software, which lasts weeks to months rather than the months to years cycle of more traditional medical products. FDA started the Software Pre-Cert Pilot Program to explore the idea of "precertifying" SaMD developers as reliable manufacturers of high-quality, safe, and effective digital health devices while providing appropriate patient safeguards in order to enable innovation in health care and accelerate time to market.

Since then, FDA has continued to release more details about their thinking about this initiative.[68] It envisions this as a voluntary alternative to the traditional PMA/510(k) pathway, thereby allowing the Center for Devices and Radiological Health (CDRH) to "assess the safety and effectiveness of software technologies without inhibiting patient access to these technologies."[69] The precertification program's goal is to provide more streamlined and efficient oversight over manufacturers that demonstrate a "robust culture of quality and organizational excellence (CQOE)" and will commit to monitor real-world performance of their products.[70] A key component of the CQOE will be to incentivize companies to transparently verify product performance (safety, effectiveness, and performance in the real world) across the total product life cycle (TPLC). The latest update from January 2019 continues to expand on how this will extend beyond traditional Quality Management Systems, incorporating clinical responsibility, cybersecurity practices and real-world performance analysis.[71] The update also recognized the need that users of SaMD have for transparency on the extent of pre-market review (if any) and post-market monitoring of the SaMD product they are using.

The current vision for the program consists of four interdependent components:

1. **Excellence Appraisal and Precertification:** FDA (or potentially a third-party auditor) would precertify a company or business unit within a company as demonstrating pre-defined "excellence principles." FDA envisions two levels of precertification, with companies or business units with a limited track record of products in the health care space being designated as level 1 and companies with a proven track record of successfully marketing and maintaining health care products obtaining level 2 status.

2. **Review Pathway Determination:** FDA has proposed to leverage the IMDRF framework for risk categorization of SaMD (described above), which could be modified by the core functionality of the SaMD (including functionality critical to maintaining the stated performance and safety profile). The precertification level of the company and the risk subtype would form the basis for the level of pre-market review (if any) that would be required, although key technological attributes and other considerations could also affect the pathway determination.

3. **Streamlined Pre-Market Review Process**: For those products that would require review, FDA envisions precertified entities being able to opt-in to an alternative process that allows FDA to interactively work with the company to use the information received during the excellence appraisal and pathway determination to streamline review of product-specific regulatory information. The goal is to facilitate iterative early engagement during development in order to

shorten the time between when the device is finalized and when the device is authorized by FDA to enter the marketplace.

4. **Real-World Performance:** FDA will identify real-world performance analytics (a range of acceptable data domains and analytic methodologies) it would expect precertified companies to perform after the product enters the marketplace, throughout the rest of the product lifecycle.

Regulation of Diagnostic Test Kits Developed & Deployed within Health Systems

Laboratory-developed testing services and procedures (LDTs) are designed, developed, validated, and provided by clinicians in a health system provided directly to their patients.[72] Similarly, in-house-developed CDS software (whether AI-enabled or not) could be considered a programmed process based on that system's unique knowledge of protocols, performance characteristics, and means of analysis that provides clinical information to clinicians and their patients. Assuming these in-house programs are not distributed commercially, they could be considered the practice of medicine using similar arguments to those used around traditional LDTs. [73]

FDA regulates commercial clinical test kits that are mass produced and distributed to many different laboratories beyond the manufacturers control through the 510(k), PMA, and De Novo pathways, like other medical devices.[74] In contrast, LDT procedures are subject to the Clinical Laboratory Improvement Amendments (CLIA), overseen by the Centers for Medicare & Medicaid Services (CMS).[*] A similar differentiated oversight approach for in-house-developed CDS software may be sought by health care delivery systems for their patients' clinical care to ensure appropriate oversight and standards if it is determined that these systems do not constitute medical devices.

Current Data Privacy Considerations around AI-Enabled Software

Developers of AI-enabled CDS software, including DxSS, need to be aware of the various federal and state laws that protect health information privacy as they look to acquire potentially large quantities of patient data for use with machine learning algorithms and understand how patient data may need to be protected as the AI-enabled software is processing it for decision support.

Health Insurance Portability and Accountability Act

The U.S. Department of Health and Human Services' (HHS) Office for Civil Rights (OCR) is the primary enforcement agency for the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which established national standards regarding the protection of patient health information. HIPAA is technology agnostic and is able to accommodate existing or innovative health technology as long as the compliance requirements are met. Detailed information, guidance, and toolkits developed by ONC and OCR and other HHS agencies for consumers, providers, and developers to understand and comply with HIPAA requirements can be found on the ONC website.[75]

The HIPAA Privacy Rule aims to safeguard all protected health information (PHI), or all "individually identifiable health information" from being used and disclosed without the patient's consent. [76] This rule applies to "covered entities," which are health plans, health care clearinghouses, and any health care provider who transmits health information in electronic form in connection with transactions, as

---

[*] As LDTs have become more common, FDA has voiced concerns about the potential for more oversight. In 2017, FDA issued a discussion paper expanding upon potential LDT policies such as a risk-based approach to oversight, independent pre-market review for tests, risk classification activities, third party review, adverse event reporting, exemptions, public availability of performance information, and so forth.[77]

well as their business associates.[*] Development and testing of SaMD using data from patients that have consented to its use is permitted, and HIPAA requires that patients be given the right to direct any covered entity to transmit a copy of their medical records to a designated person or entity of the individual's choice.[78]

There are exceptions to the requirement of patient consent under HIPAA, however. Provider systems can use PHI for health care operations, including quality assessment and improvement activities, developing clinical guidelines, and conducting population-based activities relating to improving or reducing costs.[79] The internal development and use of AI-enabled software could be used to facilitate these functions. HIPAA also allows sharing of de-identified data (data in which specified individual identifiers are removed such that the information left neither "identifies nor provides a reasonable basis to identify an individual").[80] De-identified data, however, could potentially sacrifice some of the AI's effectiveness by making it difficult for the program to match data from different databases,[81] as well as potentially removing important demographic or geographic information that might be relevant to effective diagnosis.[82]

Alternatively, and if certain conditions are met, rules-based AI or software that uses locked models also could be developed, tested, and validated using research that utilizes a waiver of HIPAA Authorization through an institutional review board. Section 3024 of the 21st Century Cures Act provided FDA with the authority to permit consideration of such research in their regulatory decision-making, and FDA issued a guidance regarding this change in July 2017.[83]

The HIPAA Security Rule established national standards around the protection of patient information created, received, used, or maintained by a covered entity.[84] The Security Rule will need to be taken into account if PHI is used in the development and testing of CDS software, including DxSS. In addition, if the use of DxSS requires data to be sent off-site for processing, the Security Rule would again apply. Encryption and HIPAA-compliant databases and cloud services can be used to mitigate risk of a breach, but an "accurate and thorough" assessment of risks to PHI will be required.

European Union's General Data Protection Regulation[†]

The European Union's General Data Protection Regulation (GDPR) implements new protections to increase the control that individuals have over their data, enabling them to demand access to or request deletion of their information. The GDPR has a combination of regulations in place establishing what may amount to a "right to an explanation" when it comes to machine learning algorithms.[85,86] In particular, Article 22 outlines rights and obligations around the use of automated decision-making and grants individuals the right to object to any decision made about them if that decision is purely based on automated processing.[87] Individuals also have the right to obtain information about the logic involved in the automated decision-making system, its significance, and any resulting consequences.[88] Diagnostic tools previously not considered medical devices may be considered medical devices under the new regulations if they have a purpose of "prediction and prognosis."[89]

---

[*] A business associate is considered any person or organization, other than a member of a covered entity's workforce, that performs certain functions or activities on behalf of, or provides certain services to, a covered entity that involve the use or disclosure of individually identifiable health information.[90]

[†] While this paper is primarily interested in US-based laws and regulations regarding the use of DxSS in the US, we mention the European Union's GDPR because it may affect training data collection and use, as well as serve as a blueprint for future federal and state laws.

Several states have their own data privacy protection laws and regulations that must be followed. Thus, companies that plan to operate in multiple states will need to ensure their compliance with each state's laws. For example, California's medical privacy laws apply to vendors of individuals' personal health records, while HIPAA applies only if the vendor is a covered entity or a business associate of a covered entity. California also provides greater ability for individuals to sue in the event of a data breach.[91] In the summer of 2018, California passed the Consumer Privacy Act (CPA). While HIPAA-protected data are not affected by the CPA, this legislation may still affect CDS software if it makes use of other types of data such as information from consumer devices including activity trackers, voice assistants, or sleep monitors.

Federal Trade Commission Regulations

The Federal Trade Commission (FTC) regulates misstatements in trade, requiring that all statements about a product must be based on reliable scientific evidence.[92] FDA and FTC have a joint Memorandum of Understanding (MOU), a non-binding agreement between parties, to coordinate efforts to protect consumer rights. The MOU delineates that FDA will have primary responsibility in regards to the labeling of foods, drugs, devices, and cosmetics, with the FTC retaining primary jurisdiction over regulating the veracity of their advertising.[93] Thus, diagnostic programs, AI or otherwise, must be able to substantiate their claims of effectiveness. Though the FTC has not precisely defined what evidence is required, the agency applies a fairly high standard.[94] For example, the FTC questioned a mole-detecting app's effectiveness because it could not diagnose as effectively as a trained dermatologist, even though evidence showed the app to be more accurate than a typical primary care physician, and later prohibited the company from making any misleading or unsubstantiated claims about the health benefits or efficacy of the product.[95,96]

The FTC also has a regulatory role when a company claims that patient or consumer information will be secure but then fails to take reasonable security measures. Currently, the FTC does not have a specific rule for what qualifies as reasonable, instead comparing each potential violator to evolving industry standards.[97]

Intellectual Property Protections

Intellectual property protections may pose another obstacle for companies looking to invest in this space and develop AI-embedded health technologies, with two Supreme Court patent eligibility cases at the forefront of current debate. In *Mayo Collaborative Services v. Prometheus Labs, Inc,* the Supreme Court held that a series of steps to determine the appropriate dosage of a medicine was not patentable because the steps did not involve an inventive step beyond identifying a law of nature.[98] Later, in *Alice Corp. v. CLS Bank International*, the Supreme Court held that using computer software to implement a goal that is otherwise not patentable does not automatically make the goal patentable.[99]

Because diagnosing disease is not patentable, and because it cannot automatically become patentable by implementation of it through software, DxSS faces significant questions as to its patentability. [100,101] Patent law is designed to spur innovation, so if diagnostic tools' patentability is in question, some are concerned that diagnostic research investment will decrease.[102] However, empirical evidence does not indicate that a decrease has taken place to date.[103]

Section III: Key Priorities for Advancing Safe & Effective AI-Enabled CDS Software & DxSS

AI-enabled CDS software in general, and DxSS specifically, has the potential to transform health systems and revolutionize the way care is delivered. For this to happen, however, stakeholders will need to address multiple concerns around the development, regulation, and adoption of safe and effective SaMD. In this section, we identify and discuss priorities issues that could benefit from additional clarity and, in some cases, consensus, in order to harness these technologies safely and effectively, while incentivizing continued innovation in this space.

