AMERICAN
OSTEOPATHIC ASSOCIATION

511 2nd Street, NE
Washington, DC 20002
312.202.8000

osteopathic.org

April 4, 2025

Martin A. Mackary, MD, MPH
Commissioner
U.S. Food and Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993

Re: FDA Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations [Docket No. FDA-2024-D-4488]

Dear Commissioner Mackary,

The American Osteopathic Association (AOA), on behalf of the more than 197,000 osteopathic physicians (DOs) and medical students we represent, appreciates this opportunity to provide the U.S. Food and Drug Administration (FDA) with information and policy recommendations to sustain and enhance artificial intelligence (AI) device innovation in the United States through the total product lifecycle (TPLC).

Innovation in AI-driven technologies holds great promise for improving patient care through promoting better outcomes, increasing efficiency, supporting physician decision-making, and enabling physicians to spend more time with patients. Osteopathic physicians, who are trained in patient-centered, whole-person care, play a critical role in fostering innovation, whether through development of new technologies, leading product adoption and implementation in their care settings, or using innovative care tools in their practice. We applaud FDA for developing this guidance on marketing submissions for AI enabled devices, which will drive greater product safety and reliability while supporting innovation.

We understand that the proposed guidance and appendix resources, including the Model Card in appendix E, outline how FDA will evaluate applications and recommendations for submission, and they are not explicit requirements for submissions. We believe that many of the recommendations should be made mandatory to promote transparency and advance uptake of AI-devices in the medical field. The AOA recommends that FDA prioritize policy and guidance development that increases transparency for consumers, and the following recommendations:

- Incorporate more assessment and oversight to mitigate bias;
- Establish data transparency for users to understand the product and its intended use;
- Improve standards for data quality used to validate AI-device performance; and
- Strengthen requirements for the premarket phase to avoid or reduce product recalls; and
- Align FDA AI-device TPLC requirements with example Model Card.

The AOA recognizes the importance of carefully balancing policies that ensure product safety and effectiveness while permitting developers sufficient flexibility to innovate and create unique products. FDA must ensure oversight that provides clear and structured requirements around product performance transparency, safety, and effectiveness. In the absence of federal leadership on these issues and adequate oversight, states will adopt regulations, which would create a patchwork of regulations for developers to follow. Requiring developers to comply with 50 differing sets of AI-enabled device regulations would discourage innovation and increase costs on developers, in difference to FDA's commitment to "least burdensome"

511 2nd Street, NE
Washington, DC 20002
312.202.8000
osteopathic.org

principles. Additionally, for broader adoption of novel technologies to take place, patients and physicians must be assured that products are safe and reliable. It is with this perspective that the AOA urges FDA to finalize its proposed guidance and offers input on how FDA can strengthen policy to support safe AI innovation and adoption in healthcare.

Device Description

A device description is intended for developers to provide information during marketing submission to help FDA understand the general characteristics of an AI-enabled device, including the intended use, expected device clinical workflow, clinical setting use, features of the model, and design of the AI-enabled device. In the proposed guidance, FDA recommends but does not require a list of device description elements, such as the intended use environment or the intended users and their desired level of training. These elements are vital to understanding how the device should be used safely and effectively, yet developers are not required to report most of these details to FDA.

While the device description is intended to allow FDA to determine the safety and effectiveness of an AI-device, we are concerned that current standards and requirements are not strong enough to ensure FDA has a robust understanding of an AI-device, its function, and its data training, particularly whether a device can achieve the expected performance level. Although FDA has approved over 950 medical devices driven by AI, as many as 43% lack clinical validation data in their FDA submissions,[1] and at least 211 products have been recalled.[2] Products marketed for clinical settings are not performing at the level of performance expected, and a lack of understanding around a device's true potential due to incomplete or limited descriptions may contribute to the underperformance. **The AOA encourages FDA to mandate descriptions of all of the information included in the example Model Card to be submitted to increase transparency and understanding of AI-device function and effectiveness, as well as to foster trust and safety for physicians and patients. We urge FDA to strengthen regulatory requirements for device descriptions and clinical validation data to ensure FDA can assess and oversee AI-devices at the point of market submission to reduce product recalls after market introduction.**

