FDA U.S. FOOD & DRUG ADMINISTRATION
CENTER FOR DEVICES & RADIOLOGICAL HEALTH

EXECUTIVE SUMMARY FOR THE DIGITAL HEATH ADVISORY COMMITTEE MEETING

Total Product Lifecycle Considerations for Generative AI-Enabled Devices

November 20 – 21, 2024

TABLE OF CONTENTS

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

Disclaimer: This Executive Summary is for discussion purposes only and does not represent draft or final guidance. It is not intended to propose or implement policy changes regarding regulation of generative artificial intelligence-enabled devices. In addition, the references cited herein are for informational purposes only and should not be construed as endorsements.

Introduction

While interest in generative artificial intelligence (GenAI) tools across the health care sector has expanded rapidly, there remain open questions on the approach to regulating GenAI-enabled products that may fall within FDA's jurisdiction, including, but not limited to, medical devices. For purposes of this document, 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 (FD&C Act), in which GenAI methods or models are integral to the device's output or functionality. Like artificial intelligence (AI)-enabled products, the capabilities of GenAI-enabled products may offer unique benefits to patients and public health, but also bring new regulatory complexities for FDA to address. As with all medical devices, FDA's regulatory oversight applies to GenAI-enabled products that meet the definition of a device; such oversight is risk-based, taking into consideration the product's intended use and technological characteristics. Further, FDA has long promoted a total product life cycle (TPLC) approach to the oversight of medical devices, including AI-enabled devices, and has committed to developing regulatory approaches for these devices using current authorities as well as exploring options that may require new authorities. This commitment has become increasingly relevant for 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. A TPLC approach is likely to remain important to the management of future, safe and effective GenAI-enabled medical devices. In this executive summary, we focus on FDA's approach to the oversight of GenAI-enabled devices, which shares many similarities with FDA's approach for AI-enabled devices in general. This executive summary also discusses the risks of GenAI, some of which may be broadly applicable to AI, and current challenges to regulation of AI- and GenAI-enabled devices across the TPLC.

How Gen AI Works

Before GenAI, developers have been incorporating AI and machine learning (ML) models within medical devices. These non-generative AI/ML models have been used to perform tasks such as

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

image segmentation,[1] classification,[2] biomarker extraction,[3,4] and risk prediction.[5] Non-generative AI/ML models are generally considered reproducible, even if not fully explainable or transparent.

GenAI refers to the class of AI models that mimic the structure and characteristics of input data to generate derived synthetic content, and can include images, videos, audio, text, and digital content.[6] GenAI, like non-generative AI/ML, learns patterns from data; however, unlike non-generative AI/ML, GenAI models are generally meant to create new data that resembles the data it learned from, rather than primarily to identify patterns to make accurate predictions. GenAI models can analyze input data and produce contextually appropriate outputs that may not have been explicitly seen in its training data.

Importantly, GenAI models are frequently developed on datasets so large that human developers typically cannot know everything about the dataset contents during development. In contrast to the datasets used to develop other AI/ML models, datasets for GenAI model development can be intentionally broad and may not be initially tailored to a specific task. GenAI models derive highly complex relationships between elements within the development data that can be applied to and further optimized for one or potentially multiple tasks. While the GenAI model may generate an output that is applicable to a specific area of interest, the derived relationships are between different words, pixels, or other elements of language found in the development data. Therefore, the output reflects probabilistic predictions about the derived relationships, which can sometimes be informed by information in the development dataset that is not necessarily relevant to the specific area of interest. GenAI models that can be broadly applied to multiple tasks are often referred to as foundation models.

Current foundation models are typically not created for an individual product, nor are they generally intended for use as a device as that term is defined in section 201(h) of the Federal Food, Drug, and Cosmetic Act (FD&C Act). Application developers, who are usually different

---

[1] Seo H, Badiei Khuzani M, Vasudevan V, Huang C, Ren H, Xiao R, Jia X, Xing L. Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. Medical Physics. 47(5):e148-e167. https://doi.org/10.1002%2Fmp.13649

[2] Islam KR, Prithula J, Kumar J, Tan TL, Reaz MBI, Sumon MSI, Chowdhury MEH. Machine Learning-Based Early Prediction of Sepsis Using Electronic Health Records: A Systematic Review. Journal of Clinical Medicine. 12(17):5658. https://doi.org/10.3390/jcm12175658

[3] Mansur A, Vrionis A, Charles JP, Hancel K, Panagides JC, Moloudi F, Iqbal S, Daye D. The Role of Artificial Intelligence in the Detection and Implementation of Biomarkers for Hepatocellular Carcinoma: Outlook and Opportunities. Cancers (Basel). 15(11):2928. https://doi.org/10.3390%2Fcancers15112928

[4] Dhillon A, Singh A, Bhalla VK. A Systematic Review on Biomarker Identification for Cancer Diagnosis and Prognosis in Multi-omics: From Computational Needs to Machine Learning and Deep Learning. Archives of Computational Methods in Engineering. 30:917–949. https://doi.org/10.1007/s11831-022-09821-9

[5] Alballa N, Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. Informatics in Medicine Unlocked. 24: 100564. https://doi.org/10.1016/j.imu.2021.100564

[6] Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Executive Order 14110 of October 30, 2023, Section 3(p), 88 FR 75191.

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

entities than foundation model developers, may adapt a single foundation model for various applications, including applications that may meet the device definition. Methods used to train a foundation model for a specific task are rapidly emerging and changing as researchers focus on this space. Some researchers are exploring the use of GenAI models for tasks similar to those that may currently be accomplished with non-generative AI/ML models, such as segmentation or classification, while others are exploring the use of AI or GenAI models for newer functionality, such as broad image analysis and providing clinical diagnoses over a broad set of disciplines. Even when an AI or a GenAI model is trained specifically with medical information, that information may cross many medical disciplines and sources within the health care sector. Therefore, the model may not only generate a specific type of output but be able to do so across several clinical disciplines. For example, a model trained to identify tumor tissue on a biopsy image could potentially be used for several different tissue types and imaging modalities.

