November 15, 2021

Attn: Nooshin Kiarashi, PhD
Center for Devices and Radiological Health
U.S. Food and Drug Administration
10903 New Hampshire Ave.
Silver Spring, MD 20993

**Re: (FDA-2019-N-1185) Virtual Public Workshop - Transparency of Artificial Intelligence/Machine Learning-enabled Medical Devices; Comments of the American College of Radiology**

The American College of Radiology (ACR)—a professional association representing over 40,000 diagnostic radiologists, interventional radiologists, nuclear medicine physicians, radiation oncologists, and medical physicists—appreciates the opportunity to provide input to the Food and Drug Administration (FDA) regarding the Oct. 14, 2021, public workshop on transparency of artificial intelligence/machine learning (AI/ML)-enabled medical devices (FDA-2019-N-1185). The following comments focus on radiology device considerations of interest to providers, although we recognize that patient stakeholders may have additional transparency needs.

**Need for Enhanced AI Transparency**
The performance of AI/ML-enabled radiology software may be significantly reduced and could result in patient harm when AI models are used in actual patient care which may expose them to input data that are different from the original training and validation environment. This brittleness in model performance from facility to facility is becoming more notable as the use of AI /ML-enabled software has become more common in real-world practice settings. It has been attributed to practice variations in patient populations, image acquisition/input devices (which can differ in terms of manufacturer, model, and generation), and imaging protocols, among other factors.

Providers—including decision-makers responsible for technology acquisition, as well as radiologist end-users of the software devices—must be empowered to make better informed purchasing decisions. This includes the ability to readily discover meaningful product labeling and performance data to inform adoption and use considerations. Furthermore, patients and the public should also be able to access this information to enhance the overall trustworthiness of AI/ML in clinical use.

Finally, the performance of AI models may degrade over time as equipment and protocol changes occur in a given practice, and developers need to work openly with end-users to ensure there is a mechanism to monitor the performance of their models over time.

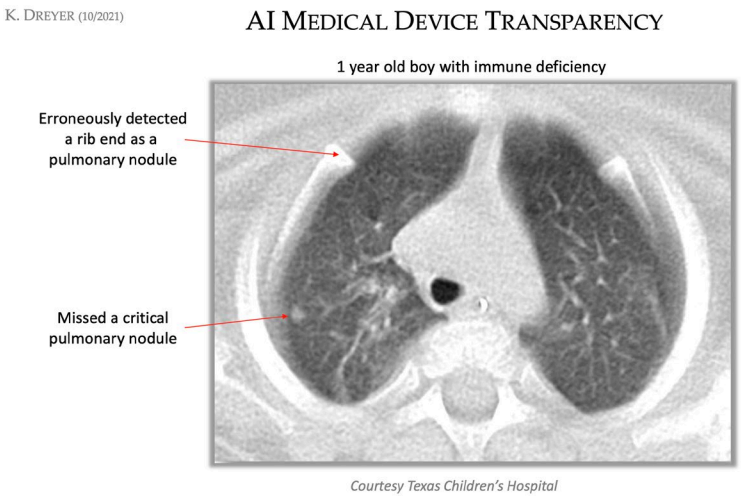
**Performance Concern Reporting To FDA**
Although there has been an increase in anecdotal examples and reports in the medical literature documenting inconsistent performance of some AI/ML models in clinical use, we suspect that the agency is not seeing a parallel increase in reports to the Medical Device Reporting (MDR) mechanism. FDA's traditional processes for soliciting and collecting information on problems with medical device safety and effectiveness are not well-suited to learning about potential issues with AI/ML-enabled radiology software devices. For example, the limitations and requirements of the MDR paradigm in 21 CFR §803

would fail to capture most instances of software underperformance due to adverse event harm prerequisites and the need for manufacturers to concur with the attribution of the software to the reported adverse event. Additionally, as none of the available AI/ML-enabled radiology software has been authorized for autonomous or unsupervised use, instances of poor model performance are typically recognized and corrected at the point of service by the radiologist before leading to patient harm. There are no production systems in widespread use today where disagreements between AI and end-users can be easily recorded, or ways to relay underperformance content (including specifics on the circumstances) back to the FDA in a structured manner.

To better understand software underperformance, the FDA should establish an alternative mechanism for compiling performance concerns from real-world users of AI/ML-enabled software devices, including those that do not trigger reportable adverse events. The reported information should be accessible to providers and other consumers to heighten public awareness of potential product issues and trends, for example via a national registry for real-world monitoring.

**Insufficient Transparency Concerns – Pediatric Radiology Example**

Radiology providers serving subsets of the general population, such as pediatric patients, face significant challenges in identifying suitable AI/ML-enabled software for use in their patients' care. During the October 14 workshop, ACR's presenter, Dr. Keith Dreyer, showed examples of how AI/ML models designed for lung nodule detection underperformed in pediatric patients.




These failures could be due to the use of low dose CT protocols in children as compared to adults, differences in the body habitus of children as compared to adults, misregistration related to breathing, or differences in the morphologic features of lung nodules in children as compared to adults; an AI/ML model that was not trained using datasets from pediatric patients and protocols might not be, or even expected to be, generalizable to a pediatric population. Furthermore, since most pediatric imaging is done in facilities caring for both children and adults, AI/ML-enabled software functions not evaluated for pediatric use could unwittingly be used by providers on children or could indirectly impact the timeliness or quality of pediatric care in multi-age care settings (e.g., because triage algorithms may only affect prioritization of [positive] adult studies, thereby indirectly deprioritizing pediatric studies by extension).