First, this section will discuss the evidentiary needs for increased adoption of these technologies. Providers need to be able to assess the value and return on investment of these technologies, including how a specific piece of software will affect their patients' health outcomes, the total cost of an episode of care, and the system's workflow.

Next, this section will examine challenges around effective risk assessment of these products, including the ability of regulators and users to understand what AI-enabled SaMD is doing, whether labeling is able to assist in understanding when the software is or is not appropriate for use, the relevant risks and benefits of locked vs continuously learning SaMD, assessment of liability in the use of this types of SaMD, and methods to evaluate performance over time.

Finally, this section will explore challenges around the use of data to develop, use, and monitor performance, including mitigation of bias, assessment of the scalability of AI-enabled DxSS, protection of patient privacy, and cybersecurity.

Priorities for Improving Adoption of DxSS by Provider Systems

Priority	Questions to Consider	Potential Solutions
Demonstrating the Value of DxSS to Provider Systems	Where do clinicians perceive the greatest need and value for diagnostic support?	• Identify clinical domains with high diagnostic uncertainty or where knowledge requirements are quickly changing • Identify areas with potential for significant clinical productivity gains
	Does the software improve outcomes, increase quality, reduce costs, or improve provider work-life?	• Developers will need to show clinical and/or economic evidence using data from a population representative of the health system • Developers that can show reduced costs within episodes of care (e.g., reducing the need for other types of diagnostic tests, targeting conditions that are more easily treated when caught early) may be particularly attractive to provider systems that use value-based payments
	Does the DxSS fit into the existing workflow?	• Developers should engage a range of perspectives in the design and development process (e.g., front-line physicians and data experts)
	Can the real-world performance of the system be evaluated?	• Software will need the functionality to show real-world performance and clinical utility of DxSS products over time • Methods will be needed to access "gold standard" comparison

Determining the Right Time and Location to Deliver Support	When is the right time and location to deliver diagnostic support or recommendations?	• Integrate software into the existing workflow
		• Research on whether:
		  ▪ Recommendations are more likely to be considered if the clinician has not already come to a decision?
		  ▪ DxSS is more effective when it runs automatically for each patient, or only when the clinician perceives a need for it?
		  ▪ Trust in the system affects perception of usability and overall effectiveness?

Securing Health Plan Coverage and Payment for DxSS	Will use of DxSS be reimbursable?	• Establish which DxSS features and performance outcomes encourage coverage and payment and when coverage is necessary for provider adoption • Developers will need to show clinical and/or economic evidence using data from a population representative of the payer's customers • Public and private payers should provide clarity on the evidence needed for different use cases

Adoption of AI-enabled DxSS is a two-stage process in which the provider system decides to adopt the software into the workflow and the individual physicians trust the software enough to have it augment their clinical decision-making. This section goes over evidentiary needs to encourage adoption of these technologies. Risk assessment and the degree to which provider systems and clinicians can understand what the software is doing is also highly pertinent to increased adoption, and will be discussed more fully in the next section starting on page 25.

Demonstrating the Value of DxSS to Provider Systems

Marketplace adoption of DxSS products generally requires evidence that demonstrates where this technology is most needed and can provide the greatest value. Evidence that use of the software meets one or more of the Quadruple Aims (i.e., improve outcomes, increase quality of care, reduce costs, or improve provider work-life) within an episode of care can also help drive adoption.

Ideal settings for DxSS to bring value are clinical domains or settings of care that have demonstrated frontline desire for diagnostic support. Such domains could include settings where clinical uncertainty is high or knowledge is quickly changing, where a need exists to reduce specialist referrals or other types of diagnostic tests, or where software can demonstrate significant clinical productivity gains. Provider systems taking on accountable care arrangements and alternative payment models that better support overall improvements in patient outcomes and total costs of care, such as Accountable Care Organizations (ACOs), may be particularly interested in SaMDs. While these technologies can potentially create more efficient diagnostic and care pathways, such as reducing the need for imaging studies, diagnosing sepsis early, or preventing hospital readmissions, evidence demonstrating these impacts will be needed to justify use. A 2017 JASON report on "Artificial Intelligence for Health and Health Care" identified evidentiary needs as a critical concern, finding that changes to standard of care normally require robust peer-reviewed research and warning against the use of poorly-validated AI-enabled software. The report recommends creating approaches to testing and validating AI-enabled software that use conditions that differ from the training set.

On an individual product level, DxSS software packages will need to verify the accuracy of the product with data that reflect the patient population served by the individual health systems considering adoption of the product. In addition, health systems will likely require methods to evaluate the real-world performance and clinical utility of DxSS products over time in the context of their patient population, including improvements in pre-specified health, quality, or cost measures for that population.

Finally, DxSS needs to be designed to fit into the workflow of the setting of care where it is to be deployed and potentially validated for accuracy within individual systems, particularly if the data the software uses varies among provider systems (e.g., EHR data). An effective DxSS design can be accomplished by engaging a range of perspectives in the development process, including front-line physicians equipped with therapeutic expertise and knowledge of the workflow and patient populations, as well as data experts.

Determining the Right Time & Location to Deliver Support

Getting the right information to the right person at the right time and location is critical for effective CDS software in general, but perhaps even more so for DxSS. Clinicians expect highly discriminative, high-value guidance, which requires that developers understand when is the "right time" and where is the "right location." Ideally, CDS software would be integrated seamlessly within the clinical and administrative workflow in a way that reduces cognitive burden and does not add new or further encumbrances.

More research is needed on when, where, and how CDS software in general, and DxSS specifically, should deliver information and recommendations for maximum effectiveness. Should CDS software act proactively or should it prompt clinicians to reconsider a decision? Some work on this subject has suggested that CDS should be timed in such a way that clinicians are not required to repeat thought processes, but diagnostic support may differ from other types of CDS.[104] When in the clinical workflow are recommendations generated by CDS software most useful? Should recommendations only be provided at the request of the physician? Types of DxSS where clinicians have the option of turning on the software only when they perceive a need for it may be more acceptable, similar to how they would order other types of diagnostic tests or would consult a specialist.

Timing of recommendations will be partially dependent on when all the required data have been entered into the system as well as any processing time required, which may be an important limitation for AI-enabled software. But if recommendations can be made during the patient encounter, additional considerations remain. If a recommendation is delivered too early or often, alert fatigue may develop, particularly if a delivered recommendation is deemed obvious, if it takes up valuable space on the computer screen, and/or the recommendation requires additional "clicks." If the recommendation is made later in the workflow, the software is tasked with changing the clinician's mind rather than simply providing recommendations to consider. The effectiveness of both techniques will be dependent on the overall trust users have in the accuracy of the system.

Securing Health Plan Coverage & Payment for DxSS

As stated above, the return on investment (ROI) is clear for some DxSS technologies, particularly those that have demonstrated impact on improving provider system efficiency and better enabling providers to meet key outcome and cost measures that are tied to their reimbursement from payers. Further driving adoption and increasing the ROI for these technologies will also include payer coverage and reimbursement to the provider systems. A useful first step would be to establish which DxSS features and performance outcomes will be most valued by payers, as well as the level of validation needed.

Currently, DxSS that are used in a manner that is similar to how diagnostic tests are used (i.e., ordered specifically for individual patients as confirmatory evidence for a suspected condition or as a screening test for a defined population) have a clear pathway for coverage and payment, contingent on evidence of clinical usefulness. An example of such a test is IDx-DR, which is deployed in a primary care office to screen for diabetic retinopathy.

However, DxSS that has been integrated into EHR systems or other clinical workflows and meant to be used on every patient regardless of the physician's need for such support may be viewed differently. Payers are likely to be hesitant to pay individually for services provided to all patients, although the case could be made to pay for a constellation of narrow DxSS services when the product performs similarly to a traditional diagnostic tests (e.g., MaxQ AI).

Clarity is needed around use cases for DxSS that public and private payers would consider appropriate for coverage. Costs involved with using the software would ideally be low compared to the savings the product produces. However, models could include savings from using the software to automate treatment approvals, saving timing on writing and reviewing authorizations.

Priorities for Assessing Patient Risk of AI-Enabled Software

Priority	Questions to Consider	Potential Solutions
Understanding the Effect of Interpretability and Explainability on Assessment of Risk	How does interpretability and explainability affect the evaluation of whether CDS software is a medical device?	• Seek regulatory clarity regarding the information needed for clinicians to "independently review the basis for [software] recommendations," which is required for software products to fall outside the definition of a medical device
	How does interpretability and explainability affect the benefit-risk assessment of SaMD generally and DxSS specifically?	• DxSS by definition "informs" or "drives" decision-making by a clinician, which are lower-risk activities than "treating/diagnosing" independently • Seek regulatory clarity on how interpretability or explainability may affect FDA risk determinations, both in the current regulatory regime and within the proposed pre-certification pathway
	How does interpretability and explainability affect the adoption of DxSS by health care systems?	• Research how trade secrecy may affect developers' willingness to disclose how the software works or the training data used • Determine whether provider systems will demonstrate a significant preference for interpretable or explainable systems over "black box" AI-enabled software
	How does interpretability and explainability affect the use of DxSS by clinicians?	• Clinicians' training should evolve so that they can function effectively in an AI-embedded landscape • Determine what information is required to engender trust and use of DxSS, such as the clinical evidence behind the recommendation, which input data were most influential, the certainty of the recommendation, and any limitations of the model that may pertain to their patient
Assigning Allocation of Liability	Should the interpretability or explainability of the software affect the distribution of liability for AI-enabled software?	• Balance the allocation of liability to ensure that entities and persons with the most knowledge of risks, and best positioned to mitigate risk, are incentivized to do so
Reimagining Labeling	Are traditional medical device labelling requirements the correct approach for AI-enabled software?	• Re-evaluate traditional approaches to labeling to address unique concerns around AI-enabled SaMD • Determine if a new risk framework or labeling classification system could better depict algorithmic safety and efficacy
Understanding the Risks and Benefits in Using Locked vs Continuously Learning Models	What are the relative risks of using continuously learning algorithms vs locked models?	• Explore the relative risks of continuous learning algorithms beginning to behave in unexpected ways (such as introducing bias or lowering performance in smaller patient subgroups) vs locked models unexpectedly degrading in performance • Determine whether use of automated verification and validation tests and/or automated performance tracking may be able to mitigate these risks
Evaluating Performance over Time Effectively	Can SaMD performance be effectively evaluated using real-world evidence (RWE)?	• While the software itself would have some of the data needed to monitor performance, other real-world data (such as claims data or patient-reported outcomes) may be required, particularly to collect high-quality information on outcomes • Determine the accessibility of the data needed to effectively monitor performance of SaMD

Understanding the Effect of Interpretability & Explainability on Assessment of Risk

AI-enabled SaMD, including AI-enabled DxSS, can be complex and the ability to "understand" what a SaMD product is doing may mean different things depending on the context. The degree to which regulators and clinicians can understand how the software works and the information that clinicians are given regarding the basis for individual recommendations will affect stakeholders' assessments of risk to patients when clinicians use this software. These risk assessments will have implications during regulatory clearance/approval, adoption consideration, clinical use, and legal liability.