AI models, including GenAI models, can also exhibit hallucinations, i.e., the generation of erroneous or false content to meet the programmed objective of fulfilling a user's prompt.[7] The complexity of the models, including model architecture and the large corpus of data typical of GenAI models, can be a factor that leads to such hallucinations.[8] Thus, while a potentially notable benefit of GenAI is that it can generate outputs that are applicable to a specific area of interest from a variety of different data types, or that it can generate outputs that are relevant to a broad number of tasks, GenAI can also present potential risks that may require varying levels of risk controls for different applications, as is true of other technologies. For example, hallucinations, particularly those that may appear to be authentic outputs to users, may present a significant challenge in certain health care applications where highly accurate, truthful information is critical.

For medical devices, using GenAI to generate content that may help identify a possible clinical diagnosis, treatment solution, and new associations in complex medical data for patients, health care professionals, and others, could greatly benefit health care. However, at times, this same ability of GenAI to tackle diverse, new, and complex tasks may contribute to uncertainty around the limits of a device's output. When insufficiently controlled, this uncertainty can translate to difficulty in confirming the bounds of a device's intended use, which can introduce challenges to FDA's regulation of GenAI-enabled devices. For example, it may be challenging for FDA to apply a risk-based approach to classify a GenAI-enabled device and determine the applicable regulatory requirements if the device's intended use is not well-defined. Further, it will be important that adequate evaluation methods and risk controls are available to ensure GenAI-enabled devices remain safe and effective across the TPLC. While some of the evaluation methods and risk controls available to ensure AI-enabled devices remain safe and effective may also be used for GenAI-enabled devices, the interest in and needs of GenAI-enabled devices

[7] National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf

[8] Reddy GP, Pavan Kumar YV, and Prakash KP. Hallucinations in Large Language Models (LLMs). 2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream). 2024: 1-6. https://doi.org/10.1109/eStream61684.2024.10542617

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

may also accelerate the development of new evaluation methods or risk controls. As with all devices, FDA follows a risk-based approach with consideration of the product's intended use and technological characteristics to provide reasonable assurance of their safety and effectiveness.

For the purposes of this executive summary, we are using the term "GenAI-enabled device" to refer to a device as that term is defined in section 201(h) of the FD&C Act, in which GenAI methods or models are integral to the device's output or functionality. Said differently, GenAI methods or models play a critical role in the device's primary functions or directly enable its output.

Potential GenAI Applications in Health Care

There is a broad range of current and potential implementations of GenAI in health care; some of these implementations form products that are medical devices, while others are not within FDA's jurisdiction. These implementations may include administrative functions, such as facilitating clinical documentation or obtaining insurance pre-authorization.[9] GenAI models may also be used to facilitate medical training and simulations, as these models have the ability to create realistic patient simulations that adapt in real-time, helping health care professionals practice in a controlled environment.[10] Research also describes the utility of GenAI models in generating a clinical diagnosis, or the potential to diagnose and treat mental health conditions, such as through the use of chatbots,[11] which may meet the definition of a device. Because the basis of many GenAI models are foundation models that can be broadly applied to multiple tasks, a GenAI model trained to identify tumor tissue on a biopsy image could use the same foundation model as a GenAI model trained to provide a chatbot functionality to discuss the image and pathology report for that sample. GenAI models can also use free text or multimodal inputs from users, rather than structured data formats typically used by non-generative AI/ML models. This capability can enable GenAI-enabled products to accept a greater variety of possible inputs for its use. For example, a GenAI-enabled chatbot could use free text or multimodal inputs describing discrete laboratory data and vital signs, patient-specific medical data, and descriptions of the patient's scenario to generate a differential diagnosis from a unique combination of inputs that may not have been encountered during development.

As we will explore in greater detail below, the rapid rise of interest in GenAI may present challenges to FDA's laws and regulations. There is the potential that certain safeguards and new approaches to the evaluation of AI- and GenAI-enabled devices may be developed in the future, which may be helpful so that FDA can ensure the safety and effectiveness of these

[9] Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digital Medicine. 6, 120. https://doi.org/10.1038/s41746-023-00873-0
[10] Sardesai N, Russo P, Martin J, Sardesai A. Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anesthesia training. Postgraduate Medical Journal. 100(1882): 237–241. https://doi.org/10.1093/postmj/qgad137
[11] Sezgin E, McKay I. Behavioral health and generative AI: a perspective on future of therapies and patient care. NPJ Mental Health. 3(25). https://doi.org/10.1038/s44184-024-00067-w

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

devices. Additionally, while FDA has long advocated for a TPLC approach to devices to assure their safety and effectiveness over their life of use, AI- and GenAI-enabled devices bring the necessity of such an approach to the forefront and raise important questions about how to implement and evaluate the lifecycle considerations effectively. These considerations are crucial to both regulators and developers with the shared goal of protecting patients and benefiting public health.

Lifecycle Management Approach to AI

Since the 1960s, Lifecycle Management (LCM) has been a structured process for managing software and remains particularly important today for AI-enabled software, including GenAI-enabled software, which may be intended to be updated frequently over their use life. Modern Software Development Lifecycles (SDLCs) embody LCM principles, offering a structured process for planning, designing, implementing, testing, integrating, deploying, maintaining, and eventually retiring software. Building on this concept for GenAI-enabled and AI-enabled software broadly, a process tailored to the AI Lifecycle could consider aspects of software development that are more particular to AI software development, such as an added emphasis on data collection and preparation, training and validation, evaluation and testing, deployment and integration, and monitoring of AI models, as examples (Figure 1). Some of these process steps (for example, evaluation and testing and monitoring) may be particularly important for AI models, including GenAI models, as discussed below. Considering a Lifecycle Management approach to AI could be useful for performance-centric considerations (e.g., correctness and robustness) as well as broader considerations (e.g., ethics, fairness, and privacy considerations) that may be highly important to AI-enabled products, including GenAI-enabled products.

