June 3, 2019

Dr. Norman Sharpless, M.D.
Acting Commissioner
U.S. Department of Health and Human Services
Food and Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993-0002

Submitted electronically via http://www.regulations.gov

RE: Docket No. FDA-2019-N-1185; "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 Commissioner Sharpless:

AMIA is pleased to provide input that will inform the U.S. Food and Drug Administration's (FDA) current thinking on the regulation of AI/ML-based SaMD. We support development of a regulatory Modification Framework for AI/ML-based SaMD and offer below some observations for consideration in developing the framework further.

AMIA is the professional home for more than 5,500 informatics professionals, representing front-line clinicians, researchers, educators and public health experts who bring meaning to data, manage information and generate new knowledge across the health and health care enterprise. As the voice of the nation's biomedical and health informatics professionals, AMIA plays a leading role in advancing health and wellness by moving basic research findings from bench to bedside, and evaluating interventions, innovations, and public policy across settings and patient populations.

AMIA commends the FDA for publishing this draft Modifications Framework and for leading a much-needed conversation on this emerging and highly important topic. Further, we applaud the FDA for offering concepts such as SaMD Pre-Specifications (SPS), Algorithm Change Protocol (ACP) and Good Machine Learning Practices (GMLP) to drive the Modifications Framework. We support these concepts and encourage FDA to engage broadly with stakeholders to refine them, especially the concept of GMLP.

This support notwithstanding, we propose the FDA modify and supplement its Framework with four critical components:

1. A stronger emphasis and acknowledgement of how starkly different continuously learning algorithms must be treated from "locked" algorithms;

2. A discussion of how new data inputs will impact the algorithm's outputs;
3. A discussion of how cybersecurity risks, such as hacking or data manipulation, may influence the algorithm's output; And
4. A discussion of how manufacturers should use evolving knowledge about algorithm-driven bias to ensure that algorithms used in affected products do not facilitate or promote such bias.

Continuously learning versus locked algorithms

While the Framework acknowledges the two different kinds of algorithms, we are concerned that the Modifications Framework is rooted in a concept that both locked and continuously learning SaMD provides opportunity for periodic, intentional updates. In particular, the ACP section assumes that periodic re-training of SaMD will occur, and that this re-training will do so under controlled circumstances where opportunities to evaluate / retest the impact of changes will occur. Our members' experience is that many AI/ML-based SaMD are intended to perform continuous updates based on real-time or near-real-time data and that the algorithms will constantly adapt as a result (see Appendix A for an example of a product more aligned with our conception of continuously learning SaMD). Figure 4 in the Framework, in specific, underscores our concern.

AMIA recommends the Modification Framework include requirements of periodic evaluation irrespective of planned updates or re-training. Further, we recommend FDA seek additional feedback to understand a basis for determining when periodic evaluation should occur. We anticipate that notifications to the FDA about changes to software could be deterministic – triggered when a threshold of data processing or algorithmic adaptations have occurred and/or the lapse of a specific time interval (e.g. a year). This could be considered akin to Genetic Shift, as Algorithm Shift.

However, there may arise conditions when the AI/ML-based SaMD's behavior changes due to real-time changes in its inputs/ACP/outputs, via Algorithm Drift. In this scenario, there needs to be some regulatory requirement that, even when the target population and indication do not change, incremental change (Algorithm Drift) in the SaMD needs to be compared with a static historical control. When the new standard drifts from the historical control by some amount (a "p-value" or a percentage, depending on the output and input), that must be a trigger for FDA review. We recommend that an annual report to the FDA indicating whether the AI/ML-based SaMD has changed as the result of Shift or Drift is an important component to "Transparency and real-world performance monitoring." We also note that a lack of change may be cause for concern among continuously learning SaMD.

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland

June 3, 2019

New data inputs' impact on algorithms' outputs

Modern AI relies more on learning from data than in the capture of existing expertise.Indeed, this gives it the power to discover new knowledge, rather than merely codify existing knowledge. However, this paradigm suffers from two weaknesses: it is vulnerable to learning from poor or biased data, thus incorporating such errors or biases, and second, it may not be able to provide cogent explanation for any decision it offers. This has been a persistent criticism of machine/deep learning approaches to AI.

We appreciate that the Framework accounts for new inputs into a SaMD's algorithm. Specifically, the Framework acknowledges that SaMD modifications could expand its intended use to new populations and sub-populations. However, patient populations in the context of AI/ML-based SaMD inputs matter greatly. We are concerned that a user of SaMD in practice would not have a practical way to know whether the device reasonably applied to their population, and therefore, whether adapting to data on their population would be likely to cause a change based on the SaMD's learning. We encourage FDA to consider a requirement for review when the SaMD's learning comes from population(s) different from its training population. A mechanism to implement this would be to require an embedded characterization of the training population in the algorithm itself, so that it may provide an estimate of its applicability in any new or marginal case.

We also note a need to view the performance of SaMD differently based on source of data inputs. For example, data inputs that come from data manually entered by clinicians, patients, or family members/caregivers pose different issues for SaMD outputs from inputs that come from fully embedded and automated devices that cannot have settings altered (e.g. entirely closed loop and based on sensed data).

A further critical question concerns the extent to which an AI-based SaMD should be able to furnish explanatory reasoning for any decision it provides or supports. In the classical form of AI, where existing expertise has been encoded, it is possible to have a chain of reasoning back to principles or data. Machine learning algorithms, however, may function in a "black box" mode, with inputs modifying the implicit circuitry with no clear traceability. It is thus vital to consider under what circumstances an AI-based SaMD should provide explanation of any decision it offers.