User Interface and Labeling

Clear and transparent device labeling allows physicians and other users to understand the product and the data used to train the product. Labeling regulation[3] states that "*adequate directions for use"* means directions under which the layman can use a device safely and for the purposes for which it is intended. For prescription devices, regulation states that labeling must contain adequate information for such use, including indications, effects, routes, methods, and frequency and duration of administration and any relevant hazards, contraindications, side effects, and precautions, under which practitioners licensed by law to employ the device can use the device safely and for the purposes for which it is intended, including all purposes for which it is advertised or represented. For physicians to be confident that a software enabled device can be used safely and as intended, labels need to contain information on the data used to train the AI device to ensure that they are using the device for the intended audience. Many tools are not validated in a local site and may not be generalizable. A tool that was trained for a large academic health system may not be as useful in a rural health population, at least not without proper disclosure to users. Disclosing this information to users with labeling allows users to understand the product and its use in relation to the intended population. **The AOA recommends that FDA explicitly requires labeling instructions for AI-devices to include a brief description of the data used to train an AI-device, including the demographics, geographic regions, and sample size.**

---

[1] Chouffani El Fassi, S., Abdullah, A., Fang, Y. et al. Not all AI health tools with regulatory authorization are clinically validated. *Nat Med* (2024).
[2] Muehlematter et al. "FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks." *The Lancet Digital Health*, Sep 2023
[3] 21 CFR Part 801

511 2nd Street, NE
Washington, DC 20002
312.202.8000
osteopathic.org

In 2024, HHS Office of Civil Rights (OCR) modified regulations for section 1557 of the Affordable Care Act to require physicians and other provider entities to monitor how point-of care decision support tools perform across different populations. This modified regulation made physicians and provider entities responsible when decisions using these products result in biased outcomes. If the AI software is trained with data that incorporates race, sex, age, disability, or other demographic factors, physicians are required to make "reasonable efforts" to mitigate risk of discrimination when the tool is used for health programs or activities. Because physicians often do not have sufficient information on device training and performance, they are likely to be more hesitant to adopt new, innovative tools due to the current state of transparency and liability. In a 2024 survey, 87% of surveyed physicians stated physician liability for errors in AI models would impact adoption of AI tools into their practice[4]. **While AOA continues to urge HHS to rescind this regulatory change, stronger and clearer labeling requirements would provide physicians with the information needed to reasonably understand how the software was trained and validated and encourage physician uptake of AI-devices.**

Risk Assessment

A detailed risk assessment provides FDA with a clear picture of an AI-device's potential malfunctions and opportunities for failure from user error. Developers are expected (but not required) to identify risks and develop a plan to control appropriate risks. The proposed guidance specifically calls out risk related to users' understanding of device and output information as a consideration, and it outlines usability evaluation considerations to minimize human error. However, the guidance falls short on including risk mitigation specifically related to bias. If the risk assessment does not include considerations for bias and plans for mitigating bias, then the AI-device likely may not be safe for use on patients. **The AOA emphasizes the importance of risk assessments in the premarket phase and encourages FDA to pursue risk assessment regulatory requirements in alignment with the information outlined in the "Data Management" section of this guidance as well as the "Risk Management" section of the example Model Card, which includes bias risks.**

Data Management

Data management describes data collection, processing, annotation, storage, control, and use. This characterization helps FDA to understand how the AI-device was developed and trained. FDA emphasizes in the proposed guidance that data management is an important area for identifying and mitigating bias, which can be done through the use of representative training data. **This is an important step in abating bias, and FDA should incorporate bias mitigation across the TPLC, as mentioned throughout our comments**.

The type of data used to develop and train a device can impact performance, safety, and possible bias. From 2015-2020, the majority of AI-devices approved by FDA were evaluated with retrospective data[5], which means the devices' predictions and recommendations were not tested on real patient assessments in clinical settings. Instead, the products were evaluated based on how the device might have performed on historical cases. Most of the devices included in the study were compared against a physician's performance without AI, and in many cases, the data was collected from only one or two sites. It is important to highlight that bias in an AI model does not only refer to outputs related to particular patient demographics, but more broadly also refers to how "models can be over-trained to recognize features of images that are unique to specific scanners, patient subpopulations, or clinical sites but have little to do with generalizable patient anatomy, physiology, or condition" as noted in the guidance. This over-training can then result in unintentional, unforeseeable outcomes for different patient populations based on demographic, geographic, and other factors. As a result, it is essential that models are trained on representative data.

[4] AMA Physician AI Sentiment Report. 2024. Available here.
[5] Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D., and Zou, J. "How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals". *Nature*. 2021.

