JAMES L. MADARA, MD
EXECUTIVE VICE PRESIDENT, CEO

ama-assn.org
t (312) 464-5000

June 3, 2019

The Honorable Norman E. Sharpless, MD
Acting Commissioner
U.S. Food and Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993

Re: Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning
(AI/ML)-Based Software as a Medical Device (SaMD), Discussion Paper and Request for Feedback

Dear Acting Commissioner Sharpless:

On behalf of the physician and medical student members of the American Medical Association (AMA), I
appreciate the opportunity to provide comments on the *Proposed Regulatory Framework for
Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device
(SaMD), Discussion Paper and Request for Feedback* (Discussion Paper). As part of the AMA's prior
comments on the proposed SaMD Precertification Program, the AMA urged the U.S. Food and Drug
Administration (FDA) to include a separate framework for AI enabled systems. We applaud this step
forward and we would welcome the opportunity to meet with you and other stakeholders to address the
key issues we have outlined below. In brief, we urged the FDA to establish accepted nomenclature and
definitions and appropriate risk stratification for AI systems. While the Discussion Paper provides an
outline of key topics, it does not provide standard nomenclature and continues to utilize the International
Medical Device Regulators Forum (IMDRF) risk table without modification, which does not account for
the added dimensions of risk that ML systems in particular may present. We urge the FDA to prioritize
addressing these two items as well as the comments we have outlined below.

In June 2018, the AMA's House of Delegates adopted policy on healthcare augmented intelligence.[1] The
policy provides that the AMA will leverage its ongoing engagement in digital health and other priority
areas for improving patient outcomes and physicians' professional satisfaction to help set priorities for
health care AI. Consistent with the policy, the AMA continues to identify opportunities to integrate the
perspective of practicing physicians into the development, design, validation, and implementation of
health care AI. Furthermore, the policy provides that the AMA will promote development of thoughtfully
designed, high-quality, clinically validated health care AI that:

- Is designed and evaluated in keeping with best practices in user-centered design,
particularly for physicians and other members of the health care team;
- Is transparent;

---

[1] The term artificial intelligence and augmented intelligence are utilized interchangeably by the AMA. However, the
term augmented intelligence was adopted by the AMA's House of Delegates as it more accurately reflects that
machines should be designed to complement humans and scale human capacity. Assistive and autonomous AI are
considered subcategories augmented intelligence.

- Conforms to leading standards for reproducibility;
- Identifies and takes steps to address bias and avoids introducing or exacerbating health care disparities including when testing or deploying new AI tools on vulnerable populations; and
- Safeguards patients' and other individuals' privacy interests and preserves the security and integrity of personal information.

Finally, AMA policy provides that we are to advocate for appropriate professional and governmental oversight for safe, effective, and equitable use of and access to health care AI. To that end, we would stress that all of these should be directly addressed by the FDA as priority areas for developers as these represent a clear statement of what a key end-user require as essential for the adoption of AI systems—particularly ML systems in clinical practice.

AMA Comment: Clarify if this Discussion Paper is Specifically Addressing ML

We urge the FDA to specify that this Discussion Paper exclusively addresses ML systems, as opposed to the full gamut of AI systems. The Discussion Paper is not clear in this regard and we are concerned that it will create confusion where there is already significant complexity. We discuss the need for standard terms and nomenclature below which would also provide helpful clarification of the scope of this document.

AMA Comment: Provide Appropriate Balance concerning Benefits and Risk

We, like the FDA, believe that AI systems generally and ML systems in particular may serve as transformative tools that will aid this nation and others around the globe in addressing increasing stress on health care systems at a time when human and financial resources are not likely to keep pace with demand. However, we strongly urge the FDA to balance the positive language exemplified in the introduction and interspersed throughout the document with appropriate identification of risk. This Discussion Paper selectively highlights the benefits of ML systems specifically and minimizes or fails to mention the risks unique to ML systems deployed for clinical applications. As a regulatory document, it is essential to be balanced. For example, the risk of bias in various forms are well-known pitfalls associated with ML, yet nowhere does the Discussion Paper mention the word "bias" or how the FDA intends to ensure ML systems are equitable. Notably, the introduction highlights the importance and value of "vast amounts of data generated during the delivery of health care every day" as an unassailable good without noting potential risks. Yet, this data utilized by ML for clinical applications also presents risks that should be addressed in a regulatory document. Enclosed is the report that was also considered at the time that AMA's policy was adopted in June 2018. It provides critical context as follows:

> There is a popular tendency to see AI as, at best, a form of neutral, "objective" decision making, a pristine mathematical process that takes only "the facts" into account, independent of human judgment [15,16,17]. The statistical process of AI specifically seeks to derive a rule or procedure from a body of data that explains that data or is able to predict future data [18]. An AI derived algorithm "is only as good as the data it works with" [19,20]. The data sets on which AI algorithms are trained are created by human agents and are imperfect. The research, patient care, and insurance records available as training data sets for health care AI can be highly variable, reflecting the different purposes for and processes by which

they were created [1,21]. Clinical trials systematically include or exclude participants with certain characteristics; patient charts and insurance records capture information only from those individuals who have access to the health care system and rarely contain information about exposure to environmental toxins. Different data sets focus on different kinds of information to the exclusion of other possible data points, and records capture and preserve information with varying degrees of accuracy. One of the most significant implications for end users of AI systems is that these systems sets can, invisibly and unintentionally, "reproduce and normalize" the biases of their training data sets [16,17]. In health care, the result can be models that "reflect the conditions only of the fortunate" and yield "an aggregate understanding of health and illness that fundamentally excludes the marginalized" [21] in a way that risks exacerbating existing health disparities.

