Alliance for AI in Healthcare (AAIH)
1340 Smith Ave, Suite 400
Baltimore, MD 21209
Tele: (410) 779-1245

December 23rd, 2019

Public Comment to Docket #FDA-2017-D-6569: Clinical Decision Support Software Draft Guidance

Dear FDA,

While we have compiled a combination of **general and specific comments**, we first want to recognize FDA's key role in this ecosystem. We applaud the Center for Devices and Radiological Health (CDRH) for its leadership in **regulatory policy related to software**. We also appreciate the **agency's efforts to clarify** jurisdictional lines, including the issuance of this **new draft guidance for public comment**.

The Alliance for Artificial Intelligence in Healthcare (AAIH) welcomes the opportunity to provide comments on FDA's Clinical Decision Support Software Draft Guidance issued September 2019. AAIH is an **international, not-for-profit, multi-stakeholder advocacy organization** (501(c)(4)), launched in 2019 to **promote scientific, regulatory, and legislative initiatives** necessary to facilitate the development of, access to, and implementation of **Artificial Intelligence (AI) powered healthcare solutions.**[1] AAIH is comprised of over 25 organizations that use or develop AI for **biomedical research, development, and clinical applications**. Our members include leading-edge **start-ups**; established **pharmaceutical, diagnostics and device manufacturers**; **academia**; and **technology and infrastructure** developers.

AAIH takes the lead on the **sector's most pressing and significant issues**, **fostering research, development, investment, and commercialization** of transformational medical care for **patients worldwide**. Our mission is to enable the **advancement and use of artificial intelligence** in healthcare to improve patients' lives. As an organization, AAIH is bringing together industry, academia, research institutions, government, NGOs, and other stakeholders to **support innovation** and the **thoughtful and responsible application** of AI in the development of new medical interventions. By convening diverse stakeholders with a wide range of relevant expertise and viewpoints, we are **working to establish responsible, ethical, and reasonable principles** for the **development and implementation** of **AI in healthcare.**

It is out of that **dedication to our mission** that we submit our comments today.

Respectfully Submitted,

On behalf of AAIH, led by the Federal Engagement and Regulatory Affairs Committee (FERAC):

Annastasiah Mudiwa Mhaka, PhD (Co-Founder, mudiwa@theaaih.org)
Aaron Chang (Strategy & Technical Advisor, achang@theaaih.org)

[1] "AAIH-About Us." AAIH. https://www.theaaih.org/faqs

I. The Potential for AI in Clinical Care

With recent increases in computing power and networking speed, decreases in the cost of data storage, the surge in research developing better and different types of algorithms, and the widespread awareness of high profile successes, there is great enthusiasm for the potential for AI to transform health care. As examples, AI could help predict acute decompensation among patients with chronic illnesses, forecast the occurrence and pattern of epidemics, guide drug discovery and inform drug development, and facilitate personalized medicine. Almost all major healthcare organizations and life science companies are currently employing or investigating applications that utilize AI to better meet their objectives.

One way AI has the potential to improve healthcare is by generating new tools to (1) assist health care professionals in providing high quality care to patients, and (2) empower patients to make better decisions in their health care and the health care of their loved ones. Clinical Decision Support (CDS) software developed using AI could reduce misdiagnoses, delayed diagnoses, unnecessary treatment, and the associated costs to patients and the health care system in general. AAIH encourages FDA to adopt policies that support the responsible, ethical development and incorporation of AI in medical care to improve the lives of patients. Given the nascency and dynamism of the field, we encourage FDA to avoid overly prescriptive regulation that may become quickly outdated or may otherwise stifle positive innovation. We urge FDA to consider the potential benefits of AI in medicine as the agency develops regulatory policy, including this guidance.