511 2nd Street, NE
Washington, DC 20002
312.202.8000
osteopathic.org

While the FDA acknowledges these issues in the guidance, we are concerned that it does not ensure sufficient data management oversight. The guidance outlines the types and descriptions of data that developers should submit for each data management activity, but we urge FDA to go further in **developing more stringent requirements to address the data quality and data types used to mitigate bias. While outlining recommendations for information that should be included in submissions is an important step forward, we urge FDA to build on this guidance in developing regulation that will ensure AI-devices can function safely and correctly in the real world. The AOA urges FDA to require explicit public disclosure of the types of data with which products were tested or validated**.

Validation

Validation in the TPLC is used to ensure the device will perform as intended. The type of data and testing site are critical to understanding how the model can perform and be used across different populations. This information is valuable for FDA and for AI-device users. As we note under Data Management, data used during performance validation should be collected across numerous sites and be representative of the diverse population in which they are intended to be used.

FDA should also require data transparency to encourage uptake of AI-devices in clinical settings. End-users of AI-devices, including physicians and healthcare facilities, may be hesitant to purchase or adopt an AI-device if they do not know what data was used to train the device (e.g. the source of the data, whether it is real patient data or synthetic), or whether the product was validated via trials or other means. Public Submission Summaries are required for 510(k) devices, and summaries of premarket approval (PMA) and De Novo device decisions are published on the FDA website. **We encourage FDA to adopt transparency requirements for performance validation and publicly publish more detailed summaries to encourage trustworthy AI-device use in clinical settings**.

Device Performance Monitoring

A unique feature of AI-devices is the ability to evolve as it performs. This feature makes it more challenging to predict long-term performance and safety if developers make continuous updates based on data collected after market introduction. Any changes to the AI algorithm or its parameters must be evaluated for safety and efficacy. For devices requiring PMA or premarket notification (510(k)), manufacturers are required to submit updates to the FDA if any updates or modifications could significantly impact the device's intended use or performance[6]. This could include submitting a new premarket application if the changes are substantial enough to alter the intended use or core functionality of the device. FDA has guidance on Predetermined Change Control Plans (PCCPs), offering a less burdensome opportunity for manufacturers to modify devices in an iterative process that improves performance while maintaining the same level of safety and effectiveness. PCCPs cover modifications that could "significantly affect" the safety or effectiveness of the device, and the guidance covers modifications that are implemented automatically (through continuous learning), manually (involving steps that require human input, action, review, and/or decision-making), or a combination of both.

However, the guidance falls short on promoting and enforcing protections to combat bias that often develops as algorithms perform. The current PCCP guidance and proposed AI-device guidance heavily rely on the developer during the premarket process and PCCP submission to anticipate where bias may arise. The developer may be able to consider how the AI-device would deviate or develop bias, but until the device functions in the real world, manufacturers may not know exactly what issues may arise. A study found that a commercial algorithm—used to identify and assist patients with complex health needs—displayed significant bias after deployment[7]. The algorithm identified Black patients needing extra care at nearly half the rate it

[6] Section 515C of the FD&C Act (21 U.S.C. 360e-4)
[7] Obermeyer, Z. et al. "Dissecting racial bias in an algorithm used to manage the health of populations". *Science*. 2019.

identified White patients. The study concluded that the bias occurred because the algorithm used health costs to identify need. Less money is spent on Black patients with the same level of need as White patients[8], so the algorithm falsely concluded that Black patients were healthier than White patients that are equally sick. While some forms of underperformance or bias are predictable through proper diligence, simulation, and use of diverse datasets, some data shifts or biases are difficult to anticipate. Other performance issues arise simply from model design flaws. AI-device performance postmarket should be closely monitored to ensure patient safety and reduce underperformance, bias or other unintended outcomes. **FDA should establish stronger post-market evaluation requirements for manufacturers to monitor and detect issues to maintain performance over time. PCCPs should not be permitted or approved without some human intervention or review of performance in the updates.**

Conclusion

The AOA looks forward to continuing to work with FDA on developing policy to ensure safe and effective AI-device development and deployment. We truly appreciate the opportunity to submit comments on this proposed guidance. Should you have any questions regarding our comments or recommendations, please contact John-Michael Villarama, Vice President for Public Policy at jvillarama@osteopathic.org at any time should we be able to support your efforts.

Sincerely,

Teresa A. Hubka, DO, FACOOG (dist.)
President, AOA

Kathleen S. Creason, MBA
Chief Executive Officer, AOA

[8] Ben, J. et. Al. "Racism and health service utilisation: A systematic review and meta-analysis". Plos One. 2017.