We urge you to consider the remainder of the AMA report that is enclosed. We also strongly recommend that the FDA ensure that future Agency discussion documents and guidance fully reflect the reality of risk and benefits inherent to ML. In addition, the AMA's House of Delegates will be considering additional AI policy the second week in June. If this policy is adopted, it will include additional important elements that we would like to more fully elaborate on as an addendum to these comments.

AMA Comment: Developer Goals Tied to Patient Outcomes

A fundamental principle missing throughout the document that the AMA strongly urges the FDA to incorporate as a foundational requirement: tying developer goals to patient outcomes. We appreciate that the Agency acknowledges that ML systems involve learning from data and hence present unique considerations in addition to SaMD generally. To address these, the Agency outlines good machine learning practices (GMLP) best practices (e.g., data management, feature extraction, training, and evaluation). GMLP considerations include:

- Relevance of available data to the clinical problem and current clinical practice;
- Data acquired in a consistent, clinically relevant and generalizable manner that aligns with the SaMD's intended use and modification plans;
- Appropriate separation between training, tuning, and test datasets; and
- Appropriate level of transparency (clarity) of the output and the algorithm aimed at users.

However, the foregoing do not include references to patient outcomes. For example, in addition to the data training risks described in the AMA policy, ML systems may learn to detect features that are closely associated with a diagnosis, but that are divorced from improvements in clinical outcome. Mirroring a diagnosis or a clinical practice itself does not ensure beneficial patient outcomes. A framework that does not tie clinical outcomes research to GMLP could lead to AI systems that deviate from outcome-based clinical standards. Similarly, the Algorithm Change Protocol includes important components, but again fails to reference or tie these provisions to patient outcomes. The risks associated with deviating from outcomes-based clinical research could be more pronounced for ML. The overall framework offered in the Discussion Paper should be reframed to ensure this is incorporated and consistently addressed.

AMA Comment: Standardize Nomenclature and Terminology

The AMA strongly urges that the FDA work with the National Institute of Standards and Technology (NIST) to convene standards setting bodies, the AMA and national medical specialty societies along with think tanks and companies committed to working on standardizing the nomenclature for health care AI in order to drive consensus on shared definitions. The definitions offered by the FDA for ML locked models and continuous learning systems do not seem to comport with the definitions offered by experts who the AMA has consulted extensively over the past year. In addition, the Discussion Paper does not consistently use terms either. It is not evident if the FDA is discussing discontinuous learning systems (also referred to as batched learning ML systems) or continuous learning systems that are being deployed directly into clinical care. This confusion is exacerbated by the following excerpt, but is not limited to this section of the Discussion Paper:

> To date, FDA has cleared or approved several AI/ML-based SaMD. <u>Typically</u>, these have only included algorithms that are "locked" prior to marketing, where algorithm changes likely require FDA premarket review for changes beyond the original market authorization. However, not all AI/ML-based SaMD are locked; some algorithms can adapt over time. The power of these AI/ML-based SaMD lies within the ability to continuously learn, where the adaptation or change to the algorithm is realized after the SaMD is distributed for use and has "learned" from real-world experience. Following distribution, these types of continuously learning and adaptive AI/ML algorithms may provide a different output in comparison to the output initially cleared for a given set of inputs. (Emphasis added.)

The first question this raises concerns the word "typically." Are there any continuous learning systems that are currently deployed that have received FDA approval, clearance, or *de novo* authorization? If so, we urge the FDA to identify these as important use cases that can be carefully considered. Second, we need to understand if this constitutes human subject research. Also, this characterization of ML systems assumes that the system may improve over time, but the opposite, that it becomes less optimized as a result of continuous learning process could also happen. This must be adequately addressed.

This section also makes reference to:

> The highly iterative, <u>autonomous</u>, and adaptive nature of these tools requires a new, total product lifecycle (TPLC) regulatory approach that facilitates a rapid cycle of product improvement and allows these devices to continually improve while providing effective safeguards. (Emphasis added.)

This conflates two concepts—one related to whether the system is a continuous learning system and the other the autonomy of the system. There are locked ML models that are autonomous (for example, the IDx-DR system). And, there could be continuous learning systems that are assistive (for example, clinical decision support to identify sepsis risk).

We again recommend that the FDA utilize terms and definitions contained in the <u>Current State and Near-term Priorities for AI Enable Diagnostic Support Software in Health Care</u> report issued by the Duke Margolis Center for Health Policy. In addition, the AMA is engaged in standards AI workgroups of the

Association for the Advancement of Medical Instrumentation and the British Standards Institute, the Consumer Technology Association, and IEEE. Standards bodies have varied definitions currently, but it is clear these will need to be fine-tuned. And, the updates should result in consistent terms and definitions across standards bodies. We also urge the FDA to take a leading role to facilitate shared understanding of the terms: transparency, explainability, bias, reproducibility, assistive, autonomous, and automated among other terms. We have found it common that stakeholder misunderstandings are rooted in not having shared understanding of basic terms.