There are several ways to disclose or describe how AI-enabled software works:

- Model disclosure—the exact function(s) which are used to compute how all inputs are weighted and combined to produce the outputted recommendation (i.e., the actual code being used in the software).[*] Depending on the complexity of the function, this may not be comprehensible to a human.
- Algorithmic disclosure—the code for the learning algorithm and parameters used to create the software as well as the training data used (i.e., all the information required to recreate the model). Algorithmic disclosure does not affect if the resulting model is interpretable or explainable.
- Interpretability—a human-comprehensible explanation of exactly how the model combines and uses inputted data to come to a specific recommendation. Also referred to by computer scientists as "model transparency".
- Explainability—a human-comprehensible explanation of how a 'black box' model is statistically likely to have come to a specific recommendation.

Interpretability and explainability are terms used in multiple overlapping ways by computer scientists. As defined here, these may be technical descriptions that do not require that recommendation be based in peer-reviewed evidence. Therefore, information may need to be presented in different ways for an individual clinician to understand and trust AI-enabled DxSS. This is further explored starting on page 28.

Evaluation of Whether the Software is a Medical Device

The language in the 21st Century Cures Act that exempts certain CDS software from FDA authority also requires that the software allows a "health care professional to independently review the basis for its recommendations so that it is not the [manufacturer's] intent that such health care professional rely primarily on any of such recommendations to make a clinical diagnosis or treatment decision regarding an individual patient." FDA's draft CDS guidance states that rules-based AI could meet this standard by using publicly available clinical practice guidelines, published literature, FDA-approved labels, etc., but did not explain what the standards might be for data-based AI. Therefore, while human-incomprehensible disclosure alone would not appear to be appropriate, it is less clear what level of explanation would be required, including whether the explanation given is truly the basis for the recommendation or is only a statistically likely explanation. FDA's draft guidance also states "a practitioner would be unable to independently evaluate the basis of a recommendation if the recommendation were based on non-public information or information whose meaning could not be expected to be independently understood by the intended health care professional user." This again suggests neither model nor algorithmic disclosure will be adequate information if the model is not

---

[*] In the case of rules-based or locked data-based models, this is reasonably straightforward, although the code may be too complex to be comprehensible to humans. In the case of continuously learning AI, the code used to make recommendations changes continually, so models need to be associated with individual software uses.

comprehensible to the intended user. It also suggests that models trained with proprietary datasets would also not meet this standard.

Software that only uses AI to clean and structure data inputs may be able to pass the 21st Century Cures Act requirement simply by allowing the user to review the cleaned and structured data and then stepping the user through whatever rules-based algorithm was used to make the recommendation. Similarly, software that uses interpretable algorithms could step through the algorithm's decision-making process but only if the underlying model used published information. It is less clear exactly what information might be required to be provided to a medical provider for "explainable" AI-enabled software to be considered exempt from the definition of medical device.

AI-enabled SaMD that does not allow clinicians to independently review the basis for its recommendations can still be marketed after undergoing any required FDA regulation (approval, clearance, notification). Nevertheless, the ability to avoid the burden of regulatory oversight might incentivize developers toward interpretable AI DxSS.

FDA Review of SaMD

Currently, the IMDRF SaMD guidance uses the significance of information provided by SaMD to the health care decision and the state of the health care situation or condition as the major contributors to categorizing risk (see page 15). FDA's latest update on the pre-certification program stated that product-level elements that would contribute to the review pathway determination includes "an explanation of how the software works" as well as instructions and limitations on use" and the "critical features/functions of the SaMD that are essential to the intended significance of the information" to decision-making.[105] Within the streamlined review process description, FDA lists that one of the product-specific elements will be the clinical algorithm, including mechanism of action, but does not specify how the depth of review may change based on the information provided.

Model disclosure and details about the training, testing, and validation data are likely to be required during FDA's clearance/approval process. If the model is incomprehensible to humans or the software is using a continuously learning algorithm, algorithmic disclosure may be more pertinent than model disclosure. While recent court decisions have indicated that trade secrecy is a concern for manufacturers, FDA has a long history of safeguarding this type of information for device manufacturers.[106,107]

More guidance is needed from FDA to understand if interpretability or explainability will factor in to AI-enabled SaMD categorization both in the current regulatory environment and in the pre-certification pathway. Importantly, this guidance will need to differentiate how FDA will consider risk in regards to the degree of interpretability/explainability given in the submission package to the FDA reviewer as well as the degree of interpretability/explainability the end user will receive. For example, several recent clearances of AI-enabled SaMD that screen for diabetic retinopathy and atrial fibrillation are neither interpretable nor explainable to the end user (who just receives the recommendation), although FDA may have more complete information. Guidance would also be useful on the relevant risk assessment of interpretability vs explainability. If two products had equivalent levels of diagnostic accuracy, it would be reasonable to assess interpretable models as lower risk compared to explainable models (similar to a medical product in which the mechanism of action is well understood) which should have the effect of incentivizing the development of these types of models. FDA has stated that it plans to develop and evaluate a support tool to help organizations determine if their products meet the definition of a medical device, if it falls under enforcement discretion, or what would be the appropriate risk categorization.[108]

Adoption of AI-Enabled DxSS

As mentioned above, adoption of AI-enabled DxSS is a two-stage process, in which the provider system decides to adopt the software and the individual physicians decide to use it. Interpretability and explainability of DxSS will have an impact at both steps of the process.

Risk assessment will be complicated if developers are reluctant to disclose trade secrets, and it is likely that the incentives are strong to keep training data private or proprietary to compile with data privacy protections and/or secure and maintain a competitive advantage. Trade secrecy concerns will also likely preclude model or algorithmic disclosures to the provider systems using the software. However, the decision by a provider system to adopt of a piece of SaMD will likely require detailed explanations of how the software works, as well as diagnostic sensitivity, specificity, and other pertinent details. The number of inputs used by the software can impact both interpretability and explainability, as can the algorithm type (e.g., deep learning is generally less interpretable than decision trees or other forms of machine learning, although explainability may still be possible).* It is still unknown whether provider systems will demonstrate a significant preference for interpretable or explainable systems over 'black box' AI-enabled software, if the systems have equivalent accuracy. In the case of increased accuracy, is a reduction in interpretability acceptable? There also may be differences in how provider systems think about the systems developed in-house vs commercial products and about higher- vs lower-risk software.

Provider systems will also be cognizant that overly technical descriptions of the software will likely be of little to no value to clinicians using the software. Individual clinicians using AI-enabled software should not be expected or relied upon to comprehend the capabilities and capacity of AI systems in general, or its ability to aid with a patient's diagnosis in particular.[109] Software that can communicate to the clinician information such as clinical evidence behind the recommendation, the input data that were highly influential of the final recommendation, the certainty of the recommendation and any limitations of the model that may pertain to their patient, in a way he or she can quickly understand within the timeframe and context of a patient encounter, will likely be preferred over otherwise equivalent product.

Just as clinicians know that a particular drug does not always work, they will need to understand that AI-enabled software will not be perfect, or equally dependable in every instance. Clinical utility will only be realized if users are able to understand, trust, and manage AI technologies. With the introduction of AI into CDS, clinician training will need to evolve, in medical school curriculums or through continuing medical education so that clinicians function effectively in an AI-embedded landscape. However, this education is unlikely to be sufficient for model disclosure or even interpretability to be a useful tool for risk reduction within the context of a patient encounter. What developers may think is explainable may not translate clinically, and some clinicians may be more interested in the operating characteristics (similarly to how most people do not know how a car's engine works but they do know how to drive them in varying conditions).

Another way to think about "understanding AI" is how it helps clinicians comprehend the critical input factors and limitations of the software, including the patients to which the software can be safely applied. In addition to being an important aspect of FDA's assessment of risk and labeling requirements, whether AI-enabled DxSS provides this type of information (and the quality of the information) will be of interest to provider systems deciding which software to adopt and to clinicians when they do not automatically concur with the recommendations provided. In cases where an AI-enabled DxSS conflicts

---

* There are multiple research efforts underway on interpretable AI-enabled models for health care as well as groups working to make non-interpretable machine learning methods explainable, known as XAI.[110,111,112]

with the clinician's diagnosis, knowing that the input data being used by the software are accurate and that the software is applicable to patients like theirs (as well as the overall validated accuracy of the software) can help establish the trust and acceptability needed for the clinician to seriously consider the recommendation. Another method that may increase trust and acceptability is if the software displays an indication of the certainty or uncertainty of the recommendations.

Establishing the minimum level of understanding that the end user needs to effectively apply DxSS to patient care can help with clinical uptake and instituting a culture of trust based on credibility of the model and its underlying data will solidify clinician and patient belief, or confidence, in DxSS.

Assigning the Appropriate Allocation of Liability for AI-Enabled SaMD

An appropriate distribution of liability among developers, entities that mandate or otherwise compel use (i.e., health systems or commercial payers), and end users can be expected to have an effect on a health care system's evaluation of the benefits and risks of AI-enabled software and, therefore, their adoption of DxSS. A balance will need to be struck such that allocation of liability ensures that entities and persons with the most knowledge of risks and best positions to mitigate risk are incentivized to do so as opposed to those with the least knowledge of risk and ability to mitigate. The allocation of liability should be followed by appropriate incentives to ensure safe development of products and that injured patients are compensated while also spurring innovation, motivating manufacturers to enter the space, and encouraging providers to use AI technology.

The introduction and uptake of AI-enabled CDS software by health care systems will depend upon relevant laws and regulations on medical liability, which vary widely across the United States and around the world.[113] As the real-world integration of AI continues, the scrutiny of liability for harms caused by AI will increase.[114] Existing tort law expects and compensates based on foreseeable, or intended, harm. For DxSS, harm caused by a poor recommendation if implemented is foreseeable. However there is complexity in identifying which entity is responsible for understanding the quality of the recommendation. For example, a DxSS recommendations may be foreseeably problematic by clinicians who understand the medical decision being made. However, forms of AI-enabled DxSS which are not interpretable or explainable may leave clinicians with little or no method by which to obtain a modicum of foreseeability over which recommendations are low-quality.[115] Paradoxically, the unforeseeability of AI decisions is foreseeable.[116] While some commentators argue against making designers strictly liable, because such high liability could deter innovation, other commentators argue that the solution is to make manufacturers or designers of AI devices strictly liable for all harm caused.[117] Volvo, for example, has embraced this logic for AI by announcing that they will cover any damage caused by their autonomous vehicle software in the future.[118] Similarly, the developer of IDx-DR has testified before the FTC that it accepts liability for IDx-DR and has medical liability insurance as a result.[119] However, it should be noted that both of these situations involve autonomous decision-making by the software, rather than augmented expert clinical decision-making.