II. Explainable AI

A considerable focus of AI research today is developing frameworks and methodologies for understanding and explaining the decision systems within AI software functions. The ability of a human to understand an algorithm's decision system is often called "explainability." As articulated in AAIH's primer on AI in healthcare[2], *also attached:*

> Machine learning specifically yields difficult to interpret models – even for the data scientists who create them. Machine learning algorithms may and often do arrive at predictions in a different way than humans. So, when a complex model makes a prediction, it may not be clear to humans *why* that prediction was made. Sometimes this obscurity is referred to as a "Black Box": data goes into the model and an output is produced – how it is produced remains a mystery to humans[2], *see page 16.*

The conventional wisdom among software engineers and data scientists developing and using machine learning tools today is that, in general, there is a tradeoff between explainability and accuracy. Models for phenomena that are more complex, and therefore difficult to explain, tend to have greater accuracy, while simpler models could be easier to explain, may be less accurate. AAIH members are putting significant resources into developing methods for explaining complex algorithms for use in healthcare. For example, AAIH members who have begun evaluating the explainability of models by validating algorithm outputs with key medical experts in order to understand and map a model's biological underpinnings. Other methods of enabling explainability include Shapley values and IBM's AIX360.

In regulating AI-based software functions that are medical devices, FDA should understand the importance of explainability, its limitations, and its potential tradeoffs. We recommend FDA apply the principles of benefit-risk and the least burdensome framework that guides FDA's oversight of medical devices more broadly. Explainability is essential in some circumstances, based on the software function's intended use and other factors, while in other

[2] "A Primer on Artificial Intelligence in Healthcare." AAIH. Sep 2019. https://www.theaaih.org/publications

circumstances, accuracy will be paramount and worth the loss of explainability. For example, algorithms developed by machine learning that meet the criteria in section 520(o)(1)(E) of the FD&C Act will be explainable. Algorithms that are not explainable can and should be appropriately validated using other methods to provide a reasonable assurance of safety and effectiveness.

By analogy, FDA approves drugs with unknown mechanisms of action, even over-the-counter drugs like acetaminophen, when the use of those drugs is supported by substantial evidence of safety and effectiveness. AI-enabled software that is a medical device should be treated similarly, aligned with statutory standards for medical devices. Patients will be best served if FDA allows paths to market for both non-explainable and explainable AI, as appropriate for the intended use.

III. Comments on FDA's Clinical Decision Support Software Draft Guidance

The comments we provide on the draft guidance are specific to the use of AI in developing Clinical Decision Support software. We believe that certain explainable algorithms developed through machine learning meet the criteria in section 520(o)(1)(E). FDA's final guidance should include how the statutory language applies to machine learning algorithms. As it stands, the draft guidance is silent on machine learning, providing more regulatory certainty for traditional algorithms than machine learning algorithms. This could encourage the proliferation of algorithms that are less advanced than what technology affords, distorting the marketplace in a way that does not benefit public health. In addition, we look forward to being part of a broader conversation with FDA and other stakeholders to ensure that health care providers continue to exercise independent judgment when using Clinical Decision Support software that is intended to assist in, but not substitute for, human decision making.

A. We recommend FDA clarify how section 520(o)(1)(E)(iii) of the FD&C Act applies to algorithms developed through machine learning.

FDA provides four elements to its interpretation of section 520(o)(1)(E)(iii). We agree with FDA that non-device CDS should describe, in plain language, (1) the purpose or intended use of the software function and (2) the intended user. Our comments related to the other two elements follow.

1. The inputs used to generate the recommendation

We recommend FDA clarify how ***"inputs used to generate the recommendation"*** apply to an explainable algorithm developed through machine learning. Some software functions developed using AI could have certain fields that a health care professional, patient, or other user would enter in, that are determinative of the recommendation. We understand those fields to be ***"inputs used to generate the recommendation."*** Other software functions could scrape existing patient-specific data sources, e.g., an electronic health record. Only a subset of that data will be determinative for a given recommendation. Moreover, if such an algorithm is developed using machine learning, the training set could include published literature, FDA-approved labeling, or other publicly available sources. Neither the electronic health record nor the publicly available sources seem to be ***"inputs used to generate the recommendation"*** as contemplated by the draft guidance.