AMA Comment: IMDRF Does Not Address Dimensions of Risk of ML

The AMA has observed in our prior comments on the proposed SaMD Precertification Program that the IMDRF was not developed to consider the added dimensions of risk presented by ML. This is particularly the case if the FDA intends to develop a pathway for continuous learning systems that do not utilize batch learning (which would not allow for validation). IMDRF also does not account for whether a system is autonomous with a human able to intervene, fully autonomous without the ability of a human to intervene, or assistive where a human is required. In addition, the risks associated with a system that continuously learns in the field is another dimension of risk that is not accounted for in the IMDRF risk categorization table. While we understand the desire to advance international harmonization, the IMDRF must be modified to address these issues as these dimensions present clear and unambiguous elements of risk not currently contemplated by the IMDRF framework.

AMA Comment: Labeling–A Number of Options

Key stakeholders have presented labeling suggestions for AI systems that we are actively considering. An example is provided below. We have not completed our review and consideration of the options presented. However, this work will be underway over the next several months as we consider how this should harmonize for labeling in the context of SaMD that does not include AI/ML. We urge the FDA to carefully consider that it appears that the elements will need to be adapted for AI/ML, but should be consistent.

Augmented Intelligence Facts

IDx-DR

AI Description	
Autonomous *	**Diagnostic ***
Disease	Diabetic Retinopathy, inclusive of macular edema
Intended for	Adults with diabetes
Prescribers	Healthcare professional
Operator qualifications	Self-certification with provided training program
Companion equipment	Topcon NW400

AI Design		
Continuous learning *: No	Gag order *: No	
Clinical Standard of Care	American Academy of Ophthalmology	2017 Diabetic Retinopathy PPP

AI Performance Data		
Initial Regulated Release	Study	Level I*
	Reference Standard	RS Level A*
	Risks of Bias	I – 0; II – 0; III – 0: IV – 0*
	Record	www.nature.com/articles/s41746-018-0040-6

Safety*	Efficiency*	Equity*
Sensitivity **87.2%**	Specificity **90.7%**	Diagnosability **96.0%**
No significant differences by race, ethnicity, or age	No significant differences by race, ethnicity	No significant differences by race, ethnicity, or age

Current Release	*Initial release is current.*

AI Regulatory Status
Regulated by FDA

Class: II[1]	Reference: DEN180001[1]	Rx Only: Yes[1]

) Augmented Intelligence Guidelines 201
[1] FDA CFR 21

Conclusion–Ongoing Dialogue

The AMA appreciates the FDA's engagement with stakeholders. We would welcome meeting with you along with a select number of experts that the AMA has had ongoing intensive discussions over the past year on a host of health care AI issues. Please contact Shannon Curtis, Assistant Director, Division of Federal Affair, Shannon.Curtis@ama-assn.org or (202) 789-8510.

Sincerely,
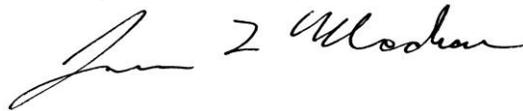
James L. Madara, MD

Enclosure

Augmented intelligence in health care*

Interest in augmented intelligence (AI) and its potential to dramatically impact medicine is growing rapidly among Congress, federal agencies, and other health care stakeholders. As a leader in American medicine, our American Medical Association (AMA) is uniquely positioned to ensure that the evolution of AI in medicine benefits patients, physicians, and the health care community. This report contains baseline policy to guide AMA's engagement with a broad cross-section of stakeholders and policymakers to ensure that the perspective of physicians in various practice settings informs and influences the dialogue as this technology develops.

Ensuring the appropriate implementation of AI in health care will require that stakeholders forthrightly address challenges in the design, evaluation, implementation, and oversight of AI systems. Through its strategic partnerships and collaborations, the AMA has the capacity to help set priorities for health care AI; integrate the perspective of practicing physicians into the design, validation, and implementation of high-quality, clinically valuable health care AI; and promote greater understanding of the promise and limitations of AI across the health care community. A strong tradition of advocacy well positions our AMA to explore the legal implications of the emerging technologies of AI in health care and advocate effectively for appropriate professional and governmental oversight for safe, effective, equitable use of and access to health care AI.

AMA policy

As a leader in American medicine, our American Medical Association (AMA) has a unique opportunity to ensure that the evolution of augmented intelligence (AI) in medicine benefits patients, physicians, and the health care community. To that end our AMA will seek to:

- Leverage its ongoing engagement in digital health and other priority areas for improving patient outcomes and physicians' professional satisfaction to help set priorities for health care AI.

- Identify opportunities to integrate the perspective of practicing physicians into the development, design, validation, and implementation of health care AI.

- Promote development of thoughtfully designed, high-quality, clinically validated health care AI that:

    — is designed and evaluated in keeping with best practices in user-centered design, particularly for physicians and other members of the health care team;
    — is transparent;
    — conforms to leading standards for reproducibility;
    — identifies and takes steps to address bias and avoids introducing or exacerbating health care disparities including when testing or deploying new AI tools on vulnerable populations; and
    — safeguards patients' and other individuals' privacy interests and preserves the security and integrity of personal information.

- Encourage education for patients, physicians, medical students, other health care professionals, and health administrators to promote greater understanding of the promise and limitations of health care AI.

- Explore the legal implications of health care AI, such as issues of liability or intellectual property, and advocate for appropriate professional and governmental oversight for safe, effective, and equitable use of and access to health care AI.

---

* Content derived from Augmented Intelligence (AI) in Health Care (Annual Meeting 2018)

What is health care AI?