Current medical device law suggests that degree of regulation (i.e. which pathway used for approval/clearance) required by FDA for marketing affects the liability exposure to the manufacturer. *Riegel v. Medtronic, Inc.* held that medical devices approved through FDA's PMA pathway are immune from tort liability in state court both from direct damages caused by the product as well as indirect damages (such as failure to warn) through the "pre-emption doctrine."[120] However, devices that go through the 510(k) pathway are not protected by the pre-emption doctrine, based on an earlier court decision *Medtronic, Inc v. Lohr.*[121]

The degree of interpretability and/or explainability, and well as how pertinent information is displayed alongside recommendations is likely to be important to clinicians as they consider the amount of liability

they may be willing to accept, especially when DxSS results differ from their expert clinical decision-making. Clear labeling that is accessible to clinicians during use will be critical for clinicians to assess and mitigate risk.

Reimagining Labeling to Better Promote Understanding of AI-Enabled SaMD

Once a product has entered the market, FDA has historically sought to address device-specific safety issues through labeling requirements. Labeling for medical devices describes the indication for use based on the features of the product and the clinical evidence on which it was trained and validated. AI-enabled software, particularly software that uses continuously learning algorithms, will challenge this existing regulatory paradigm and raise the question of whether the current paradigm for labeling is the correct approach for AI-enabled software.

The relative newness of AI-enabled SaMD likely means that the effectiveness of approaches to labeling will need to be evaluated and should evolve over time. A new risk framework or labeling classification system may need to be created in order to better define a process that correctly depicts algorithmic safety and efficacy. For example, the risks of using an AI-enabled product off label may not be obvious or apparent to clinicians if they do not understand the input data requirements, applicable populations, etc.

Understanding the Risks and Benefits in Using Locked vs Continuously Learning Models

The decision to use continuously learning algorithms versus "locked" AI-enabled models in the software requires a serious understanding of the risks of each alternative. As of now, FDA has only cleared AI-enabled software that uses locked models for their recommendations. This is due to the risk that continuously learning algorithms may learn in unexpected ways, introducing bias or lowering performance in smaller patient subgroups.

However, as previously discussed, a "locked" algorithm performance may also unexpectedly degrade if the data entered into the software begins to differ significantly from the training data in quality, completeness, or content due to changes in clinical guidelines or billing requirements, new treatments, upgrades in medical imagers/sensors, etc.[*]

AI-enabled SaMD that makes use of either continuous learning or locked models in the recommendation generation stage will need to clarify how it continues to learn once "in the field," particularly the process in which new examples are labeled. This would be a particular problem for continuous learning software, where in many cases it is not immediately clear what the "true" diagnosis may be.[†] However, locked algorithms will also need to continue to collect new labeled examples over time in order to prevent performance degradation over time, as discussed above. Product labeling for updates should

---

[*] This is a particular concern with EHR data inputs, which can change substantially over time. The criteria for diagnoses can change, new treatments can be introduced, the precision or completeness of data may change due to automatic pulls of data from medical devices or standard-of-care changes—all of these changes may affect the performance of a locked algorithm which depends on those data. For example, if an algorithm uses the presence of particular prescription drugs as a contributor to its diagnosis recommendation (possibly because of known side effects), a new drug with the same potential side effect would not be accounted for in the locked algorithm.

[†] There are functions where AI-enabled continuous learning within SaMD may be more straightforward. For example, if data-based AI is used to clean and structure the input data (e.g., natural language processing of free text notes) then the software could display the resulting structured data and allow clinicians to correct any mistakes.

indicate to users how these new examples are labeled if the process differs from that of the original training dataset.

Finally, it is critical for FDA and software adopters to understand how software updates are verified and validated before being sent out to the field. Continuous learning algorithms may make use of automated verification and validation tests, using a pre-specified testing set to ensure that its performance after each change stays within pre-specified thresholds of accuracy and sending out alerts and/or preventing the changes if those boundaries are exceeded. However, locked models could use similar types of automated testing, particularly if the updates happen frequently or are setting-specific. FDA guidance states that quality systems which "demonstrate that the manufactured device meets the change in design specifications (or the original specifications, if no change was intended)" can be used to reasonably ensure the safety and effectiveness of SaMD that has been modified.[122] These quality systems should include risk assessment followed by successful, routine verification and validation activities, but it isn't clear if humans need to be involved in the verification and validation assessments unless an alert is generated. In addition, more clarity is needed to understand when modifications or updates to AI-enabled SaMD will require submission of a new 510(k) or a supplemental PMA to FDA, and when these quality systems will suffice.

Evaluating Performance over Time Effectively

Due to the challenges laid out in the section above, post-market performance surveillance will be very important with these AI-enabled CDS products. FDA currently has the authority to require post-market surveillance of the products it approves or clears for marketing. The proposals regarding the pre-certification program of SaMD anticipate strong post-market requirements. Regular real-world evaluations of performance in the marketplace could mitigate the risks posed by software. FDA clearances or approvals could include post-market requirements to actively monitor performance of the SaMD, such as regular reporting and/or setting performance measure thresholds that would require immediate notification of FDA if missed.

However, while the software itself would have some of the data needed to monitor performance, other real-world data (such as claims data or patient-reported outcomes) may be required, particularly to collect high-quality information on outcomes. It could be useful to have a Sentinel-type surveillance system that reports back on a regular basis with performance metrics. FDA has stated its commitment to the National Evaluation System for health Technology (NEST) becoming a strong partner in such a system.[123] Having mutually agreed upon pre-specified accuracy thresholds[*] that must continue to be met to ensure the safety and effectiveness of AI-enabled products may be a prerequisite to gaining and maintaining clinician confidence that product performance is stable or improving rather than degrading over the product's lifecycle. In the future, there may be a role for a range of stakeholders—including medical associations, algorithm developers, hospitals, health systems, and third-party certifiers—to monitor AI technologies over time for compliance with the threshold performance standard.[124]

---

[*] These thresholds are similar conceptually to the thresholds discussed in the previous section in regards to software updates. However, standards for collection of output data (i.e., the "correct" diagnosis labels) may differ and performance analysis may be monitored by the health care system rather than or in addition to the software manufacturer. The data and analyses used for real-world performance analytics therefore may be different and separate from any data being used to continue to train the software in preparation for updates.

Priorities for Using Data Responsibly

Priority	Questions to Consider	Potential Solutions
Mitigating Bias	How can stakeholders ensure that AI-enabled software does not perpetuate or introduce biases?	• Develop best practices in determining the "gold standard" for labelling the training and testing datasets • Policymakers should support the development of and access to research databases of labeled health data, in addition to data at levels that exceed standard clinical assessment
	How can developers ensure that they have diverse representation of patients in their training sets?	• Developers will need to account for various factors such as demographics, social risk factors, and cultural norms • Developers should ensure that they have appropriately diversified training sets that have been annotated for subgroup analysis
Evaluating Algorithmic Adaptability	How can developers of DxSS and potential adopters of the software gauge the "brittleness" of the product?	• Developers will need to understand how variability in input elements may affect software performance • Provider systems need to understand their own data, particularly how workflow and other issues may affect the accuracy and completeness of critical input elements • Understand how input data types affect the performance of algorithms across multiple sites
	How can developers of DxSS anticipate the expanded use of their products and build in efficient scalability?	• Develop common data models and data dictionaries for specific clinical domains that are reflective of specific diseases and their underlying biology • Research how input data types vary across health care systems, and how that may affect data element accuracy and completeness.
Maintaining Patient Privacy and Ensuring Ethical Use of Data	How can data be shared for research purposes, without stunting innovation, while preserving individual privacy?	• Increased security standards, establishment of certified third-party data holders, and regulatory limits on downstream uses of data • Promote better stewardship of data by establishing a national framework that helps guide companies using personal data
Safeguarding AI-Enabled SaMD from Cyberthreats	How can developers account for cybersecurity risks?	• Integrate cybersecurity into the design, development, implementation, and maintenance of the system • Developers will need to evaluate and mitigate cyberthreats to their product's performance to meet federal regulations using existing guidance

In order to adequately train the model behind AI-enabled CDS software, large volumes of data are often needed. There are critical challenges around understanding how the data used in developing software can affect the suitability of using that software with individual patients and in different care settings. Stakeholders also need to come to consensus on the ethics in using and storing data responsibly, as well as appropriate safeguards to protect against cyberattacks.

Mitigating Bias

While there is legitimate excitement that AI may have a significant role in helping to reduce health disparities, AI-enabled CDS software (including DxSS) could potentially exacerbate such disparities, depending on the population used to create and test the algorithms. Ensuring that AI-enabled software does not perpetuate or introduce biases, either by perpetuating inequalities or introducing new discrepancies, is a critical challenge. In order to mitigate algorithmic bias, developers will need to take into account the variability in risk within and across populations, ensuring that the labeled examples being used for training the model are robust enough to support diagnosis of the population on which the CDS software will be used, and that the labeling clearly informs the user of the populations for which results may be less reliable.

Anticipating and accounting for introduced bias will require a deep understanding of the labeled examples being used. This includes understanding the representation of the types of patients that the software will be analyzing. Factors that may need to be accounted for include, but are not limited to, the patient profile (e.g., age, gender, race, and social risk factors) and cultural norms (e.g., procedure options and patient preferences, trust in clinician, expressing symptoms). Existing training datasets may not be inclusive of these factors and, moving forward, developers should ensure that they have appropriately diversified training sets that have been annotated for subgroup analysis if they wish to see widespread uptake of their software.

In addition, the labeled examples themselves also will need to be examined for existing biases in certain populations, such as misdiagnosis of heart failure or myocardial infarction in women. AI-enabled software developed with "data that encode human bias will reproduce, not eliminate, the bias."[125] As such, the "gold standard" used to label the training and the testing sets is critical. As the previously mentioned 2017 JASON report states, AI algorithms can only perform at levels matching their training sets. The report found that access to high-quality labels on data sets is a significant barrier to the development and evaluation of AI-enabled software for CDS, and it recommended that policy makers support the development of and access to research databases of labeled health data, as well as labeling data at levels that exceed the standard clinical assessment (e.g., using biopsy results to label dermatological images).[126]

Evaluating Algorithmic Adaptability

It may be important for developers of DxSS to anticipate expanded use of their products in the future and consider how system-dependent, or "brittle," the algorithms driving the software are. The brittleness of algorithms is an important consideration if the same model or approach will later be able to be used at multiple sites and must transfer with ease.

Because of workflow differences between provider systems, clinical data in general and EHR data in particular are considered extremely variable. Levels of conformance, accuracy, and completeness will differ by site and by data field. Details that may be recorded in some settings of care may not be recorded in others. Characteristics of the providers and the provider systems that entered the data used for training can also introduce discrepancies if they differ from the setting or location where the software will be used. Regional and setting-of-care differences in clinician profiles and preferences (e.g.,

workflow, training, treatment resources, individual bias), location (e.g., prevalence of disease, local outbreaks), provider system (e.g., workflow, billing), and payer requirements (i.e., what information is required for payment) can introduce biases to an AI algorithm. Custom EHR implementations used by many large health care systems as a result of these workflow differences mean that data elements and definitions can be quite variable as well. Until common data models and data dictionaries are developed for specific clinical domains that are reflective of specific diseases and their underlying biology, artisanal-style curation of the input elements will often be needed, meaning fewer input variables may be desirable and/or individual sites will need customized systems.