All facilities, including those that provide care to specific subsets of patients such as children, should have access to pertinent product information (see "Radiology Provider Data Needs" subsection) prior to engaging with vendors regarding purchasing decisions. This step combined with additional activities outside the scope of the Oct. 14, 2021 workshop (e.g., FDA resource prioritization programs being used

to advance pediatric radiology AI innovations, etc.), should help ensure providers are paired with appropriate solutions for the specific patient populations they serve.

**Importance of Provider-Level Evaluations of Software**
End-users have a responsibility to evaluate AI/ML-models prior to deploying them for clinical use (or during an initial acceptance phase of clinical deployment) and to monitor the performance over time. Key to that evaluation will be determining whether the model functions as expected in their local patient populations using their own equipment and protocols. While many, including the ACR, advocate pre-deployment testing at the facility, most facilities do not have the resources to evaluate and test all models for each AI use case they are considering. Moreover, this requires IT infrastructure not commonly found in clinical imaging departments. Having manufacturers expose pertinent practical data to end-users via publicly accessible documentation would significantly streamline the evaluation process for technology decision-makers and end-users.

Additionally, manufacturers should work with end-users to monitor the performance of AI\ML models throughout the software's lifecycle. Some manufacturers currently offer feedback collection capabilities in the user interface of the software, but any such feedback may not be accessible to the manufacturer's other customers or prospective purchasers. A more sophisticated approach for relevant device types would be to analyze radiology report data from all cases processed by a given AI, and to check for concordance or discordance with the end-user. Then, this data could be captured in a registry providing practice-level and practice-to-practice analytics. This level of transparency would help warn FDA, manufacturers, and customers of potential performance issues.

**Radiology Provider Data Needs**
Critically important to transparency are the performance testing dataset characteristics for population demographics including age, sex, race, and ethnicity; number of facilities; acquisition devices, including manufacturer and model; and source of ground truth. Such data could even be made accessible in a public database that allows searching by multiple parameters. Ideally, the following information would also be included in publicly accessible documentation, perhaps linked from the database entry for each AI/ML-enabled device:

- Population demographics
    - Training and performance testing dataset summaries to enable extrapolation of representativeness, generalizability, and applicability to providers' patient populations.
    - Warnings or statements about patient subgroups and/or input devices that were not part of training and subsequent performance evaluations.
    - Contraindications for device use, for example if there is potential risk of patient harm because of certain demographic data such as age.
    - Statement regarding authorization specifically for use in children, including a description of the evidence that does/does not support use in children, or if there is a lack of such evidence.
- Acquisition devices
    - Manufacturer, model, version, protocols, contrast.
- Performance measurements
    - Summaries (with links to additional details) regarding performance testing methodologies and related data, including information about the representativeness of readers (e.g., sample sizes, qualifications, etc.) that participated in reader studies, if applicable.
    - Summaries of testing/validation parameters and datasets used during software evaluations (for the same reasons stated above).

    - Sensitivity, specificity, positive and negative predictive values.
- Findings metrics
    - Conspicuity, prevalence, ground truth methods.
- Intended user qualifications (e.g., "qualified radiologist", etc.).
- Enhanced product identification
    - Use case definition.
    - CAD classification instead of, or in addition to, product codes.
- Summaries (with links to additional details) regarding post-market monitoring methodologies, metrics, and/or reporting approaches that should be known by providers to help ensure continued real-world performance.

This information should be concise, understandable, and readily discoverable by interested providers unfamiliar with FDA regulatory processes. To that end, participants at the Oct. 14 workshop described a "food label"-style high-level summary of information to simplify pertinent data that may otherwise be tucked away in complex regulatory documents. A high-level summary should minimally include key dataset characteristics used for performance testing as listed above. The content of the high-level summary could possibly dynamically change based on the category of stakeholder seeking information (e.g., purchaser, intended user, researcher, patient, etc.). Given the differences in pediatric medical needs, it could include a highly visible "pediatric use section" to help promote pediatric patient safety and effectiveness during use. FDA could either present this information in a centralized location or work with groups representing various stakeholders, including the ACR, to provide high-level summaries to their respective communities.

Manufacturers could further promote effective communication by recommending their customers identify key clinical champions within each customer site to ensure that the instructions for use and other key product information reaches the end-users. This would help address scenarios in which this information may be provided to technical implementation team but not to physicians responsible for clinical implementation.

**Summary of ACR Recommendations**

FDA should enable radiology provider access to pertinent information about AI/ML-enabled device training and performance evaluation. This information should be accessible and understandable to technology acquisition decision-makers, radiologist end-users, and other stakeholders. FDA should establish a new reporting mechanism to capture and provide access to potential AI performance issues that are not reportable via the MDR mechanism.

As always, the ACR welcomes the opportunity for further communication with FDA regarding transparency enhancements and AI/ML-enabled medical device oversight in general. Please contact Gloria R. Romanelli, JD, ACR Senior Director of Legislative and Regulatory Relations, at gromanelli@acr.org, or Michael Peters, ACR Government Affairs Director, at mpeters@acr.org or (202) 223-1670 with any questions or concerns.

Sincerely,

Howard B. Fleishon, MD, MMM, FACR
Chair, Board of Chancellors
American College of Radiology