The development and refinement of a machine learning algorithm involves *featurization*, during which the model identifies *features* of data that are determinative of the output. For example, age, gender, and specific disease biomarkers could be the features of a patient that underly the given recommendation of a given Clinical Decision Support algorithm that inputs a patient's electronic health record. For some algorithms developed through machine learning, features will be complex and not human understandable. The biggest reasons for this are (1) the number of features extracted from the data could be in the millions for a given patient; and (2) features may not be

independent of each other, e.g., high blood pressure could be a feature whose weight in the recommendation could vary depending on a patient's A1c level. If clinical decision support software uses explainable AI, the software should be able to identify and describe, in plain language that the user can understand, the features of the data that the algorithm uses to generate the recommendation as well as the relationship among those features. We recommend FDA clarify that "inputs" can also mean "features" for algorithms developed through machine learning.

2. The basis for rendering a recommendation

Machine learning algorithms have the power to analyze large amounts of data to identify relationships with greater robustness and fidelity than traditional algorithms. For example, a doctor may have a patient with an unknown disease; an algorithm developed through machine learning using vast amounts of reputable, publicly available sources could identify the relationship between the patient's symptoms, a genetic variant, a rare disease, and a clinical study testing a treatment for that disease, all supported by sources available to health care professionals for independent review. Because of the mathematical nature of machine learning algorithms, each recommendation would have a quantifiable degree of certainty that the recommendation is correct. For some patients and clinical decisions, an algorithm might provide a recommendation with a high degree of certainty, while for others, various options might have similar likelihood of success. For example, an algorithm might recommend a diagnosis with 95% confidence for a given patient with a certain set of symptoms and relevant characteristics, and two possible diagnoses with 70% confidence each for a second patient with similar symptoms and different other characteristics.

While FDA's guidance on what it means to ***"describe the basis for a recommendation"*** may apply to traditional algorithms and to some explainable AI, we propose FDA adopt a slightly different, though similar, framework for most algorithms developed through machine learning. For example, "***a description of the logic or rationale used by an algorithm to render a recommendation"*** would not help a user of an algorithm developed through machine learning to be able to independently review the basis for the recommendation. Such a software function could instead provide:

(1) a list of the specific publicly available sources supporting the specific recommendation;
(2) the features of the patient used to generate the recommendation;[3] and
(3) an estimate of the certainty of the recommendation, e.g., the expected accuracy for a categorical prediction or a confidence interval for a numerical value.

With the list of specific publicly available sources, coupled with the features of the patient that the algorithm used to generate the recommendation, the health care professional has the information he/she needs to independently review the basis of the recommendation. The healthcare professional can review the supportive sources and understand the relevant characteristics of a given patient that lead to the recommendation. In addition, the estimate of certainty allows the health care professional to use his/her own professional judgment to decide whether and how to incorporate the algorithm's recommendation into the clinical decision-making process for a given patient. As a result, the software function meets the statutory criteria of not intending the health care professional to rely primarily on the recommendation to make a clinical diagnosis or treatment decision.

[3] We recognize that we also recommended earlier in this comment that FDA clarify that "inputs used to generate the recommendation" can mean "features used to generate the recommendation" where an algorithm is developed using machine learning. For certain machine learning algorithms, we propose that only three elements are necessary to meet the statutory text of section 520(o)(1)(E)(iii) of the FD&C Act: (1) intended use; (2) intended user; and (3) basis for rendering the recommendation. This is true so long as the basis for rendering the recommendation includes (a) a list of the specific publicly available sources supporting the specific recommendation; (b) the features of the patient used to generate the recommendation; and (c) an estimate of the certainty, e.g., the confidence interval, of the recommendation.

We anticipate that methods enabling explainable AI will improve and proliferate over time, given the active research in the area. Therefore, we recommend that FDA leave open the possibility than an algorithm could have other legitimate ways of ***"describ[ing] the basis of the recommendation"*** in a way that the intended user would understand, and that could meet the criteria in section 520(o)(1)(E).