Medical imaging and sensor data are more standardized but potential differences should nevertheless also be considered. For example, a study done by Mount Sinai researchers compared performance of an algorithm meant to detect signs suggestive of pneumonia in X-rays using training data from within a hospital system versus training data from elsewhere. In three out of five natural comparisons, performance was significantly lower if the training was done using external data.[127]

Developers should consider the incremental value of each data input, balancing any demonstrated increase in diagnostic accuracy with the burden of collection of that element in workflow. For EHR inputs, developers should consider how and why each input element is entered into the EHR and how that may affect accuracy and completeness, comparing and contrasting different types of provider systems that may use their software.

If an algorithm is proven to be truly valuable to clinicians, they may be willing to change workflow in return for improved decision-making. This may lead to improved data quality (i.e., reliability, completeness) over time (which ironically could degrade the performance of a locked algorithm over time), as explained on page 30.

Maintaining Patient Privacy and Ensuring Ethical Use of Data

Machine learning requires vast amounts of health care data to accurately determine a patient's diagnosis. Understanding how to train AI algorithms using patient data while preserving an individual's privacy remains a challenge. Enabling data sharing for research and validation purposes while better protecting data subjects' privacy is a significant concern. Data-based AI may exacerbate this concern because it often requires large amounts of fairly comprehensive personal health data to be most effective, and it may provide insights into a person's health that result in additional privacy problems.[128]

In 2017, the United Kingdom's data protection regulator found the Royal Free NHS Foundation Trust in violation of patient data rights when it shared protected patient information with Google's DeepMind, prompting renewed efforts to increase transparency through a patient and public engagement strategy as well as formal data-sharing agreements.[129] The introduction of AI-enabled SaMD may give stakeholders the opportunity to reimagine privacy protections in ways that ensure privacy at the right level for data sharing and storage for various purposes. Among other possibilities, increased security standards (described below), trusted third-party data holders, and limitations on downstream uses of data may help protect privacy while enabling innovation.

A common method for protecting the privacy of individuals is de-identification. However, de-identifying patient data before aggregation can contribute to an inherent loss of value from the data. In addition, aggregating data from multiple institutions or data sources may allow re-identification as the number of data points on each individual increases, posing additional challenges to the preservation of patient privacy throughout the entire research and development process. Though technical methods exist to reduce risk of re-identification (e.g., differential privacy and homomorphic encryption),[130] a national framework, similar to the one established by the UK Anonymization Network, can help guide companies

using personal data by promoting proper stewardship practices or encouraging the more conscious use of patient data when companies are training AI systems.[131] As an example of a real-time data audit, Google's DeepMind has engineering verification systems that provide technical assurance about what is occurring with each discrete piece of health data at each process point within its AI system, adding an entry to a special digital ledger each time interaction with the data occurs.[132]

Safeguarding AI-Enabled SaMD from Cyberthreats

Fortifying cybersecurity across an AI system will require an understanding of pertinent cyber threats and establishing an understanding of who controls and authorizes the use or transfer of health data. It is anticipated that cybersecurity risks will vary depending on the status or function of the data (i.e., training vs testing data), and where it is located within the system (i.e., stored vs streamed through). Therefore, to safeguard data and to ensure personal information is handled appropriately, cybersecurity will need to be integrated into every aspect of the design, development, implementation, and maintenance of a system.[133]

Companies also will be expected to evaluate and mitigate cybersecurity threats to their product's performance in order to meet federal regulations. An October 2018 FDA guidance seeks to address effective cybersecurity management practices in medical devices, providing recommendations to industry on how to decrease the risk of patient harm by reducing the exploitability of devices by employing a risk-based approach to design and development, assessing risks across the TPLC, ensuring maintenance and continuity, and promoting a culture of trust.[134]

Conclusion

In the 1950s, Alan Turing wrote a paper predicting that machines would be capable of artificial intelligence by the 21st Century.[135] Now, AI systems and applications are ubiquitous and are embedded into almost every industry today, including health care. AI-enabled DxSS, as a subset of CDS, has the potential to equip clinicians, staff, patients, and others with the knowledge they need to enhance overall health and improve outcomes by supporting their decision-making processes, helping them arrive at a correct diagnosis faster, reducing unnecessary testing and treatments otherwise resulting from misdiagnosis, and reducing the amount of pain and suffering by facilitating earlier treatment initiation.

For this opportunity to be realized, the real challenges holding back safe and effective innovation in this space need to be addressed, and consensus standards need to be developed. Through the background information provided here as well as identification and discussion of the priority near-term challenges, we hope to be able to further the dialogue to move this field forward.

Appendix A: Key Terms, Definitions, and Acronyms

Acronyms

ACO	Accountable Care Organization
AI	Artificial Intelligence
CDRH	Center for Devices and Radiological Health
CDS	Clinical Decision Support
CLIA	Clinical Laboratory Improvement Amendments
CMS	Centers for Medicare and Medicaid Services
CPA	Consumer Privacy Act
CQOE	Culture of Quality and Organizational Excellence
DxSS	Diagnostic Support Software
ECG	Electrocardiography
EHR	Electronic Health Record
FDA	Food and Drug Administration
FD&C Act	Federal Food, Drug, and Cosmetic Act
FTC	Federal Trade Commission
GDPR	General Data Protection Regulation
HHS	Department of Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act
IMDRF	International Medical Device Regulators Forum
LDT	Laboratory Developed Tests
ML	Machine Learning
MOU	Memorandum of Understanding
NEST	National Evaluation System for health Technology
NPV	Negative Predictive Value
OCR	Office for Civil Rights
ONC	Office of the National Coordinator for Health Information Technology
PDS	Patient Decision Support
PHI	Protected Health Information
PMA	Pre-market Approval
PMN	Pre-market Notification
PPV	Positive Predictive Value
RWE	Real World Evidence
SaMD	Software as a Medical Device
SiMD	Software in a Medical Device
TPLC	Total Product Life Cycle

Key Terms and Definitions

Algorithmic Disclosure	The code for the learning algorithm and parameters used to create the software as well as the training data used (i.e., all the information required to recreate the model). Algorithmic disclosure does not affect if the resulting model is interpretable or explainable.
Artificial Intelligence	A term with many definitions, but in this paper refers to a machine's ability to mimic the cognitive functions of the human mind.
Augmented Intelligence	AI meant to enhance the capabilities of human clinical-decision-making rather than replace the human.
Black Box	Refers to software that does not explain how the input data are analyzed in order to come to a recommendation. This can be because the algorithm/model is too complex to be understood by humans (refer to Section III) or because the functionality is considered proprietary.
Business Associate	A person or entity that performs certain functions or activities that involve the use or disclosure of protected health information on behalf of, or provides services to, a covered entity.[136]
Clinical Decision Support Software	Software that provides "clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care" and is used to support health care providers in diagnosis, treatment decisions, and population health management.[137]
Continuous Learning Algorithm	An AI algorithm that continuously modifies itself in real-time as it is being used. Therefore, the output of the system for the same input data may be different over time.
Covered Entity	Defined in the HIPAA rules as (1) health care provider, (2) health plans, and (3) health care clearinghouses.[138]
Data-Based Artificial Intelligence	More commonly known as "machine learning," this was described by Arthur Samuel in 1959 as a computing system that has "the ability to learn without being explicitly programmed."[139] Refer to Appendix D for more information.
Diagnostic Support Software	A subset of CDS software that is specifically designed to support a clinician in coming to a diagnosis.
Disclosure	Release, transfer, provision of access to, or divulging in any manner of information outside the entity holding the information.
Enforcement Discretion	FDA has the power to exercise discretion, or judgement, on whether to enforce their rules and authorities on certain low-risk products. When they choose not to enforce their authority, it is referred to as "enforcement discretion".
Explainability	A human-comprehensible explanation of how a 'black box' model is statistically likely to have come to a specific recommendation.
General Artificial Intelligence	Designed to fully mimic human reasoning or intelligence in any context, and is capable of conscious, abstract thought.

General Data Protection Regulation	Set of regulations in the European Union that implement new protections to increase the control that individuals have over their data, enabling them to demand access to or request deletion of their information.
Health Insurance Portability and Accountability Act of 1996	Established national standards regarding the protection of patient health information and is primarily enforced by the Department of Health and Human Services' (HHS) Office for Civil Rights (OCR).
Interpretability	A human-comprehensible explanation of exactly how the model combines and uses inputted data to come to a specific recommendation. Also referred to by computer scientists as "model transparency".
Labeled Examples	A type of training data that pairs input data with what the designer would consider the correct "output" for each example.
Locked Model	A function/model that was developed through data-based AI methods, but does not update itself in real time (although supplemental updates can be made to the software on a regular basis).
Model Disclosure	The exact function(s) which are used to compute how all inputs are weighted and combined to produce the outputted recommendation (i.e., the actual code being used in the software). Depending on the complexity of the function, this may not be comprehensible to a human.
Narrow Artificial Intelligence	Refers to AI designed to address a specific application area (e.g., strategic games, autonomous vehicles), as opposed to "general AI".
Negative Predictive Value	The probability that individuals with a negative test result don't have the disease.[140]
Patient Decision Support	Similar to CDS software, but is designed to support the layperson in the decision about whether to seek medical advice and/or in following their wellness or treatment regimens.
Positive Predictive Value	The probability that individuals with a positive test result have the disease.[141]
Rules-Based Artificial Intelligence	An algorithm that answers a sequence of preprogrammed criteria using "if X, then Y"-type rules with the patient's data, leading to a discrete output. Refer to Appendix D for more information.
Software as a Medical Device	Software intended to be used for one or more medical purposes and to perform these purposes without being integral to the hardware of a medical device.[142]
Software in a Medical Device	Software that is integral to the hardware of a medical device.[143]

Appendix B: Relevant Legislation

Section 3060 of the 21st Century Cures Act

Signed into law in December 2016, the 21st Century Cures Act includes a host of provisions that are aimed at improving the FDA review process for drugs and medical devices. Specifically, Section 3060, "Clarifying Medical Software Regulation," strives to reduce uncertainty involved with FDA regulation of medical software. The following types of software are not medical devices:

- software intended to provide administrative support of a health care facility;
- software for maintaining or encouraging a healthy lifestyle (unrelated to the diagnosis, prevention, or treatment of the disease or condition);
- software that serves as electronic patient records (including patient-provided information) given that the software is not intended to interpret or analyze patient records or medical data for the purpose of diagnosis, prevention, or treatment;
- software intended for "transferring, storing, converting formats, or displaying clinical laboratory test or other device data and results, findings by a health care professional with respect to such data and results, general information about such findings, and general background information about such laboratory test or other device, unless such function is intended to interpret or analyze clinical laboratory test or other device data, results and findings."[144]

These changes codified into law that certain digital health technologies, like mobile health applications that are intended for sustaining or encouraging a healthy lifestyle without providing a diagnosis, currently fall outside the scope of FDA's regulation. Previously, FDA had exercised enforcement discretion on this type of software due to the low risk these devices pose to the general public.[145] In 2015, CDRH announced that it would be exercising enforcement discretion and would not regulate medical device data systems in order to promote necessary access.[*][146] Therefore, the fourth exclusion officially codifies FDA's position regarding the transfer, storage, conversion, and display of medical device data.