To address these concerns there should be strong requirements regarding transparency and availability of the original and update training data set's characteristics. Further, the FDA should develop an exhaustive list of data characteristics, such as training set population, to enumerate the dimensions for intended use. Especially when continuously learning algorithms are applied to different populations or rely on different types of data inputs (e.g. manual v. automated) from those inputs they were originally trained, there is a need for users to understand the potential impacts of new inputs or impacts to the SaMD's intended use.

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland

Cybersecurity in the context of AI/ML-based SaMD

Again, the Framework discusses FDA expectations regarding "risk assessments," but it fails to discuss how modifications to SaMD algorithms may be the result of breaches of cybersecurity and the need to make this a component of periodic evaluation. We encourage FDA to consider how cybersecurity risks, such as hacking or data manipulation that may influence the algorithm's output, may be addressed in a future version of the Framework. For example, we could envision a need for specific types of error detection geared towards preventing a system adaptation to an erroneous signal. Detection of data that may have either been corrupted or manipulated should be a priority.

Evolving knowledge about algorithm-driven bias

The Framework rightly considers algorithms within the context of the device development environments in which algorithms will be designed and the medical environments in which SaMD-based products will be used. However, a growing body of knowledge indicates that even in the absence of intended discrimination, bias against persons of particular ethnicities, genders, ages, socioeconomic backgrounds, physical and cognitive abilities, and other characteristics may occur. We recommend that FDA develop guidance about how and how often developers of SaMD-based products test their products for such biases and adjust algorithms to eliminate identified biases.

* * *

Finally, we note that there is a need for FDA to help all stakeholders explore additional dimensions of AI/ML-based SaMD, including:

- The "ingredients," or components of AI/ML SaMD, including how the SaMD was trained;
- The SaMD's appropriate application relative to other populations; and
- The capacity for SaMD to be used in a manner considered "off label," inadvertently;
- The "explain-ability" / interpretability of ML models;
    - For example, if a patient is classified as high risk initially, but following an appropriate update to the model is classified as low risk, users of the SaMD will need explanation that would help them assess contributory factors (perhaps with some acknowledgment of the change in the model).
- Common data sets or testbeds for high-priority areas, such as sepsis prediction, to benchmark and determine performance parameters;
- The extent to which an SaMD may support, supplement, or supplant human decision making, including possible legal and regulatory ramifications; and

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland

June 3, 2019

- Use of AI/ML-based SaMD as combined in complex environments or paired with technology that are used in patient care and can drive numerous solutions such as NLP Platforms, ASR technologies, and Artificial Neural Networks.

Together, further inquiry will help improve FDA's ability to regulate SaMD and help potential users understand the intentions/limitations of SaMD through detailed labeling.

As the FDA endeavors to better understand this space, AMIA offers its support and the support of its members to help regulators achieve the dual goal of patient safety and innovation. Should you have questions about these comments or require additional information, please contact Jeffery Smith, Vice President of Public Policy at jsmith@amia.org or (301) 657-1291. We look forward to continued partnership and dialogue.

Sincerely,

Douglas B. Fridsma, MD, PhD, FACP, FACMI
President and CEO
AMIA

Appendix A: Example Scenario of Continuous Learning SaMD

Example #4: Diagnostic Support Q & A Tool

Description of SaMD: A SaMD uses advanced NLP to allow it to digest vast amounts of medical literature in order to compile a knowledge base of signs and symptoms of diseases that can be used to respond to user inquiries in complex diagnostics scenarios, e.g. patients with unusual constellations of symptoms that may reflect either a rare diagnosis, or an uncommon presentation of a more common diagnosis.

The system is initially trained on well-regarded medical textbooks and journals, and has also been trained to process EMR data to identify unusual features of a patient's data. The system has been further trained to compute patient similarity metrics such that it can identify patients similar to the one in question to use as a reference point.

The system is validated by expert users and goes into production in such a way that it continually ingests newly published literature as well as daily EMR updates from the institution where it is used and offers suggestions to a user in a Q & A format, attempting to identify missing data elements that may be important, such as missing medication or diagnostic information.

SPS: The manufacturer expects the system to utilize data in ways not specifically programmed in order to adapt to changing real-world situations and medical reports.

- Allows the algorithm to make differential diagnostic suggestions to medical staff based on Q & A interaction.

ACP: For these modifications, the ACP details methods for real-world data collection, including inclusion and exclusion criteria, reference standard information, and comparative and statistical analysis for performance testing. The ACP also details the analytical validation for performance improvement, as well as the clinical validation for determining high-confidence cases. The manufacturer follows GMLP.

Modification Scenario 4A: Improved diagnostics due to continually updated patient data, consistent with SPS and ACP

A patient in New Mexico with symptoms of lupus has presented to a provider, but the diagnostic tests are all negative so the user accessed the SaMD's interface to see what information may help determine the correct diagnosis.

June 3, 2019

The SaMD concurs with the initial diagnosis based on the patient's presenting symptoms, however, it also checks for similar patients and finds that several other patients presented with the same symptoms, and were also initially negative for Lupus, but two of whom were eventually discovered to have Lyme Disease. The system indicates that late stage Lyme Disease would also be consistent with the symptoms, however the vector for Lyme disease is not found in New Mexico, and the patient's medical record did not have any information about travel history to support such a diagnosis, so this was initially ruled out as an option. Since the SaMD detected a high frequency of Lyme Disease among the cluster of patients, it suggests inquiring about this patient's travels. Visits to areas of the country where Lyme Disease is prevalent are confirmed, the patient is tested for Lyme disease and the correct diagnosis is made.

The manufacturer re-validated the algorithm based on the accumulated real-world data, as described in the ACP, which improved the SaMD accuracy in identifying a potential diagnosis. The system had updated its own algorithms as a result of the patient similarity analysis. The modified algorithm can be marketed without additional FDA review.