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

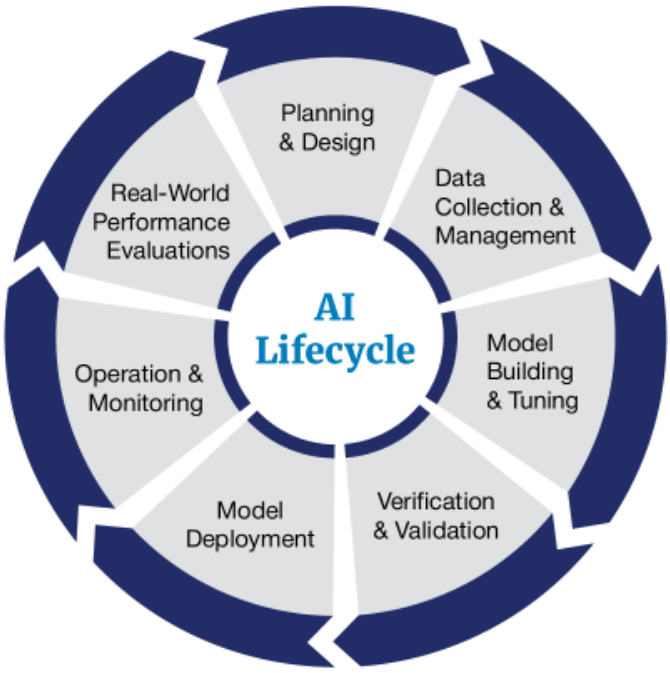

Figure 1. AI Lifecycle

An example of a general AI Lifecycle, as shown in Figure 1, can be further described as follows:

- **Planning and Design** – In this phase, the goals, scope, and requirements of the AI project would typically be defined, and the architecture and algorithms to be used would typically be designed.
- **Data Collection and Management** – In this phase, the necessary data for training and testing the AI model would typically be gathered and organized to ensure its quality and relevance.
- **Model Building and Tuning** – In this phase, the AI model would typically be developed and refined using the collected data, including selection of appropriate algorithms, feature engineering, and performance optimization.
- **Verification and Validation** – In this phase, the accuracy and reliability of the AI model would typically be assessed through testing and validation processes to ensure it meets the desired standards and objectives.
- **Model Deployment** – In this phase, the trained AI model would typically be integrated into the intended environment or application, making it accessible for real-world use.
- **Operation and Monitoring** – In this phase, the AI-enabled product would typically be tested to ensure that it operates smoothly and reliably in the production environment after deployment. This would typically involve tasks such as monitoring,

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

troubleshooting, bug fixing, and making necessary updates or enhancements to the system to address issues that arise during its operational lifetime. This phase would also typically include maintenance of the model's functionality, including addressing any issues that arise.

- **Real-World Performance Evaluation** – In this phase, the performance of the AI-enabled product would typically be assessed in real-world conditions to determine how well it meets its intended objectives and user requirements. This would typically involve evaluating various aspects of the system's performance, including accuracy, efficiency, scalability, robustness, and user satisfaction.

In general, consideration of the AI Lifecycle for GenAI-enabled devices, and AI-enabled devices broadly, may be one important way for manufacturers to approach managing their devices throughout the TPLC. Additionally, the AI Lifecycle can be used as a helpful model to identify where new techniques, approaches, or standards may be needed to assure adequate management of these new technologies across the TPLC. As discussed further below, this can be important for understanding the valid scientific evidence that may be needed to ensure continued safety and effectiveness of these devices over their lifecycle and whether these needs may be different from other devices. For example, considerations included as part of the Real-World Performance Evaluation phase of the AI Lifecycle can help manufacturers effectively monitor the postmarket performance of their AI-enabled devices. This can be particularly important for GenAI-enabled devices, which may benefit from or require monitoring plans to assure their continued safety and performance over time, given their intrinsic nature to generate new content. As such, consideration of the AI Lifecycle and its management may be helpful to address some of the current challenges for management of AI-, including GenAI-, enabled devices.

FDA's Oversight of GenAI-Enabled Devices

As with all medical devices, FDA's regulatory oversight applies to GenAI-enabled products that meet the definition of a device, following a risk-based approach with consideration of the product's intended use and technological characteristics. While the application of some aspects of FDA's laws and regulations to GenAI-enabled products is straightforward, for example, what the term device means or FDA's device classification schema, there are also some aspects where the application of FDA's current laws and regulations is more challenging. For example, because of the evolving nature of many GenAI-enabled products, it can be challenging to determine how its intended use may align within the scope of FDA's current digital health policies. Below, we explore these topics to describe FDA's potential approach to oversight of GenAI-enabled devices, and some of the outstanding uncertainties.

The term, "device" is defined in section 201(h) of the FD&C Act as an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is, among other criteria, intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention

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

of disease.[12] Certain software functions are specifically excluded from the device definition by section 520(o) of the FD&C Act, which include, for example, certain software functions intended for administrative support of a health care facility, or those intended for maintaining or encouraging a healthy lifestyle and are unrelated to diseases or conditions, to name a few. Other software functions are also not devices, because they do not meet the definition of a device even if they were not specifically excluded by section 520(o) of the FD&C Act.