Clinical Laboratory Improvement Amendments (CLIA)

The Clinical Laboratory Improvement Amendments (CLIA) of 1988 regulate laboratory testing and requires clinical laboratories to be certified by their state, as well as CMS, before they can accept human samples for diagnostic testing (laboratories can also obtain multiple types of CLIA certificates based on the types of diagnostic tests they are performing).[147] Three agencies, FDA, CMS, and Centers for Disease Control and Prevention (CDC), are responsible for the oversight of CLIA.

FDA may categorize tests based on complexity, reviews requests for Waiver by Application, and develops rules/guidance for CLIA complexity categorization. CMS issues laboratory certificates, collects user fees, conducts inspections and enforces regulatory compliance, approves private accreditation of organizations for performing inspections, approves state exemptions, monitors laboratory performance on Proficiency Testing (PT) and approves PT programs, and publishes CLIA rules and regulations. CDC provides analysis, research, and technical assistance, develops technical standards and laboratory practice guidelines, conducts laboratory quality improvement studies, monitors proficiency testing practices, develops and distributes professional information and education resources, and manages the Clinical Laboratory Improvement Advisory Committee (CLIAC).[148]

---

[*] According to a February 2015 guidance issued by FDA, a medical device data system (MDDS) is "a hardware or software product that transfers, stores, converts formats, and displays medical device data."

Appendix C: Regulatory Pathways

Pathway 1: Pre-Market Approval

The pre-market approval (PMA) process is widely considered the most intensive device marketing application and is for Class III (high-risk) devices. PMAs require valid scientific evidence—in the form of technical, non-clinical laboratory, and clinical investigations—demonstrating the device's safety and effectiveness prior to obtaining FDA approval. However, there are exceptions for devices that can be reclassified as "De Novo," thereby foregoing the rigorous PMA application for a more expedited process.

Pathway 2: De Novo

Alternately, the De Novo pre-market review pathway can be used for new types of medical devices that are thought to be low- to moderate-risk, and do not demonstrate substantial equivalence to any previously FDA-approved predicate device. In other words, the De Novo process allows the reclassification of a novel Class III device, with no predicate, to a Class I or Class II device. As part of the submission process, the manufacturer or developer may send a De Novo application to FDA (either directly or after submission to 510(k) application) and, if granted, is able to establish a new "device type." That device is then eligible to serve as a predicate for new medical devices, where appropriate.

Pathway 3: Pre-Market Notification aka 510(k) Application

The pre-market notification (PMN) process, also known as a 510(k) application, is an expedited pathway to market for low- to moderate-risk (Class I or Class II) medical devices. The process allows for devices to obtain marketing authorization if the sponsor can demonstrate substantial equivalence to a predicate device.[*] Once FDA determines that the device has an approved predicate, the applicant can proceed with the 510(k) application submitted at least 90 days prior to the anticipated marketing date.

Pathway 3a: Class I and Class II with General Controls

General controls are basic provisions, or minimum requirements, that apply to all medical devices in order to ensure their safety and effectiveness.[149] In the PMN/510(k) review process, Class I and Class II devices with general controls require the applicant to submit necessary materials, such as device and company registration, PMN, records and reports, good manufacturing practices, along with the required user fees to CDRH. Once cleared, the agency issues a 510(k) number within 15 days of receipt of the application and indicates the initiation of the regulatory review process. The application is then directed to the appropriate review division within CDRH where the application is pre-screened for completeness. After this phase, the application moves on to a more thorough review to determine whether the new device is eligible for the 510(k) submission. FDA has an additional 45 days to conduct this more substantial review after which the application will be approved or denied.

---

[*] Showing substantial equivalence to a predicate device, or existing legally marketed device, is a required component of the 510(k) application process. Substantial equivalence means that the novel device is "at least as safe and effective as the predicate." While the devices do not need to be identical, a device is considered substantially equivalent to a predicate if:
- it has the same intended use and technological characteristics as the predicate; or
- it has the same intended use, though it may have differing technological characteristics that do not raise different safety and effectiveness concerns, and demonstrates equal levels of safety and effectiveness as a legally marketed device.

Pathway 3b: Class II with Special Controls

Some Class II devices are subject to special controls, meaning that FDA has determined that general controls alone are insufficient to providing reasonable assurance that the device is safe and effective and extra information is needed to establish assurance and these are usually device-specific. Special controls may include information on performance standards, post-market surveillance, patient registries, labeling requirements, pre-market data requirements, and other guidelines.[150]

Pathway 4: "Exempt" Status

If a Class I device qualifies for "exempt" status, a PMN application and FDA clearance are not required before the device can be legally marketed within the United States. The manufacturer is required to list their product with FDA and register their company.[151]

Appendix D: A Deeper Dive into AI Algorithms

Rules-Based AI

Rules-based algorithms can be complex, but are conceptually straightforward tools and are often represented by a conventional flowchart diagram. The algorithm answers a sequence of preprogrammed criteria using "if X, then Y"-type rules with the patient's data, leading to a discrete output. Many of the CDS software programs use "expert-systems" that create rules from information on product labels, clinical guidelines, and journal articles. These systems then step through each rule with the data on the patient it is analyzing. However, rules-based algorithms can be very complex depending on the number of variables (or discrete inputs) and the number of rules that must be stepped through. The IBM computer nicknamed DeepBlue that beat chess champion Garry Kasparov in 1997 used a massive rules-based decision tree algorithm that chose each move by going through the possible outcomes.[152] There is some overlap between rules-based algorithms and data-based algorithms, however, in that big data can be used to find the correlations that may then be used to develop rules-based algorithms.

An advantage of using rules-based decision tree algorithms is that they are highly interpretable, even if big data methods were used to create the rules the algorithm is based on.

Data-Based AI

Data-based AI, more commonly known as "machine learning," was described by Arthur Samuel in 1959 as a computing system that has the ability to learn without being explicitly programmed.[153] Machine learning techniques can be further broken down into subcategories (i.e., supervised, unsupervised, reinforcement learning) based on underlying algorithmic methodology.[154] As mentioned in the main document, this paper focuses on supervised learning which uses statistical methods to find a function that best fits the training data as they are what are most likely to be used in the development of DxSS. Training data consists of a (generally) large set of "labeled examples," which are sets of data inputs paired with associated outputs. For example, a machine learning algorithm might be developed using a training set of thousands of ECG recordings, each one labeled as to whether that recording included an episode of atrial fibrillation or not.

A very simple analogy to developing a data-driven algorithm is something people perform manually all the time: graphing a single set of inputs (or independent variables) on an x-axis, with associated outcomes (dependent variables) on the y-axis. Statistical line-fitting methods can compute a function that can be used to predict an outcome from new sets of inputs. The line-fitting function is the data-based algorithm in this analogy, because training data were used to derive the solution to classify new inputs, while the computed equation of the line is the locked model.

The power of machine learning is, given enough computational power, that one can perform much more complicated statistical analyses with an unlimited number of input variables. However, it is possible to "over-fit" the data, which may result in the derived algorithm not being generalizable to new data inputs. For this reason, data-based machine learning algorithms need to be verified with testing sets (sometimes called validation sets) of labeled examples which were not used to train the algorithm. These testing sets are used to calculate how often the algorithm correctly diagnoses an individual patient by generating the rate of false-positives and false-negatives.

There are many different types of machine learning algorithms, including linear and logistic regression, support vector machines, decision trees, boosted trees, random forest, and deep learning/neural networks.

Different types of algorithms are more suitable for different types of problems, similar to how certain statistical methods are more appropriate for certain types of analyses. The degree of interpreability and explainability for machine learning algorithms differs depending on the methodology used.

[1] National Academies of Sciences, Engineering, and Medicine. (2015). "Improving Diagnosis in Health Care." *The National Academies Press.* Retrieved from https://www.nap.edu/catalog/21794/improving-diagnosis-in-health-care

[2] National Academies of Sciences, Engineering, and Medicine. (2015). "Improving Diagnosis in Health Care."

[3] Densen, P. (2011). "Challenges and Opportunities Facing Medical Education." *Transactions of the American Clinical and Climatological Association.* Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116346/

[4] Tcheng, J.E., Bakken, S., Bates, D.W., et al. (2017). "Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series." *National Academy of Medicine*. Retrieved from https://nam.edu/wp-content/uploads/2017/11/Optimizing-Strategies-for-Clinical-Decision-Support.pdf

[5] HealthIT.gov. (2018). "Clinical Decision Support." Retrieved from https://www.healthit.gov/topic/safety/clinical-decision-support

[6] IDx-DR. (2018). Retrieved from https://www.eyediagnosis.net/

[7] The Human Diagnosis Project. (2018). Retrieved from https://www.humandx.org/

[8] Garrison, J.A. (2003). "UpToDate." *Journal of the Medical Library Association*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC141198/

[9] Epocrates. (2018). Retrieved from https://www.epocrates.com/

[10] FDA. (2017). "Prostate Seeding Kit." *Product Classification*. Retrieved from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPCD/classification.cfm?id=5376

[11] FDA. (2018). "Non-Invasive Central Venous Manometer. *Product Classification.* Retrieved from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpcd/classification.cfm?id=958

[12] Flaxman, S.R., et al. (2017). "Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis." *Lancet Global Health.* Retrieved from https://www.ncbi.nlm.nih.gov/pubmed?cmd=Retrieve&db=PubMed&list_uids=29032195&dopt=Abstract

[13] FDA. (2018). "FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems." Retrieved from https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm

[14] Krueger, A.C. (2018). "IDx-DR." FDA. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180001.pdf

[15] FDA. (2018). "FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems."https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm

[16] FDA. (2018). "FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems."

[17] Benjamin, E.J., et al. (2017). "Heart disease and stroke statistics—2017 update: a report from the American Heart Association." *Circulation.* Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28122885

[18] CDC. (2017). "Stroke Facts." Retrieved from https://www.cdc.gov/stroke/facts.htm

[19] FDA. (2018). "FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients." Retrieved from https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm596575.htm

[20] Ochs, R. (2018). "Accipiolx." *FDA.* Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf18/K182177.pdf

[21] FDA. (2018). "FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients." *FDA News Release.* Retrieved from https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm596575.htm

[22] Nature Biotechnology. (2018). "FDA approves stroke-detecting AI software." Retrieved from https://www.nature.com/articles/nbt0418-290

[23] FDA. (2018). "Evaluation of Automatic Class II Designation for ContaCT." *Decision Memorandum.* Retrieved from https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170073.pdf

[24] Nature Biotechnology. (2018). "FDA approves stroke-detecting AI software."

[25] Ochs, R. (2018). "Viz CTP." *FDA.* Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf18/K180161.pdf

[26] Ochs, R. (2018). "Accipiolx." *FDA.*

[27] Ibid.

[28] Ibid.

[29] Ochs, R. (2018). "OsteoDetect." *FDA*. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180005.pdf

[30] Ibid.

