U.S. Food and Drug Administration (FDA)
Center for Devices and Radiological Health (CDRH)
Digital Health Advisory Committee (DHAC) Meeting

Meeting Topic: *"Total Product Lifecycle Considerations for Generative AI-Enabled Devices"*

FDA DISCUSSION QUESTIONS

November 20-21, 2024

Introduction: FDA has long promoted a total product life cycle (TPLC) approach to the oversight of medical devices, including artificial intelligence (AI)-enabled devices, and has committed to advancing regulatory approaches for these devices using current authorities as well as exploring options that may require new authorities. Such an approach has become increasingly critical for modern medical devices incorporating technologies that are intended to iterate faster and more frequently over a device's life of use than ever before. GenAI-enabled products can be intended to provide variable outputs for the same inputs, may frequently rely on models that are meant to change rapidly and often, and may query models that are not themselves medical devices. Thus, a TPLC approach is likely remain important to the management of future, safe and effective GenAI-enabled medical devices. Generative AI (GenAI) is defined as the class of AI models that mimic the structure and characteristics of input data in order to generate derived synthetic content. This can include images, videos, audio, text, and other digital content (Source: E.O. 14110). For the purposes of this Committee Meeting, we are using the term "GenAI-enabled device" to refer to a device, as that term is defined in section 201(h) of the Federal Food, Drug, and Cosmetic Act, in which GenAI methods or models are integral to the device's output or functionality. In other words, for GenAI-enabled devices, GenAI methods or models play a critical role in the device's primary functions or directly enable its output. Additionally, generative AI-enabled devices are not monolithic and can have various intended uses, such that the considerations discussed may not be generally applicable to all generative AI-enabled devices. These questions aim to understand the critical information and practices needed for a comprehensive approach to the management of risk throughout the TPLC for generative AI-enabled devices, through the lens of information the Agency will need to evaluate pre and postmarket to support the safety and effectiveness of these devices.

1. ***Premarket Performance Evaluation:*** Please discuss what specific information related to generative AI should be available to FDA to evaluate the safety and effectiveness of Gen AI-enabled devices considering, for example, that foundation models leveraged by the Gen AI-enabled device will change over time and that there may be limited information available on the training data utilized for these pretrained generative models.

    a. What information should be included as part of a device's description or characterization in the premarket submission when the device is enabled by generative AI? For example, when

U.S. Food & Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993
www.fda.gov

a human is/is not intended to be in the loop, or if a device is intended only to recall information versus generate new recommendations. What information is particularly valuable to evaluate the safety and effectiveness for devices enabled with generative AI in comparison to non-generative AI?

b. What evidence specific to generative AI-enabled devices should the FDA consider during premarket evaluation regarding performance evaluation and characteristics of the training data during the total product lifecycle to understand if a device is safe and effective?

c. What new and unique risks related to usability may be introduced by generative AI compared to non-generative AI? What, if any, specific information relevant to health care professionals, patients, and caregivers is needed to be conveyed to help improve transparency and/or control these risks?

d. Are there prospective performance metrics that are particularly suited/most informative for these technologies, given their complexity? What kind of performance metrics are needed for multimodal systems, for example text/image models where either inputs, outputs or both could be multimodal? Performance metrics will typically vary with device intended use. Examples of known metrics to support discussion may be modality-specific such as for generative text (perplexity, quantitative comparison to reference text), for generative images (Frechet Inception Distance (FID), Structural Similarity Index Measure (SSIM)), or for generative audio (Log-Spectral Distance, Perceptual Evaluation of Speech Quality), or may be functionally-based, such as frequency and types of errors made by the generative AI-enabled device.

2. ***Risk Management:*** What new opportunities, such as new intended uses or new applications in existing uses, have been enabled by generative AI for medical devices, and what new controls may be needed to mitigate risks associated with the generative AI technologies that enable those opportunities? For example, controls related to governance, training, feedback mechanisms, and real-world performance evaluation.

3. ***Post Market Performance Monitoring:*** Postmarket performance monitoring and evaluation may be important for these devices, particularly because they are non-deterministic. Additionally, after deployment, many generative AI-enabled devices will undergo continuous adjustment based on localized live data, user interactions, and changing conditions. Please discuss the aspects of post market monitoring and evaluation that will be critical to maintaining the safety and effectiveness of these devices.

    a. What specific monitoring capabilities should be considered to effectively evaluate and monitor the post market performance of generative AI-enabled devices to ensure they maintain adequate accuracy, relevance, and reliability, especially when adapting to new

U.S. Food & Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993
www.fda.gov

data?

b. What specific strategies and tools can be implemented to monitor and manage the performance and accuracy of a generative AI-enabled device implemented across multiple sites, ensuring consistency, and addressing potential regional biases and data variations compared to the device that was authorized?

c. What methods and metrics can be utilized to effectively monitor and evaluate the post market performance of generative AI-enabled devices that use a multi-layer application design, i.e., the device queries external consumer-grade AI services that are not themselves medical devices?

U.S. Food & Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993
www.fda.gov