To determine if a product meets the definition of a device, it is important to first identify the intended use of the product.[13] Intended use is defined as the general purpose of the product or its function, which is the objective intent of the persons legally responsible for labeling of a product (or their representatives), and may be shown by expressions, the design or composition of the product, or by the circumstances surrounding the distribution of a product.[14]

As described above, there are unique characteristics of GenAI that, as part of a product's design without adequate risk controls, can introduce uncertainty in the product's output and can make it difficult to determine the bounds of a product's intended use, and therefore, whether it meets the definition of a device and is the focus of FDA's device regulatory oversight. The unique characteristics of GenAI also may impact and introduce uncertainty to how FDA's existing policies could apply to GenAI-enabled products, such as the policies outlined in FDA's guidances, Changes to Existing Medical Software Policies Resulting from Section 3060 of the 21st Century Cures Act and Policy for Device Software Functions and Mobile Medical Applications, which describe FDA's approach to certain software functions that are excluded from the definition of a device or to those software functions that are medical devices and whose functionality could pose a risk to a patient's safety if the device were to not function as intended. In their current form, these policies are scoped to certain, specific intended uses and products that the Agency understands to be low risk, and there may be uncertainty whether GenAI-enabled products are within the scope of these policies. Therefore, as with any new technology, consideration of the benefit and risk of implementing GenAI for such uses is important.

For GenAI-enabled devices that *are* the focus of FDA's regulatory oversight, the unique characteristics of these devices may also impact their classification and the type of evidence FDA reviewers may need in a marketing submission to reasonably evaluate the safety and effectiveness of these devices. As with all devices, new intended uses will be subject to existing regulatory approaches, e.g., through the De Novo or Premarket Approval pathways. However, GenAI-enabled devices that introduce new intended uses, such as diagnosis across broad areas of medicine, generation of new information and images, utilization of increased levels of automation, or delivery of health care services directly to patients with reduced role for, or no oversight from, health care professionals, will create the need for new types of valid scientific

[12] See section 201(h) of the Federal Food, Drug, and Cosmetic Act.
[13] See FDA's website on How to Determine if Your Product is a Medical Device.
[14] See 21 CFR 801.4.

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

evidence for, or approaches to, their evaluation and FDA will need new evaluation strategies to understand GenAI applied in medical devices broadly.

Digital Health Policies and Premarket Pathways Applied to GenAI-Enabled Products

As described in FDA's guidance Policy for Device Software Functions and Mobile Medical Applications, FDA intends to apply its regulatory oversight to those software functions that are medical devices and whose functionality could pose a risk to a patient's safety if the device were to not function as intended. Broadly, this same approach may be reasonable to apply to GenAI-enabled products. For example, as with all software, GenAI-enabled products appear to exist on a spectrum from those that are not devices and are not within FDA's regulatory purview to those that are devices and are regulated by FDA. This spectrum of GenAI-enabled products could potentially include:

- Certain GenAI-enabled products that may not meet the definition of a device;
- Certain GenAI-enabled products that may meet the definition of a device, but are products for which FDA intends to exercise enforcement discretion; and
- Certain GenAI-enabled products that meet the definition of a device and are devices for which FDA intends to focus its regulatory oversight.

Certain GenAI-enabled products may not meet the device definition. For example, similar to an example, in FDA's guidance Policy for Device Software Functions and Mobile Medical Applications, certain GenAI-enabled products that are intended for health care professionals to use as educational tools for medical training or to reinforce training previously received are likely not devices. Other GenAI-enabled products may not meet the device definition if they are excluded from the definition of a device pursuant to section 520(o) of the FD&C Act. For example, certain GenAI-enabled products that are solely intended for administrative support of a health care facility, or those intended for maintaining or encouraging a healthy lifestyle and are unrelated to diseases or conditions, are likely not devices.

Other GenAI-enabled products may meet the definition of a device, but FDA may intend to exercise enforcement discretion for these devices because they pose lower risk to the public (meaning FDA does not, at this time, intend to enforce requirements under the FD&C Act). For example, certain GenAI-enabled products that automate simple tasks for health care professionals, or that generate health reminders or tracking tools for patients with chronic diseases, which are similar to examples described in FDA's guidance Policy for Device Software Functions and Mobile Medical Applications, may be devices, but FDA may intend to exercise enforcement discretion for these devices.

Finally, many GenAI-enabled products may be the focus of FDA's regulatory oversight as devices because the software meets the definition of a device and its functionality could pose a risk to a patient's safety if the software were to not function as intended. For example, similar to an example in FDA's guidance Policy for Device Software Functions and Mobile Medical

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

Applications, GenAI-enabled products that perform patient-specific analysis and provide specific output(s) or directive(s) to health care professionals for use in the diagnosis, treatment, mitigation, cure, or prevention of a disease or condition or that perform patient-specific analysis and provide patient-specific diagnosis or treatment recommendations to patients, caregivers, or other users who are not health care professionals are devices that are the focus of FDA's regulatory oversight.

The FDA's oversight of devices is risk-based, which means that the level of regulatory controls necessary to demonstrate a reasonable assurance of safety and effectiveness is typically matched to the level of risk of the device. Each device is assigned to one of three regulatory classes, Class I, Class II, or Class III, where each class has increasing levels of regulatory control necessary to provide reasonable assurance of device safety and effectiveness. The extent to which risks can be mitigated can affect the level of regulatory control for a particular product.

Generally, medical devices, including those that incorporate GenAI, may be subject to premarket review under one of the following regulatory pathways based on the device's classification and the degree of risk they present[15]:

- Premarket Approval (PMA), when the device is high risk
- De Novo Classification Request, when the device is low to moderate risk and there is no legally marketed predicate device
- Premarket Notification [510(k)], when the device is low to moderate risk and there is a legally marketed predicate device

Regardless of the type of premarket pathway – PMA, De Novo, or 510(k) – the principles of safety and effectiveness underlie FDA's review of all medical devices. The lowest risk devices are subject to general controls (like other devices reviewed through the above regulatory pathways), but are generally exempt from premarket review.