Computational methods and techniques for data analysis have been evolving for decades. [1,2] A number of these methods have come to be known collectively as "artificial intelligence." Artificial intelligence constitutes a host of computational methods that produce systems that perform tasks normally requiring human intelligence. These computational methods include, but are not limited to, machine image recognition, natural language processing, and machine learning. However, in health care a more appropriate term is "augmented intelligence" (AI), reflecting the enhanced capabilities of human clinical decision making when coupled with these computational methods and systems.

In December 2017, Senators Maria Cantwell (D-WA), Todd C. Young (R-IN), and Edward Markey (D-MA) and U.S. Representatives John Delaney (D-MD) and Pete Olson (R-TX) introduced S. 2217/H.R. 4625, "Fundamentally Understanding the Usability and Realistic Evolution (FUTURE) of Artificial Intelligence Act of 2017." The legislation defines "general AI" as computational methods that produce systems that exhibit intelligent behavior at least as advanced as a human across the range of cognitive, emotional, and social behaviors. In contrast, the bill defines the term "narrow AI" as computational methods that address specific application areas, such as playing strategic games, language translation, self-driving vehicles, and image recognition. Thus, these AI methods and tools for the foreseeable future are better characterized as narrow AI that augments human intelligence (augmented intelligence).

At a February 2018 U.S. House of Representatives Government Oversight Committee Subcommittee on Information Technology hearing, three national experts testified that general AI is decades away. Consistent with the foregoing, in response to a 2016 Request for Information on Artificial Intelligence issued by the White House Office of Science and Technology Policy, a technology company stated that it is "guided by the term 'augmented intelligence' rather than 'artificial intelligence'" and noted further that "[i]t is the critical difference between systems that enhance and scale human expertise rather than those that attempt to replicate all of human intelligence." [3]

Software algorithms developed using these evolving methods and techniques, coupled with proliferating sources of data (datasets) pertinent to health and medicine, offer the promise of new and more powerful ways to augment human intelligence and expertise in health care.

The American College of Radiology (ACR), which has been at the leading edge of health care AI, addressed its promise in comments to the White House Office of Science and Technology Policy in 2016:

> AI could offer various benefits to medical imaging in the future, including augmenting the capabilities of radiologists to enhance their efficiency and accuracy, as well as reducing costs by improving the appropriateness and cost-effectiveness of medical imaging utilization. The use of AI and machine learning in health care in general could be best applied to the areas of precision medicine, predictive analytics, and outcomes assessments. AI can streamline health care workflow and improve triage of patients (especially in acute care settings), reduce clinician fatigue, and increase the efficiency and efficacy of training. Moreover, shortages of medical experts to meet the needs of vulnerable and underserved populations in domestic and international settings could potentially be relieved, in part, by AI [4].

Prime AI applications include clinical decision support, patient monitoring and coaching, automated devices to assist in surgery or patient care, and management of health care systems [5]. AI in health care holds out the

prospect of improving physicians' ability to establish prognosis [6], as well as the accuracy and speed of diagnosis [6,7,8], enabling population-level insights to directly inform the care of individual patients [9], and predicting patient response to interventions [10]. The number of empirical studies of AI applications in medicine is growing rapidly [2].

What's next in health care AI?

Commercial entities are driving rapid evolution in AI across the board. In health care, the next three to five years will be marked by efforts to scale AI options involving patient-centered wearables that support clinical care, improved tools for diagnosis and physician training, and health system initiatives to improve patient care and clinical decision support [11]. The following are early examples of such efforts.

Wearable AI
Wearable monitoring devices that can transmit patient data are evolving rapidly. For example, one company has developed the Cardiogram application which is designed to work with the built-in infrared heart rate sensor of the Apple Watch to detect hypertension and sleep apnea. In a study carried out with the University of California–San Francisco that involved over 6,000 patients, the application and its machine learning system, DeepHeart, was able to detect hypertension and sleep apnea with 82 percent and 90 percent accuracy, respectively [12]. Rapid innovation is expected on this front propelled by coverage of payers, including Medicare, of remote patient monitoring and management.

New tools for diagnosis and physician training
The utilization of machine learning algorithms to enhance clinical decision making is increasing, but emerging systems take such support a step further. For example, the Human Diagnosis Project (Human Dx), organized as a tandem 501(c)(3) nonprofit and public benefit corporation, and created with and led by the medical community, allows attending physicians to ask for assistance on difficult medical cases from an online community of physicians all over the world. Responses from the medical community are combined with help from machine learning to create a synthesized collective assessment for each case. This collective insight is designed to augment clinical decision making with machine intelligence, providing useful information to physicians and patients who may not otherwise have access to specialist expertise. Human Dx also provides a platform for medical education through its Global Morning Report teaching cases. Today, residents from

over 40 percent of U.S. internal medicine residency programs have access to these cases. Human Dx vets the quality of responses by comparing how physicians solve reference training cases in order to calculate a quantitative measure of reasoning called Clinical Quotient, which is now being vetted in conjunction with the Johns Hopkins School of Medicine.

Health systems and data analytics
Applying AI to health system data to improve care is another area of rapid evolution. The University of Pittsburgh Medical Center (UPMC) has launched a system-wide effort to reduce hospital readmissions and enhance clinical decision making while a patient is receiving care. UPMC has applied machine learning to claims data to predict a patient's risk of readmission before the patient arrives. A second algorithm uses laboratory and clinical metrics extracted from clinical records to update the risk prediction every 15 minutes over the course of the patient's admission. Before discharge, if the risk prediction's two models are in conflict, UPMC uses unsupervised machine learning to come up with a set of rules that dictate which model takes precedence to inform clinician discharge decisions [13].