[31] FDA. (2018). "FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures." *FDA News Release.* Retrieved from https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm608833.htm

[32] Ochs, R. (2018). "OsteoDetect." *FDA*.

[33] Torio, C.M., & Moore, B.J. (2016). "National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013." *Agency for Healthcare Research and Quality.* Retrieved from https://www.hcup-us.ahrq.gov/reports/statbriefs/sb204-Most-Expensive-Hospital-Conditions.jsp

[34] Torio, C.M., & Andrews, R.M. (2013). "National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011." *Agency for Healthcare Research and Quality.* Retrieved from https://www.hcup-us.ahrq.gov/reports/statbriefs/sb160.pdf

[35] Rivers, E., et al. (2001). "Early Goal-Directed Therapy in the Treatment of Severe Sepsis and Septic Shock." *NEJM.* Retrieved from https://www.nejm.org/doi/full/10.1056/NEJMoa010307

[36] Yealy, D.M., et al. (2014). "A randomized trial of protocol-based care for early septic shock." *NEJM.* Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24635773

[37] Evans, D.J.W., et al. (2016). "Automated monitoring for the early detection of sepsis in critically ill patients." *The Cochrane Library.* Retrieved from https://www.cochrane.org/CD012404/ANAESTH_automated-monitoring-early-detection-sepsis-patients-receiving-care-intensive-care-units

[38] Xavier Health Organization. (2018). "Perspectives and Good Practices for AI and Continuously Learning Systems in Healthcare." Retrieved from https://www.advamed.org/sites/default/files/resource/perspectives_and_good_practices_for_ai_and_continuous_learning_systems_in_healthcare.pdf

[39] Kryzauskas, M., et al. (2016). "Is acute appendicitis still misdiagnosed?" *Open Medicine.* Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5329832/

[40] ClinicalTrials.gov. (2018). "Free Text Prediction Algorithm for Appendicitis." *U.S. National Library of Medicine.* Retrieved from https://clinicaltrials.gov/ct2/show/NCT03414853?cond=Appendicitis&cntry=SG&rank=1

[41] Ngiam, K.Y. (2018). "Bringing AI to Clinical Implementation at NUHS." *National University Hospital Singapore.* Retrieved from http://www.nextcenter.org/wp-content/uploads/2018/01/Talk-3-PPT_Ngiam-Kee-Yuan.pdf

[42] Markar, S.R., et al. (2014). "A comparative international study on the management of acute appendicitis between a developed country and a middle income country." *International Journal of Surgery.* Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24480238

[43] Ngiam, K.Y. (2018). "Bringing AI to Clinical Implementation at NUHS." National University Hospital Singapore.

[44] Boddupalli, D., Nundy, S., & Bates, D.W. (2017). "Collective Intelligence Outperforms Individual Physicians in Medical Diagnosis." Paper presented at the meeting of the Society for Medical Decision Making, Pittsburgh, PA.

[45] Abbasi, J. (2018). "Shantanu Nundy, MD: The Human Diagnosis Project." *JAMA.* Retrieved from https://jamanetwork.com/journals/jama/fullarticle/2670248

[46] Ibid.

[47] Fontil, V., et al. (2018). "Testing and improving the acceptability of a web-based platform for collective intelligence to improve diagnostic accuracy in primary care clinics." *JAMIA Open.* In press.

[48] Office of Science and Technology Policy (OSTP). (2016). "Preparing for the Future of Artificial Intelligence." *National Science and Technology Council (NSTC)*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

[49] Rajkomar, A., et al. (2018). "Scalable and accurate deep learning with electronic health records." *Nature Partner Journal Digital Medicine.* Retrieved from https://www.nature.com/articles/s41746-018-0029-1

[50] Obermeyer, Z., & Emanuel, E.J. (2016). "Predicting the Future – Big Data, Machine Learning, and Clinical Medicine. *NEJM Catalyst*. Retrieved from https://catalyst.nejm.org/big-data-machine-learning-clinical-medicine/

[51] IMDRF SaMD Working Group. (2013). "Software as a Medical Device (SaMD): Key Definitions." *IMDRF.* Retrieved from http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf

[52] FDA. (2018). "Software as a Medical Device (SaMD)." *Medical Devices.* Retrieved from

https://www.fda.gov/medicaldevices/digitalhealth/softwareasamedicaldevice/default.htm

[53] Dhar, A., Delone, M, & Ressler, D. (2018). "Reimagining Digital Health Regulation: An agile model for regulating software in health care." *Deloitte*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-ps-reimagining-digital-health-regulation.pdf

[54] 114th Congress. (2016). "21st Century Cures Act." *H.R.34*. Retrieved from https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf

[55] FDA. (2017). "Clinical and Patient Decision Support Software – Draft guidance for Industry and FDA Staff." *U.S. Department of Health and Human Services.* Retrieved from https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM587819.pdf

[56] Ibid.

[57] Ibid.

[58] FDA. (2017). "Software as a Medical Device: Clinical Evaluation." *U.S. Department of Health and Human Services.* Retrieved from https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM524904.pdf

[59] Ochs, R. (2017). "Arterys Cardio DL." *FDA*. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf16/K163253.pdf

[60] Ochs, R. (2017). "QuantX." *FDA*. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf17/DEN170022.pdf

[61] Ochs, R. (2018). "Viz.AI ContaCT." *FDA*. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf17/DEN170073.pdf

[62] Ochs, R. (2018). "Viz CTP." *FDA*. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf18/K180161.pdf

[63] Krueger, A.C. (2018). "IDx-DR." *FDA*. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180001.pdf

[64] Ochs, R. (2018). "OsteoDetect." *FDA*.

[65] Krueger, A.C. (2018). "ECG App." *FDA*. Retrieved from https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180044.pdf

[66] Ochs, R. (2018). "Accipiolx." *FDA.*

[67] FDA. (2017). "Digital Health Innovation Action Plan." Retrieved from https://www.fda.gov/downloads/medicaldevices/digitalhealth/ucm568735.pdf

[68] FDA. (2018). "Digital Health Software Precertification (Pre-Cert) Program. *Medical Devices.* Retrieved from https://www.fda.gov/medicaldevices/digitalhealth/digitalhealthprecertprogram/default.htm

[69] FDA. (2018). "Developing Software Precertification Program: A Working Model." *U.S. Department of Health and Human Services.* Retrieved from https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM611103.pdf

[70] FDA. (2018). "Developing Software Precertification Program: A Working Model." *U.S. Department of Health and Human Services.*

[71] FDA. (2019). "Developing Software Precertification Program: A Working Model (v1.0)." *U.S. Department of Health and Human Services.* Retrieved from https://www.fda.gov/downloads/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/UCM629276.pdf

[72] FDA. (2018). "Laboratory Developed Tests." *Medical Devices.* Retrieved from https://www.fda.gov/medicaldevices/productsandmedicalprocedures/invitrodiagnostics/laboratorydevelopedtests/default.htm

[73] Clement, P.D., & Tribe, L.H. (2015). "Laboratory Testing Services, As the Practice of Medicine, Cannot Be Regulated As Medical Devices." *American Clinical Laboratory Association.* Retrieved from https://www.acla.com/wp-content/uploads/2015/01/Tribe-Clement-White-Paper-1-6-15.pdf

[74] Sharfstein, J. (2015). "FDA Regulation of Laboratory-Developed Diagnostic Tests." *JAMA*. Retrieved from https://jamanetwork.com/journals/jama/fullarticle/2089188

[75] HealthIT.gov. (2018). "Privacy, Security, and HIPAA." *ONC*. Retrieved from https://www.healthit.gov/topic/privacy-security-and-hipaa

[76] HHS. (2003). "Summary of the HIPAA Privacy Rule." *OCR Privacy Brief.* Retrieved from https://www.hhs.gov/sites/default/files/privacysummary.pdf?language=es.

[77] FDA. (2017). "Discussion Paper on Laboratory Developed Tests (LDTs)." *U.S. Department of Health and Human Services.* Retrieved from https://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/LaboratoryDevelopedTests/UCM536965.pdf

[78] HHS. (2016). "Individuals' Right under HIPAA to Access their Health Information 45 CFR § 164.524." *Health Information Privacy.* Retrieved from https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html

[79] Price II, W.N. (2017). "Artificial Intelligence in Health Care: Applications and Legal Issues." *ABA SciTech Law*. Retrieved from https://repository.law.umich.edu/cgi/viewcontent.cgi?article=2932&context=articles

[80] HHS. (2003). "Summary of the HIPAA Privacy Rule." *OCR Privacy Brief.*

[81] Swanson, A., & Khan, F. (2012). "The Legal Challenge of Incorporating Artificial Intelligence into Medical Practice." *Journal of Health & Life Sciences Law.*

[82] HIPAA Journal. (2018). "De-identification of Protected Health Information: How to Anonymize PHI." Retrieved from https://www.hipaajournal.com/de-identification-protected-health-information/

[83] FDA. (2017). "IRB Waiver or Alteration of Informed Consent for Clinical Investigations Involving No More Than Minimal Risk to Human Subjects." *U.S. Department of Health and Human Services.* Retrieved from https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM566948.pdf

[84] HHS. (2017). "The Security Rule." *Health Information Privacy.* Retrieved from https://www.hhs.gov/hipaa/for-professionals/security/index.html

[85] Wachter, S., et al. (2018). "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR." *Harvard Journal of Law & Technology*.

[86] Edwards, L., & Veale, M. (2017). "Slave to the Algorithm? Why A 'Right to an Explanation' Is Probably Not the Remedy You Are Looking for." *Duke Law & Technology Review*.

[87] Thelisson, E. (2017). "Toward Trust, Transparency, and Liability in AI/AS Systems." *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence.* Retrieved from https://www.ijcai.org/proceedings/2017/0767.pdf

[88] EU GDPR. (2018). "Article 22." *PrivazyPlan.* Retrieved from http://www.privacy-regulation.eu/en/article-22-automated-individual-decision-making-including-profiling-GDPR.htm

[89] Clancy, M. (2017). "Intellectual Property Law—The Future of Patent Eligibility Analysis on Medical Diagnostics and Its Effects on Healthcare Innovation— Ariosa Diagnostics, Inc. v. Sequenom, Inc." *Journal of Health & Biomedical Law.*

[90] HHS. (2003). "Summary of the HIPAA Privacy Rule." *OCR Privacy Brief.*

[91] Electronic Frontier Foundation. (n.d.). "The Law and Medical Privacy." Retrieved from https://www.eff.org/issues/law-and-medical-privacy

[92] United States Court of Appeals, District of Columbia Circuit. (2015). "POM Wonderful, LLC. v. FTC." *Trade Cases.*

[93] FDA. (1971). "Memorandum of Understanding Between The Federal Trade Commission and The Food and Drug Administration." *MOU 225-71-8003.* Retrieved from https://www.fda.gov/AboutFDA/PartnershipsCollaborations/MemorandaofUnderstandingMOUs/DomesticMOUs/ucm115791.htm

[94] Cortez, N. (2017). "Substantiating Big Data in Health Care." *I/S: A Journal of Law and Policy for the Information Society.*

[95] Bambauer, J.R. (2017). "Dr. Robot." *U.C. Davis Law Review.*

[96] FTC. (2015). "'Melanoma Detection' App Sellers Barred from Making Deceptive Health Claims." *Press Release.* Retrieved from https://www.ftc.gov/news-events/press-releases/2015/08/melanoma-detection-app-sellers-barred-making-deceptive-health

[97] Ford, R.A., & Price II, W.N. (2016). "Privacy and Accountability in Black-Box Medicine." *Michigan Telecommunications & Technology Law Review*.