As noted above, it is important for FDA to continue to apply a risk-based approach to devices, including GenAI-enabled devices. However, some of the characteristics associated with GenAI-enabled products may introduce unique risks in comparison to other software-enabled devices and thus, may impact FDA's understanding of and approach to regulating GenAI-enabled products.

Challenges for GenAI-Enabled Devices and a TPLC Regulatory Approach

When considering the unique risks associated with introducing GenAI in medical devices, FDA faces two broad categories of regulatory challenges: first, there are challenges associated with

[15] See FDA's website on Medical Device Safety and the 510(k) Clearance Process.

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

applying a risk-based approach to classification and determining regulatory requirements for GenAI-enabled devices; and, second, for those GenAI-enabled devices that require FDA's regulatory oversight, there are challenges associated with determining the types of valid scientific evidence ("evidence")[16] for FDA's evaluation of the safety and effectiveness of GenAI-enabled devices across the TPLC.

Considerations when Applying a Risk-Based Approach to GenAI-Enabled Devices

As mentioned previously, a unique characteristic of GenAI is that GenAI models can generate "new information," and some implementations may even intentionally leverage this capacity to generate more "creative" responses. However, this may also make the GenAI model prone to hallucinations that may be difficult to identify or explain. For FDA and device users, hallucinations produced by a GenAI-enabled device can introduce uncertainty in the device's behavior, which can translate to difficulty in understanding the specific bounds of a device's intended use. For example, for a GenAI-enabled product that may be meant to summarize a patient's interaction with a health care professional, the possibility of that product hallucinating can present the difference between summarizing a healthcare professional's discussion with a patient and providing a new diagnosis that was not raised during the interaction.

Further, and as discussed earlier in this executive summary, foundation models and other GenAI tools are often designed to perform a wide range of tasks without being tailored to a specific task or use case. As such, these foundation models and other GenAI tools generally are not meant to have a specific medical device intended use. However, if a device manufacturer uses a foundation model or other GenAI tool as part of a product with a specific intended use that meets the device definition, the product that leverages the foundation model may be the focus of FDA's device regulatory oversight. In such circumstances, it may be challenging for those device manufacturers incorporating foundation models and other GenAI tools into their GenAI-enabled device to obtain detailed information about the foundation model, such as the model's attributes, architecture, training methodology, and/or datasets, given these models can be "locked" or "unlocked[17]" depending on how the foundation model developer chooses to deploy the model and to provide access to users.

Additionally, foundation models are generally provided with large amounts of data that may not be well-controlled. Therefore, the foundation model may be susceptible to bias that may be especially difficult for individual product developers to identify or mitigate for their resulting GenAI-enabled products. Limited visibility into off-the-shelf (OTS) software incorporated into medical devices already presents potential trade-offs for manufacturers. While the use of OTS software can allow the device manufacturer to focus on what is needed to run device-specific functions, OTS software may not always be appropriate for a given specific use in a medical

[16] For purposes of this executive summary, the terms "valid scientific evidence" and "evidence" are used interchangeably.
[17] The terms "unlocked model", "continual machine learning", and others can be used interchangeably.

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

device. As noted in the guidance, Off-The-Shelf Software Use in Medical Devices: "The medical device manufacturer using OTS Software generally gives up software life cycle control, but still bears the responsibility for the continued safe and effective performance of the medical device." In the context of GenAI-enabled devices, an analogous lack of software lifecycle control over an incorporated foundation model may raise certain challenges.

For example, in a scenario where the device manufacturer may have very limited control of the foundation model, changes in the overall data composition in the foundation model could impact the intended use of, or could lead to performance bias in, the GenAI-enabled device, which may lead to further uncertainty in the device's behavior. Importantly, FDA has also yet to authorize any unlocked AI-enabled device. Thus, not only can the application of third-party foundation models lead to uncertainty in a device's behavior due to a manufacturer's own limited visibility into the model's development, if incorporated "unlocked", third-party foundation models may raise further new and significant questions during FDA's review.

To summarize, as with all devices, understanding a GenAI-enabled device's intended use and full functionality is important so that FDA can make a determination of how to apply our regulatory oversight and ensure device safety and effectiveness. Further, even as these devices may present new challenges for FDA review, the bounds of such intended use and functionality will need to be adequately demonstrated to support FDA's risk-based approach to regulatory oversight.

Impact on FDA's Digital Health Policies

FDA's various guidances on digital health and software help manufacturers determine whether a product, including a GenAI-enabled product, is the focus of FDA's device regulatory oversight. However, given the aforementioned characteristics regarding GenAI and related uncertainty, it can be difficult to understand whether the product is within the scope of FDA's digital health policies.[18]

For example, consider a mobile application that is a non-GenAI-enabled product, which is intended to prompt a user to enter which herb and drug they would like to take concurrently and provide information about whether interactions have been seen in the literature and a summary of what type of interaction was reported. As described in FDA's guidance, Policy for Device Software Functions and Mobile Medical Applications, such software is generally software for which FDA intends to exercise enforcement discretion. That guidance also notes that this example, as well as other examples in that section of the guidance, are examples of software functions that may meet the definition of a medical device but for which FDA intends to exercise enforcement discretion because they pose lower risk to the public. Consider if, instead, the same product was GenAI-enabled. While the GenAI-enabled product may still be intended for the same use, if not sufficiently controlled, it is possible for the product to generate output that is 1) beyond its intended use with newly apparent device functions (e.g., directing patients to treatment using an alternative herb/drug combination due to interactions identified regarding the

[18] See FDA's Guidances with Digital Health Content.