These three relatively nascent efforts are designed to scale, but will require significant additional research and real world testing. However, they illustrate the types of initiatives beyond condition-specific efforts to enhance clinical decision support that could produce significant improvements in health care. Notably, these efforts have active engagement and support of clinicians and seek to address medical challenges and problems identified by clinicians.

Federal engagement with AI

AI has surfaced as a public policy issue at the federal level in a relatively short period of time. In 2016, the White House Office of Science and Technology hosted several public meetings on a range of public policy issues addressing AI along with a public request for information regarding potential policy directions. In Congress, the U.S. Senate Commerce Committee held a hearing titled "The Dawn of Artificial Intelligence" at which the Department Chair for Genomic Medicine at MD Anderson Cancer Center highlighted the clinical applications of AI and discussed policy implications.

Shortly thereafter, the 21st Century Cures Act was passed by Congress and became law in December 2016. The Act included provisions modifying the U.S. Food and Drug Administration's (FDA) oversight of software as a

medical device, which has implications for a number of current AI computational methods. The FDA is now actively evaluating whether a new oversight framework is needed for software as a medical device, a precursor to future oversight models.

The bipartisan "FUTURE of Artificial Intelligence Act," introduced in December 2017, provides for the establishment of a Federal Advisory Committee on the Development and Implementation of Artificial Intelligence. The legislation, if passed, would be the first effort at the federal level to provide a forum for consideration of AI public policy. In 2018, additional legislation has been introduced, and additional congressional hearings held on AI generally, with health care applications receiving particular attention.

Achieving the promise of AI in health care

Fulfilling the promise that "combining machine learning software with the best human clinician 'hardware' will permit delivery of care that outperforms what either can do alone" [14] will require that stakeholders forthrightly address challenges in the design, evaluation, implementation, and oversight of AI systems in health care. In the first instance, stakeholders across the board, not the least among them patients and physicians, must hold realistic expectations for the roles AI tools can and cannot play. Machine learning is only one of the AI computational methods and raises particularly thorny challenges. However, many of the public policy issues (including transparency and intellectual property) and clinical issues that will need to be addressed apply to other AI computational methods that are more common currently, such as natural language processing.

Designing and evaluating health care AI

There is a popular tendency to see AI as, at best, a form of neutral, "objective" decision making, a pristine mathematical process that takes only "the facts" into account, independent of human judgment [15,16,17]. The statistical process of AI specifically seeks to derive a rule or procedure from a body of data that explains that data or is able to predict future data [18]. An AI derived algorithm "is only as good as the data it works with" [19,20]. The data sets on which AI algorithms are trained are created by human agents and are imperfect.

The research, patient care, and insurance records available as training data sets for health care AI can be highly variable, reflecting the different purposes for and processes by which they were created [1,21]. Clinical trials systematically include or exclude participants with certain characteristics; patient charts and insurance records capture information only from those individuals who have access to the health care system and rarely contain information about exposure to environmental toxins. Different data sets focus on different kinds of information to the exclusion of other possible data points, and records capture and preserve information with varying degrees of accuracy.

One of the most significant implications for end users of AI systems is that these systems sets can, invisibly and unintentionally, "reproduce and normalize" the biases of their training data sets [16,17]. In health care, the result can be models that "reflect the conditions only of the fortunate" and yield "an aggregate understanding of health and illness that fundamentally excludes the marginalized" [21] in a way that risks exacerbating existing health disparities. Minority populations can be disadvantaged in the context of AI systems in a second way as well in that "by definition, there is proportionately less data available about minority
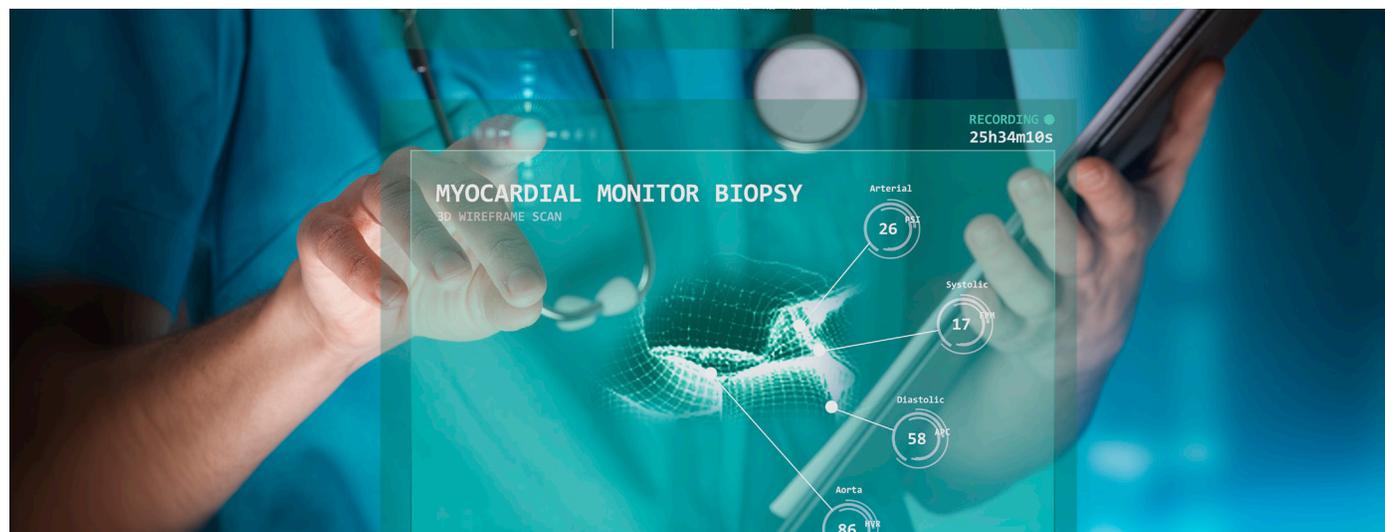
predictions," while the accuracy of decision making, a proxy for fairness, will be higher for majority groups [17]. Addressing fairness is essential, even if doing so may be costly for developers when it requires them to seek more complex decision rules [17].