3. Sources

We recommend that FDA clarify that appropriate sources for recommendations under this guidance include reputable sources related to off-label use of medical products, such as peer-reviewed literature reporting clinical studies on off-label use and information in FDA's recently launched CURE ID app. FDA stated that the CURE ID app, ***"may serve as a resource for practitioners making individual patient treatment decisions in the absence of established safe and effective options."***[4] Peer-reviewed literature on off-label use is a similarly important resource. A machine learning algorithm that intelligently synthesizes reputable sources of off-label use, including information in the CURE ID app could better assist clinicians in identifying potentially life-saving treatments for diseases without other treatment options. Insofar as such a machine learning algorithm meets the criteria in section 520(o)(1)(E), the health care professional will be able to independently review the basis for the recommendation, e.g., the specific peer-reviewed studies supporting the off-label use and/or the specific entries in CURE ID, as part of their practice of medicine. In addition to providing regulatory certainty about the use of important potential sources that could help physicians care for patients with few options, this approach is consistent with the language in current draft guidance, the statutory language in section 520(o)(1)(E) of the FD&C Act, and the First Amendment.

B. FDA's application of the IMDRF framework creates complexity and ambiguities that could be resolved through additional explanation and examples.

We agree that FDA's oversight of software should be risk-based. We also applaud FDA for working toward international harmonization of software regulation. However, our membership found the application of the IMDRF framework in this context confusing. We would appreciate clarification that deeming Clinical Decision Support to "inform" clinical management does not create an additional criterion on top of section 520(o)(1)(E). Whether software "drives" or "informs" clinical management is not always clear. The role of software function in a health decision is a gradient, not three crisp categories. Similarly, the severity of disease is a gradient. Whether a disease is "serious" or "non-serious" can be a difficult question, with border cases. We recognize that FDA is often in the position of drawing regulatory lines in grey areas. However, we question whether doing so is necessary or helpful for this particular policy.

We also note that the draft guidance does not fully align with the IMDRF framework, in that IMDRF recommends the same level of regulatory oversight when information provided by SaMD informs the clinical management of serious and non-serious health care situations or conditions. FDA's draft guidance recommends the same level of regulatory oversight when information provided by SaMD informs the clinical management of critical and serious health care situations or conditions.

Some AAIH members questioned whether health care professionals and patients should be treated differently. As time passes, we anticipate that patients will play greater roles in their own health care decisions and will be educated partners for clinicians. The proliferation of social media, apps, websites, wearables, and other software-related products in the health space has already led to an increased democratization of health information and

[4] "CURE ID App Lets Clinicians Report Novel Uses of Existing Drugs." FDA. 5 Dec 2019. https://www.fda.gov/drugs/science-and-research-drugs/cure-id-app-lets-clinicians-report-novel-uses-existing-drugs

decision making. At the same time, we recognize the important role that health care professionals play in making diagnostic and treatment decisions. For software functions for serious or critical conditions where the patient is the intended user and the patient can independently review the basis for the recommendation, an alternative distinction for FDA enforcement prioritization could be whether the health decision would be mediated by a health care professional (enforcement discretion), or whether the software intends for the patient to make the health decision on his own (FDA oversight). For example, if the software function intends for a patient to make his own decision about whether to treat cancer by undergoing dietary changes or an FDA-approved treatment regimen that his doctor prescribed, FDA should oversee that software function.

IV. Conclusion

AAIH thanks FDA for the opportunity to comment on this draft guidance. The use of AI in healthcare has great potential to improve lives and reduce costs, so long as it is deployed responsibly and with appropriate regulatory oversight. Specifically, AI-enabled algorithms are poised to support clinical decisions in transformative ways. We recognize the important role of FDA in considering the benefits and risks of new technologies and encouraging positive innovation to promote public health. We look forward to continuing to work with FDA and other stakeholders on policies and initiatives related to the use of AI in the development of FDA regulated products.