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

queried herb/drug combination) or 2) erroneous or false content regarding interactions of the queried herb/drug combination. In the latter case, while such a GenAI-enabled product may still not meet the definition of a device, it is important to consider whether the implementation of GenAI, which generates new content and information, introduces potential uncertainty and/or risk in a product where the intended use is to provide patients with access to known information. To apply this concept more generally, it is important to consider whether the implementation of GenAI introduces potential uncertainty and/or risk in a product that would otherwise, without GenAI, have been considered low risk and not the focus of FDA's regulatory oversight.

In many cases, whether or not a product is enabled with GenAI may not change considerations of FDA's oversight. However, in certain cases, e.g., if the GenAI model is insufficiently controlled, GenAI-enabled products may behave outside the stated intended use to function in a way that would be the focus of FDA's device regulatory oversight. At times, it may be helpful for manufacturers and developers to consider that a GenAI implementation of a product, in comparison to a non-GenAI implementation, may not be beneficial to public health when it could be providing erroneous or false content, spreading misinformation. Therefore, it is helpful for manufacturers and developers to consider the potential uncertainty and/or risk of implementing a specific GenAI model for the specific intended use, or in other words, to consider when GenAI may or may not be the best technology for a specific intended use.

As with other technologies applied in health care or other high-risk disciplines, it is especially important to continue to seek to apply the "right tool" for each task. When a manufacturer has determined that there is benefit to developing a GenAI-enabled device, it will be important for manufacturers to consider the evidence that may need to be generated for such an implementation for FDA's evaluation of reasonable assurance of safety and effectiveness of the device. The generation of such evidence for a GenAI-enabled device may be challenging. For example, it may be difficult to develop an accurate device description or characterization of the GenAI-enabled device if little is known about the base foundation model. There may also be challenges to generate premarket and postmarket performance data. Later in this executive summary, we describe some of the challenges that FDA foresees related to evidence generation for FDA's evaluation of safety and effectiveness for GenAI-enabled devices.

Impact on Device Classification

FDA's oversight of devices is risk-based, and many GenAI-enabled products that meet the definition of a device will be the focus of FDA's device regulatory oversight. However, as discussed above, there may be challenges to applying FDA's current risk-based approach to classification and determining regulatory requirements for a GenAI-enabled product when such products may have difficult to constrain or intentionally broad intended uses or may present uncertain performance.

As previously mentioned, GenAI is generally intended to output new data meant to resemble the data it learned from rather than being specifically trained to recall information or make predictions. Further, it could be *designed* to produce variable outputs. Both of these aspects of GenAI can lead to hallucinations -- making it challenging to demonstrate that a GenAI-based

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

product has a clearly constrained intended use and will provide accurate, consistent, and reliable outputs. FDA needs to have an understanding of the device's intended use and technological characteristics to make a determination of how to apply our regulatory oversight, specifically, how to apply our risk-based approach to device classification to ensure device safety and effectiveness.

GenAI-enabled devices, like AI-enabled devices broadly, can also challenge FDA's current device classification schema because the technological characteristics of GenAI may sometimes introduce new or different risks for a particular GenAI-enabled product, which raises new questions of safety and effectiveness that may impact a device's classification or premarket pathway, as well as the kinds of regulatory controls that may be necessary to ensure such devices are and will remain safe and effective. As it pertains to impact on a device's classification or regulatory pathway, this is most evident in FDA's largest premarket program, our 510(k) Program. As described in section 513(i) of the FD&C Act, for devices subject to 510(k) requirements, the determination of substantial equivalence includes, among other requirements, a comparison between the intended uses and technological characteristics of the predicate device and the subject device. Assuming a circumstance where a non-GenAI-enabled device (the predicate device) has the same intended use as a GenAI-enabled device (the subject device), based on the technological characteristics of GenAI and the uncertainty introduced, it is likely that the GenAI-enabled device could be found to have different technological characteristics than the predicate device, and that those different technological characteristics raise different questions of safety and effectiveness.

There are many types of regulatory or risk controls that can be helpful to ensure GenAI-enabled devices are and will remain safe and effective. It will be helpful for manufacturers to consider if and how they can control the underlying GenAI model for the GenAI-enabled device, to the extent possible, even in circumstances where its underlying foundation model may be outside of their control. Additionally, other risk controls, such as appropriate governance of the GenAI model, utility of appropriate feedback mechanisms regarding device safety, and real-world performance evaluation, can help provide assurance that the GenAI-enabled device is performing safely and effectively once in real-world use. All medical devices are subject to general controls, unless exempt, including, but not limited to medical device reporting (21 CFR Part 803), reports of corrections and removals (21 CFR Part 806), establishment registration and device listing (21 CFR Part 807), and quality system regulation (21 CFR Part 820).[19] Beyond general controls, for certain Class II devices, FDA may require special controls unique to GenAI-enabled devices when needed to provide reasonable assurance of safety and effectiveness of the device. Such special controls could include requirements for postmarket monitoring of device safety and performance or notification requirements if the GenAI-enabled device is not performing as intended.

While there may be challenges in applying FDA's current risk-based approach to GenAI-enabled products, with an appropriate understanding of the design and control of the GenAI

[19] See FDA's website on Regulatory Controls.

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

model, and its underlying foundation model, and the intended use and technological characteristics of the GenAI-enabled product, FDA can make a determination of how to apply our regulatory oversight and ensure device safety and effectiveness. Additionally, it will remain important for manufacturers and the entities that use these products to consider the risk management strategies that may be important to help ensure these devices remain safe and effective in real-world use, beyond the scope of explicit regulatory controls that FDA may establish.

For those GenAI-enabled products that are the focus of FDA's regulatory oversight as devices, there is also a second challenge pertaining to generation and review of new types of valid scientific evidence for GenAI-enabled devices over the TPLC.