Design issues also encompass how a model is evaluated, as well as relationships between the dataset used to train an algorithm and the dataset used to evaluate the algorithm. In the first instance, evaluation criteria must be clinically relevant and evaluation should be representative of how the algorithm will be applied in practice [22]. For example, evaluating a model to predict risk of hospital-acquired infection over the entire course of a patient's admission more accurately predicts how the model would be used and would perform in practice [22]. For predictive models, developers must evaluate "how far in advance the algorithm identifies positive cases." [22] From a clinician's perspective, the critical concern is "predicting events early enough for a relevant intervention to influence care decisions and outcomes." [14] Ensuring that all examples in the training dataset are earlier in time than all examples in the evaluation set helps avoid misleading results by limiting the possibility that training data could otherwise reflect structural changes in hospital population, clinical protocols, electronic health record (EHR) systems, or other factors that occurred over time [22].

Developers also have a responsibility to ensure that their work is transparent and can be reproduced by others [23,24]. Proposed guidelines for essential components of publications reporting development of predictive machine-learning algorithms include not only rationale and objectives, but, importantly, the setting, prediction problem, relevant data, and a description of the building of the predictive model [23]. Authors should also provide information about the final model and its performance, and discuss the clinical implications of the work, its limitations, and unexpected results. Scholars have further recommended creating open repositories for long-term storage, archiving, and access to datasets and code to enable replication of published findings [24].

Furthermore, the AMA's work in the area of EHRs reveals that to be useful and accepted in practice, AI systems need to be developed and evaluated in keeping with best practices in user-centered design [25]. The focus must be on users' needs and usability should be tested by participants who are demographically representative of end users [26].

Health care AI and patient privacy

Commitment to protecting the confidentiality of patient information is central to medicine's professional ethos. In this respect, AI poses a significant challenge where traditional strategies of notification and consent are no longer adequate [18]. Nor are anonymization, deletion of data, or distinguishing metadata sufficiently robust protections in the context of massive complex data sets [18,20] when machine-learning algorithms can identify a record "easily and robustly" from as few as three data points [20].

The ease of re-identification means that, in important respects, traditional expectations for health care privacy are simply no longer attainable. This significantly raises the bar on the task of ensuring the security and integrity of data. Among proposed technical solutions to the dilemma of privacy in large data sets are "blockchain-style" technology to secure data and track access or data auditing systems that allow secure verification of the contents of large data structures, such as those being explored by DeepMind Health in the UK [1]. Researchers at the University of Pennsylvania have explored the creation of publicly sharable simulated datasets that limit possible re-identification as another approach to protecting data privacy [27]. The recent revelation that the data mining firm Cambridge Analytica siphoned private data from 50 million Facebook users to target them for political campaigns raises confidentiality and privacy questions across the spectrum of digital platforms that collect and curate data. While this report establishes policy that underscores the necessity to safeguard individuals' privacy interests and preserve the security and integrity of personal information, the Board recognizes the importance of this issue and will continue to assess our policy as our AMA engages in the public debate and discourse on protecting patient information.

Implementing health care AI
The AMA's ongoing engagement with digital health offers insights for understanding, from physicians' perspectives, what is at stake in integrating AI systems into the delivery of health care. The organization's recent survey of 1,300 physicians about barriers to adoption of digital health technologies suggests that physicians are most receptive to digital health tools they believe can be integrated smoothly into their current practice, will improve care, and will enhance patient-physician relationships [28]. Coverage for liability, assurance that data privacy is protected, linkage to their EHR, and billing/reimbursement are key considerations.