[98] 132 S.Ct. 1289 Supreme Court of the United States. (2012). "Mayo Collaborative Services. v. Prometheus Labs, Inc."

[99] 134 S.Ct. 2347 Supreme Court of the United States. (2014). "Alice Corp. Pty. v. CLS Bank International."

[100] Tull, S.Y. (2018). "Patenting the Future of Medicine: The Intersection of Patent Law and Artificial Intelligence in Medicine." *Landslide*. Retrieved from https://www.finnegan.com/en/insights/patenting-the-future-of-medicine-the-intersection-of-patent-law-and-artificial-intelligence-in-medicine.html

[101] Tsang, L., Kracov, D.A., Mulryne, J., et al. (2017). "The Impact of Artificial Intelligence on Medical Innovation in the European Union and United States." *Intellectual Property & Technology Law Journal.*

[102] Ibid.

[103] Rai, A. (2018). "Intellectual Property in Precision Medicine." *Duke Law School*. Retrieved from https://precisionmedicine.duke.edu/sites/precisionmedicine.duke.edu/files/field/attachments/GPMF_Rai_02.15.2018.pdf

[104] Wong, A., et al. (2017). "Comparison of Overridden Medication-related Clinical Decision Support in the Intensive Care Unit between a Commercial System and a Legacy System." *Applied Clinical Informatics.* Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28832067

[105] FDA. (2019). "Developing Software Precertification Program: A Working Model (v1.0)." *U.S. Department of Health and Human Services.*

[106] U.S. Court of Appeals for the Federal Circuit. (2011). "Mayo Collaborative Services v. Prometheus Laboratories, Inc." *Supreme Court of the United States.* Retrieved from https://www.supremecourt.gov/opinions/11pdf/10-1150.pdf

[107] U.S. Court of Appeals for the Federal Circuit. (2013). "Alice Corporation v. CLS Bank International, et al." *Supreme Court of the United States.* Retrieved from https://www.supremecourt.gov/opinions/13pdf/13-298_7lh8.pdf

[108] FDA. (2019). "Developing Software Precertification Program: A Working Model (v1.0)." *U.S. Department of Health and Human Services.*

[109] Xavier Health. (2018). "Perspectives and Good Practices for AI and Continuously Learning Systems in Healthcare." Retrieved from https://www.xavierhealth.org/cls-working-team/

[110] Abdul, A.M., et al. (2018). "Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda." *CHI*. Retrieved from https://www.semanticscholar.org/paper/Trends-and-Trajectories-for-Explainable%2C-and-An-HCI-Abdul-Vermeulen/2a0e4eeea84f6352a6ac012f10581e5a228b6a92

[111] Chen, C., et al. (2018). "This Looks Like That: Deep Learning for Interpretable Image Recognition." Retrieved from https://arxiv.org/pdf/1806.10574.pdf

[112] Gilpin, L.H., et al. (2018). "Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning." Retrieved from https://arxiv.org/pdf/1806.00069.pdf

[113] Stanford University. (2016). "Artificial Intelligence and Life in 2030." Retrieved from https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf

[114] Price II, W.N. (2018). "Medical Malpractice and Black-Box Medicine." *Big Data, Health Law, and Bioethics.* Retrieved from https://doi.org/10.1017/9781108147972.027

[115] Ibid.

[116] Scherer, M.U. (2016). "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies." *Harvard Journal of Law & Technology.*

[117] Vladeck, D.C. (2014). "Machines without Principals: Liability Rules and Artificial Intelligence." *Washington Law Review*.

[118] Korosec, K. (2015). "Volvo CEO: We will accept all liability when our cars are in autonomous mode." *Fortune.* Retrieved from http://fortune.com/2015/10/07/volvo-liability-self-driving-cars/

[119] Abramoff, M.D. (2018). "Panel: Understanding Algorithms, Artificial Intelligence, and Predictive Analytics Through Real World Applications." *FTC Hearing #7, on Algorithms, Artificial Intelligence and Predictive Analytics.* Retrieved from https://www.ftc.gov/system/files/documents/public_comments/2018/11/ftc-2018-0101-d-0004-162932.pdf

[120] Supreme Court of the United States. (2008). "Riegel v. Medtronic, Inc." *Legal Information Institute.* Retrieved from https://www.law.cornell.edu/supct/html/06-179.ZS.html

[121] Supreme Court of the United States. (1996). "Medtronic, Inc. v. LOHR et vir." *Legal Information Institute.* Retrieved from https://www.law.cornell.edu/supct/html/95-754.ZS.html

[122] FDA. (2017). "Deciding When to Submit a 510(k) for a Software Change to an Existing Device." *U.S. Department of Health and Human Services.* Retrieved from

https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM514737.pdf

[123] FDA. (2018). "Precertification (Pre-Cert) Pilot Program: Frequently Asked Questions - #20." *Medical Devices.* Retrieved from
https://www.fda.gov/MedicalDevices/DigitalHealth/DigitalHealthPreCertProgram/ucm577330.htm

[124] Price II, W.N. (2017). "Regulating Black-Box Medicine." *Michigan Law Review.* Retrieved from
http://michiganlawreview.org/wp-content/uploads/2017/12/116MichLRev421_Price.pdf

[125] Lum, K. (2017). "Limitations of mitigating judicial bias with machine learning." *Nature Human Behavior.* Retrieved from https://www.nature.com/articles/s41562-017-0141

[126] JASON. (2017). "Artificial Intelligence for Health and Health Care." *The MITRE Corporation.* Retrieved from https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf

[127] Zech, J.R., et al. (2018). "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study." *PLOS Medicine.* Retrieved from
https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002683#sec007

[128] Ford, R.A., & Price II, W.N. (2016). "Privacy and Accountability in Black-Box Medicine." *Michigan Telecommunications and Technology Law Review.* Retrieved from
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758121

[129] Suleyman, M., & King, D. (2018). "The Information Commissioner, the Royal Free, and what we've learned." *DeepMind.* Retrieved from https://deepmind.com/blog/ico-royal-free/

[130] Guo, S. (2015). "Learning on private data with homomorphic evaluation encryption and differential privacy." *University of Buffalo*. Retrieved from https://ubir.buffalo.edu/xmlui/handle/10477/51335

[131] UK Anonymisation Network. (2018). Retrieved from http://ukanon.net/

[132] Suleyman, M., & Laurie, B. (2018). "Trust, confidence and Verifiable Data Audit." *DeepMind.* Retrieved from https://deepmind.com/blog/trust-confidence-verifiable-data-audit/

[133] Department of Health and Social Care. (2018). "Code of conduct for data-driven health and care technology." *Gov.UK.* Retrieved from https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology

[134] FDA. (2018). "Content of Premarket Submissions for Management of Cybersecurity in Medical Devices – Draft Guidance for Industry and FDA Staff." *U.S. Department of Health and Human Services.* Retrieved from *https*://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM623529.pdf?utm_campaign=FDA%20Releases%20Draft%20Recom%20on%20Premrkt%20Sub%20Manag%20Cybersecurity&utm_medium=email&utm_source=Eloqua

[135] Tan, C. (2012). "Artificial Intelligence: Will Computers Pass the Turing Test by 2029? Does it Matter?" *Harvard University.* Retrieved from http://sitn.hms.harvard.edu/flash/2012/ai/

[136] Office of Civil Rights (OCR). (2013). "Business Associates." *Health Information Privacy.* Retrieved from https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/business-associates/index.html

[137] HealthIT.gov. (2018). "Clinical Decision Support."

[138] https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html

[139] Puget, J.F. (2016). "What is Machine Learning?" *IBM Community.* Retrieved from
https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learning?lang=en

[140] LaMorte, W.W. (2016). "Positive and Negative Predictive Value." *Boston University School of Public Health.* Retrieved from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_Screening/EP713_Screening5.html

[141] Ibid.

[142] IMDRF SaMD Working Group. (2013). "Software as a Medical Device (SaMD): Key Definitions." *IMDRF.*

[143] Dhar, A., Delone, M., & Ressler, D. (2018). "Reimagining Digital Health Regulation: An agile model for regulating software in health care." *Deloitte*. Retrieved from
https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-ps-reimagining-digital-health-regulation.pdf

[144] FDA. (2017). "Changes to Existing Medical Software Policies Resulting from Section 3060 of the 21st Century Cures Act – Draft Guidance for Industry and FDA Staff." *U.S. Department of Health and Human Services.* Retrieved from

https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM587820.pdf

[145] FDA. (2016). "Examples of Mobile Apps for which the FDA will Exercise Enforcement Discretion." *Medical Devices.* Retrieved from
https://www.fda.gov/MedicalDevices/DigitalHealth/MobileMedicalApplications/ucm368744.htm

[146] FDA. (2015). "Medical Device Data Systems, Medical Image Storage Devices, and Medical Image Communications Devices – Guidance for Industry and FDA Staff." *U.S. Department of Health and Human Services.* Retrieved from
https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm401996.pdf

[147] FDA. (2018). "Clinical Laboratory Improvement Amendments (CLIA)." *Medical Devices.* Retrieved from
https://www.fda.gov/medicaldevices/deviceregulationandguidance/ivdregulatoryassistance/ucm124105.htm

[148] FDA. (2018). "Clinical Laboratory Improvement Amendments (CLIA)." *Medical Devices.*

[149] FDA. (2018). "General Controls for Medical Devices." *Medical Devices.* Retrieved from
https://www.fda.gov/medicaldevices/deviceregulationandguidance/overview/generalandspecialcontrols/ucm055910.htm#application_of_provisions

[150] FDA. (2018). "Regulatory Controls." *Medical Devices.* Retrieved from
https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/GeneralandSpecialControls/default.htm

[151] FDA. (2017). "Learn if a Medical Device Has Been Cleared by FDA for Marketing." *Medical Devices.* Retrieved from https://www.fda.gov/medicaldevices/resourcesforyou/consumers/ucm142523.htm

[152] Greenemeier, L. (2017). "20 Years after Deep Blue: How AI Has Advanced Since Conquering Chess." *Scientific American.* Retrieved from https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/

[153] Samuel, A.L. (1959), "Some Studies in Machine Learning Using the Game of Checkers." *IBM Journal of Research and Development.* Retrieved from https://ieeexplore.ieee.org/abstract/document/5392560

[154] Microsoft Azure. (2017). "Machine learning algorithm cheat sheet for Azure Machine Learning Studio." Retrieved from https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-cheat-sheet#notes-and-terminology-definitions-for-the-machine-learning-studio-algorithm-cheat-sheet