Considerations Regarding Valid Scientific Evidence for GenAI-Enabled Devices over the TPLC

FDA reviews many types of valid scientific evidence as part of its determination of reasonable assurance of safety and effectiveness for devices that require FDA's regulatory oversight. Generally, the evidence needs for AI-enabled devices will likely apply to GenAI-enabled devices, but may need to be supplemented with additional or different evidence to ensure device safety and effectiveness. For example, current evaluation approaches, such as those used to evaluate computer-assisted triage, detection, and diagnostic devices, may still be applicable for GenAI-enabled devices, albeit with additional supporting evidence. However, it may be challenging to determine the evidence that may be needed for certain GenAI implementations, in particular, for those GenAI-enabled devices that use open-ended input and output formats that are different than the typically structured format used in other AI-enabled devices, as these implementations may have novel evaluation considerations.

While the specific information provided for a GenAI-enabled device would be governed by its intended use and design, the broad applications of some GenAI-enabled devices can lead to difficulty in bounding a device's stated intended use, which creates challenges in the review process. As such, the level of detail that may be needed in a marketing submission for a GenAI-enabled device regarding the underlying device design and the performance testing requirements may be greater, depending on the specific device. In addition to the details provided about the GenAI model implementation in the device, information about a utilized foundation model may also be needed, similar to the current approach taken for incorporating OTS software into a medical device. This information could include the foundation model's attributes, architecture, and training methods and datasets, which may not be trivial for a manufacturer to obtain. Information on training methodology and datasets for the foundation model may be particularly important because, as noted previously, foundation models are generally provided with large amounts of data that may not be well-controlled. In addition, foundation models have billions of model parameters. Due to the large and complex parameter structure, small changes to the input data can lead to different outputs. For instance, slightly modifying the phrasing of the request can cause a GenAI model to generate different responses or predictions to the feedback requested. Since GenAI models can have such open-ended

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

inputs, it is unreasonable to evaluate every conceivable input that the model might encounter during deployment and is especially unfeasible when third-party foundation models do not openly disclose their parameters. This lack of transparency and the potential for emergent or unanticipated behavior may be particularly challenging to evaluate premarket. To mitigate these associated risks and premarket challenges, and to ensure continued safety and effectiveness of GenAI-enabled devices across the TPLC, it may be particularly important for premarket review to be complemented with effective and tailored postmarket monitoring strategies.

Current Premarket Evidence Needs

As part of FDA's review of a device, FDA needs to have a general understanding of the device and its design.[20] For AI-enabled devices broadly, this includes an understanding of the design specifications, data management and model development information, and characteristics of model. In particular, for GenAI-enabled devices, the types of information related to the model are the same, but the expected details could vary. While this information may not be unique to GenAI-enabled devices, this information may help address some of the considerations for GenAI-enabled devices, in particular, related to demonstrating an adequately bounded intended use and/or acceptable variability in device performance.

For GenAI-enabled devices, design details may include information such as specific design specifications (e.g., attention mechanism, model merging, etc.), model parameters (e.g., temperature, top-K, and top-p, etc.), built-in prompting strategies (e.g., in-context learning, zero-shot prompting, etc.); and prompting capabilities available to the end user. For GenAI-enabled device training, this may include information—as reasonably as possible—pertaining to data management and model development for the initial foundation model, as well as specific details on data management and fine-tuning the underlying GenAI model for the specific GenAI-enabled device. The steps to fine-tune that result in a specific adaptation of the GenAI-enabled device is important to understand since it may govern the device's intended use.

Finally, while device output characteristics are important for all AI-based devices, they can be particularly important for GenAI-enabled devices, which, as described, can output new information that is variable by design. It is important to ensure transparency to the end user for GenAI-enabled devices, which can have an increased level of complexity compared to traditional devices, and even some AI-enabled devices.[21] This may include a variety of transparent information for the end user on the GenAI-enabled device, such as the device design, how the device was tested, the level of autonomy, and how users interact with the device (e.g., through prompt engineering), which may impact the device output. It is also important to understand the level of autonomy for a GenAI-enabled device, and if, and how, it incorporates human-in-the-loop and affords the level of control to the end user.

While, as noted, many of these evidence needs are not unique to GenAI-enabled devices, robust design specifications and output characteristics may help FDA and users better

[20] See, e.g., 21 CFR 807.87, 21 CFR 860.220, or 21 CFR 814.20.
[21] See also Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles.

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

understand unique considerations for GenAI-enabled devices that may impact the bounds of the intended use and device performance.

New Methodologies for Premarket Evidence Needs

For some GenAI-enabled devices, the current methodologies for performance evaluation may still apply. Additionally, use of methodologies and strategies such as prompt engineering, quality control on types of inputs and/or outputs, and use of topic-specific foundation models, can help so that current, quantitative performance evaluation methodologies may be used to evaluate a particular GenAI-enabled device.

However, new methodologies may also need to be developed to evaluate the performance of GenAI-enabled devices broadly to ensure that they remain safe and effective. Current quantitative evaluation approaches may not provide a thorough or complete assessment of device performance and may need to be complemented with additional performance metrics. For example, new approaches for qualitative performance evaluation of GenAI-enabled devices may help to accurately characterize the underlying GenAI model of the GenAI-enabled device as it pertains to the level of autonomy, transparency, and explainability. Additionally, such new approaches may also be helpful to evaluate the performance of GenAI-enabled devices that have broad applications across medical disciplines.

The performance evaluation methodologies needed, however, would be governed by the specific intended use and design of the GenAI-enabled device, some of which may necessitate formulation of new performance metrics for certain intended uses. As with all devices, the totality of evidence, which may include premarket and postmarket evidence, can support reasonable assurance of safety and effectiveness of these devices across the TPLC.