Earlier AMA research into physician professional satisfaction found that frustrations with EHRs, especially
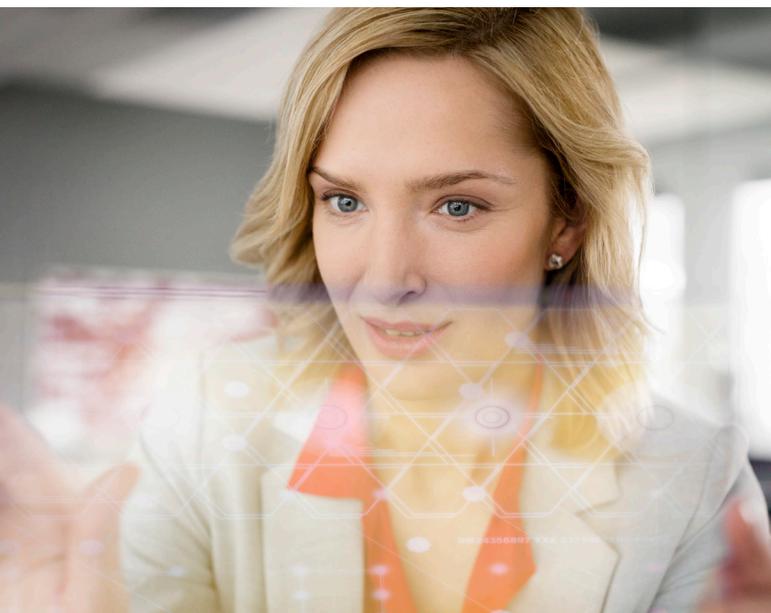
usability issues, were a major source of dissatisfaction in physicians' professional lives [29]. The findings led the AMA to identify priorities for ensuring usability in EHR systems, including, among other considerations, ensuring that EHRs are designed to meet the cognitive and workflow needs of physicians, that they support team-based care, promote coordination of care, focus on reducing cognitive workload instead of focusing simply on data collection, and incorporate end user feedback into designing and improving EHR systems [25].

AMA policies addressing the use of telemedicine similarly stress the importance of minimizing disruptive effects on patient-physician interactions, ensuring that technologies promote quality of care and safety, and, importantly, establishing mechanisms to monitor the impact of an innovation both to identify and address adverse consequences and to identify and encourage dissemination of outcomes [30,31].

To reap the benefits for patient care, physicians must have the skills to work comfortably with health care AI. Just as working effectively with EHRs is now part of training for medical students and residents [32], educating physicians to work effectively with AI systems, or more narrowly, the AI algorithms that can inform clinical care decisions, will be critical to the future of AI in health care.

Physicians need to understand AI methods and systems sufficiently to be able to trust an algorithm's predictions—or know how to assess the trustworthiness and value of an algorithm—as a foundation for clinical recommendations. The challenge may be more easily met with advances in "explainable AI," that is, algorithms

that can "explain" to users why a particular prediction is made [33,34]. Technology to predict the risk of 30-day readmission for cardiac patients being tested by Boston-based Partners Connected Health provides clinicians with a readmission prediction score and identifies the top factors contributing to that score, providing information that is actionable for clinicians [35].

A leadership role for the AMA

A component of the AMA's strategic work in 2018 and beyond has been to provide the physician perspective across health care technology sectors by promoting improved usability of and productive access to data used in medical decision making as well as respect for the patient-physician relationship. As our AMA implements this component of its strategic plan, the Board of Trustees has observed a rapidly growing interest in augmented intelligence (AI) technology in health care. In 2018, the AMA Council on Long Range Planning and Development (CLRPD) provided the Board with a primer on the history, definitions and components, and the status of AI in health care that offered a high-level look at this rapidly evolving area and its potential to dramatically impact medicine. The AMA Council on Legislation (COL) and CLRPD have observed increased interest in AI by Congress, federal agencies, and other health care stakeholders. To form a clearer understanding of the expected impact of AI technologies for patients and physicians, as well as key stakeholders who are influencing legislation and regulation in this area, the COL has met with physician experts immersed in the development and clinical integration of various health care AI technologies.

The AMA has adopted a base-level of policy on health care AI to guide AMA's engagement with a broad cross-section of stakeholders and policymakers in order to ensure that the perspective of physicians in various practice settings informs and influences the dialogue as this technology develops.

To realize its potential to support improved patient care and health outcomes and enhance physician professional satisfaction, the health care AI enterprise should be informed and guided by the expertise, experience, and leadership of physicians and organized medicine in developing and implementing these tools. Physicians are well positioned to advocate for health care AI solutions that support healthier lifestyles and reduce disease burden, improve access to care, enhance diagnostic accuracy, inform individually tailored treatment plans, and improve patient self-management,

adherence, and health outcomes. Physicians are likewise well placed to apply their experience to drive improved design and implementation of health care AI that will strengthen clinicians' relationships with patients; enhance communication among the health care team and between team members, patients, and family members; simplify the coordination of care; minimize administrative burdens; and help the health care team to better deliver care to those patients and populations in greatest need.

In addition to the work of COL and CLRPD, at the 2017 Interim Meeting all seven AMA councils met jointly with AI experts to discuss issues in health care AI. Likewise, the AMA's ongoing engagement with key stakeholders from across the spectrum of clinical care, health care administration, implementation science, and AI product development enables the organization to play a distinctive role in contributing to the overarching vision for health care
AI in the U.S.

Through its strategic partnerships and collaborations, the AMA has the capacity to offer the insight that is critical to the development of clinically sound AI systems that will enhance the quality of care and sustain the integrity of patient-physician relationships. The AMA's strong tradition of advocacy positions the organization to promote meaningful oversight of AI as it is integrated into clinical practice.

Conclusion

Patients, physicians, and the health care system in the U.S. face enormous challenges in the combined impact of a rapidly aging population, a relative decline in the working population that reduces revenue essential for safety net programs [36], and persistent high costs of care that will strain the nation's ability to support affordable, accessible, high quality care. With the engagement of physicians to identify needs and set priorities for design, development, and implementation, health care AI can offer a transformative set of tools to help patients, physicians, and the nation face these looming challenges. Given the number of stakeholders and policymakers involved in the evolution of AI in health care, it is important that our AMA not only adopt a base level of policy to guide our engagement, but equally continue to refine our policy as an organization to ensure that the perspective of physicians in various practice settings informs and influences the dialogue as this technology develops.

References

1. JASON. *Artificial Intelligence for Health and Health Care.* The MITRE Corporation, 2017. Available at https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf. Accessed February 19, 2018.

2. Chang A. AIMed: Artificial Intelligence in Medicine: Analytics, Big Data, Cloud and Cognitive Computing, Databases, and Deep Learning in Healthcare and Medicine, 2017.

3. IBM Research. *Preparing for the Future of Artificial Intelligence*, 2016. Available at http://www.research.ibm.com/cognitive-computing/ostp/rfi-response.shtml. Accessed January 30, 2018.

4. American College of Radiology. Request for information on artificial intelligence; comments from the American College of Radiology; July 22, 2016. Available at https://cqrcengage.com/acradiology/file/ywodvc80AHo/acr_comments_ostp_rfiai_7222016.pdf. Accessed February 29, 2018.

5. Stanford University. *Artificial Intelligence and Life in 2030*. September 2016. Available at https://ai100.stanford.edu/2016-report. Accessed February 6, 2018.

6. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *New Engl. J. Med* 2016;375:1216–1219.

7. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. *JAMA*. 2016;315(6):551–552.

8. Golden JA. Deep learning algorithms for detection of lymph node metastases from breast cancer: helping artificial intelligence be seen. *JAMA*. 2017;318(22):2184–2186.

9. Mandl KD, Bourgeois FT. The evolution of patient diagnosis from art to digital data-driven science. *JAMA*. 2017;318(19):1859–1860.

10. Redlich R, Opel N, Grotegerd D, et al. Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry*. 2016;73(6):557–564.

11. Council on Long-Range Planning and Development. *A Primer on Artificial and Augmented Intelligence.* Memorandum to the Board of Trustees, February 2018.

12. Tison HG, Singh AC, Chashi DA, et al. Cardiovascular risk stratification using off-the-shelf wearables and a multi-task deep learning algorithm. Presentation to the AHA Science Session, Anaheim, California. November 13, 2017

13. McCann E. UPMC's big data approach slashes readmissions. *Healthcare IT News* 2015;October 1. Available at http://www.healthcareitnews.com/news/upmcs-big-data-approach-slashes-readmissions. Accessed February 6, 2018.

14. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med* 2017;376:2507–2509.

15. Bogost I. The cathedral of computation. *Atlantic* 2015: January 15.

16. Burrell J. How the machine "thinks": understanding opacity in machine learning algorithms. *Big Data & Society* 2016;January-June:1–12.

17. Hardt M. How big data is unfair. *Medium* 2014; September: 26.

18. Executive Office of the President. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights.* 2016.

19. Barocas S, Selbst AD. Big data's disparate impact. *California Law Review* 2016;104:671–732.

20. Osoba O, Welser IV, W. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence.* 2017. Available at https://www.rand.org/pubs/research_reports/RR1744.html. Accessed February 19, 2018.

21. AI Now Institute. *The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term.* 2016. Available at https://ainowinstitute.org/reports.html. Accessed January 26, 2018.

22. Wiens J, Guttag J, Horvitz E. Patient risk stratification with time-varying parameters: a multitask learning approach. *J Machine Learning Research* 2016;17:1–23.

23. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary review. *J Med Internet Research* 2016;18(12):e323.

24. Stodden V. Enhancing reproducibility for computational methods. Presentation, National Academies of Science, Engineering, and Medicine, July 20, 2017.

25. American Medical Association. *Improving Care: Priorities to Improve Electronic Health Records.* Available at https://www.ama-assn.org/sites/default/files/media-browser/member/about-ama/ehr-priorities.pdf. Accessed January 30, 2018.

26. MedStar Health. *Why We Created the EHR User-Centered Design Evaluation Framework*. Available at https://mi2.medstarhealth.org/human-factors/research/ehr-user-centered-design-evaluation-framework/methodology/#q={}. Accessed January 29, 2018.

27. Beaulieu-Jones BK, Wu ZS, Williams C, Green CS. Privacy-preserving generative deep neural networks support clinical data sharing. bioRxiv preprint first posted online Jul. 5, 2017; doi: http://dx.doi.org/10.1101/159756. Accessed January 29, 2018.

28. American Medical Association. *Digital Health Study.* 2016. Available at https://www.ama-assn.org/sites/default/files/media-browser/specialty%20group/washington/ama-digital-health-report923.pdf. Accessed February 19, 2018.

29. Freidberg MW, Chen PG, Van Busum KR, et al. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. *RAND Health Q* 2014;3(4):1. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051918/?report=printable. Accessed February 12, 2018.

30. American Medical Association. *Code of Medical Ethics Opinion E-1.2.12, Ethical Practice in Telemedicine*.

31. American Medical Association. *Code of Medical Ethics Opinion E-11.2.1, Professionalism in Health Care Systems.*

32. Indiana University School of Medicine. Regenstrief EHR Clinical Learning Platform. Accessed February 19, 2018.

33. Kuang C. Can AI be taught to explain itself? *NYT Magazine* 2017; November 21.

34. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? *arXiv:1712.09923v1* 2017: December 28.

35. Pecci AW. "Explainable AI" could reduce readmission (and win clinicians' trust). *Health Leaders Media* 2018; January 16.

36. Kvedar JC. *The New Mobile Age: How Technology will Extend the Healthspan and Optimize the Lifespan.* 2017.