New Postmarket Evidence Needs

Beyond premarket evidence, some of which may be generated using new methodologies, postmarket evidence may also play a role in ensuring the safety and effectiveness of GenAI-enabled devices, and AI-enabled devices in general. Broadly, AI-enabled devices have the potential to undergo continuous adjustment based on localized live data, user interactions, and changing conditions. For GenAI-enabled devices, as previously described, the foundation models that may serve as their basis may be further susceptible to changing conditions. In comparison to a framework for monitoring and continuous adjustment for AI-enabled devices broadly, it could be particularly challenging to provide similar oversight and control for such GenAI-enabled devices. For example, for some GenAI-enabled devices, the device manufacturers may lack control of and visibility into the data for the underlying foundation model that may impact their device. It will be critical to monitor and evaluate AI-enabled devices, including GenAI-enabled devices, in the postmarket space in order to maintain device safety and effectiveness. Further, it is important to consider how challenges related to effective evaluation and monitoring may evolve, for example, as manufacturers begin to propose use of unlocked models, device development utilizing third-party foundation models, increasingly complex multi-layer designs, and site- or patient-specific models in their AI-enabled or devices.

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

Broadly, manufacturers may consider robust postmarket monitoring strategies as a mechanism to monitor the safety and performance of AI-enabled and GenAI-enabled devices in a proactive manner, which can also help provide FDA with assurance of the safety and effectiveness of the device over time. For example, as described in FDA's guidance Consideration of Uncertainty in Making Benefit-Risk Determinations in Medical Device Premarket Approvals, De Novo Classifications, and Humanitarian Device Exemptions, "FDA also considers the appropriateness of risk mitigations and the collection of postmarket data to address the uncertainty in the benefit-risk information," which can help mitigate premarket uncertainty. Furthermore, the guidance states that "the continuous, robust generation of evidence throughout the premarket and postmarket setting as part of a learning health care system is important to continuously refine our understanding of how medical devices are used and perform, and corresponding patient outcomes."

Therefore, as with all devices, it will be important for manufacturers to implement postmarket monitoring strategies for their GenAI-enabled devices. In particular, it will be important for manufacturers to consider how to effectively evaluate and monitor the GenAI-enabled device for its specific intended use in a way that can ensure the device accuracy, relevance, and reliability is maintained, once deployed. It is important to note that postmarket monitoring strategies can become more complex if there are various implementations of the GenAI-enabled device, for example, site-specific implementations for each GenAI-enabled device. Finally, it may be important for manufacturers to consider postmarket monitoring strategies for not only their GenAI model, but the underlying foundation model, as applicable, as changing conditions for both the foundation model and the GenAI model can ultimately impact the GenAI-enabled device safety and performance. For certain AI-enabled or GenAI-enabled devices, FDA may require postmarket monitoring as a regulatory control, depending on the intended use and risk of the device. To ensure the continued safety and effectiveness of AI-enabled and GenAI-enabled devices, it may be important for premarket evidence to be complemented by postmarket evidence gathered through robust device performance monitoring approaches.

In Summary

FDA's CDRH is committed to assuring that patients and providers have timely and continued access to safe, effective, and high-quality medical devices. As part of this mission, CDRH facilitates medical device innovation by advancing regulatory science, providing industry with predictable, consistent, transparent, and efficient regulatory pathways, and assuring consumer confidence in devices marketed in the United States.

The novel capabilities of GenAI may offer unique benefits to patients and public health, but the use and adoption of GenAI also come with specific risks and complexities that challenge FDA's approach to the regulation of devices. In particular, FDA faces challenges associated with applying a risk-based approach to classification and determining regulatory requirements for GenAI-enabled devices and, for those GenAI-enabled devices that require FDA's regulatory oversight, FDA faces challenges associated with determining the types of valid scientific

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

evidence for FDA's evaluation of the safety and effectiveness of GenAI-enabled devices across the TPLC. FDA has long promoted a TPLC approach to the oversight of medical devices, including AI-enabled devices, and has committed to advancing regulatory approaches for these devices. As the interest in GenAI tools across the health care sector has expanded rapidly, FDA notes that there remain open questions on the approach to regulating GenAI-enabled products that may fall under the purview of FDA's regulatory jurisdiction and that it is of public health importance that the Agency work with experts to address these questions in a timely manner.

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

Panel Questions

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 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, for example text/image models where either inputs, outputs or both could be multi-modal? Performance metrics will typically vary with device intended use. Examples of known metrics to support discussion may also 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

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

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 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

Appendix

The AI Lifecycle described briefly in this executive summary, and provided to help drive discussion around TPLC needs for AI-enabled devices, including GenAI-enabled devices, can be further expanded to a list of technical considerations associated with each lifecycle phase as illustrated in Figure 2. For purposes of this executive summary, these “per-lifecycle phase” sets of technical considerations are intended to be representative of such lifecycle phases for AI-enabled software broadly, but also for GenAI-enabled software.

Considerations of the AI Lifecycle can provide AI- and GenAI-enabled device manufacturers (as well as regulators, customers, and clinicians, etc.) a practical framework for the development of and assessment of adherence to AI development best practices, tooling, applicable metrics, and standards through the entire lifecycle of a device. Such a framework can provide a structured approach to compliance and quality considerations that lifecycle phase-appropriate owners and parties can apply and document by way of ensuring device quality. For example, a GenAI-enabled device manufacturer may have limited visibility on the third-party foundational model design and implementation incorporated into its device (as discussed in the executive summary). The AI Lifecycle may help to develop, deliver, and operationalize GenAI-enabled devices, beginning from the earliest stages of their development, drawing attention to phases and considerations more particular to GenAI (such as data management, model training and testing, model validation and verification, etc.) as well as more general considerations, such as security, risk management, and operationalization considerations, concerning which GenAI may present its own unique challenges. The technical considerations are not intended to be exhaustive, nor are they intended as requirements or recommendations for any regulated device.

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

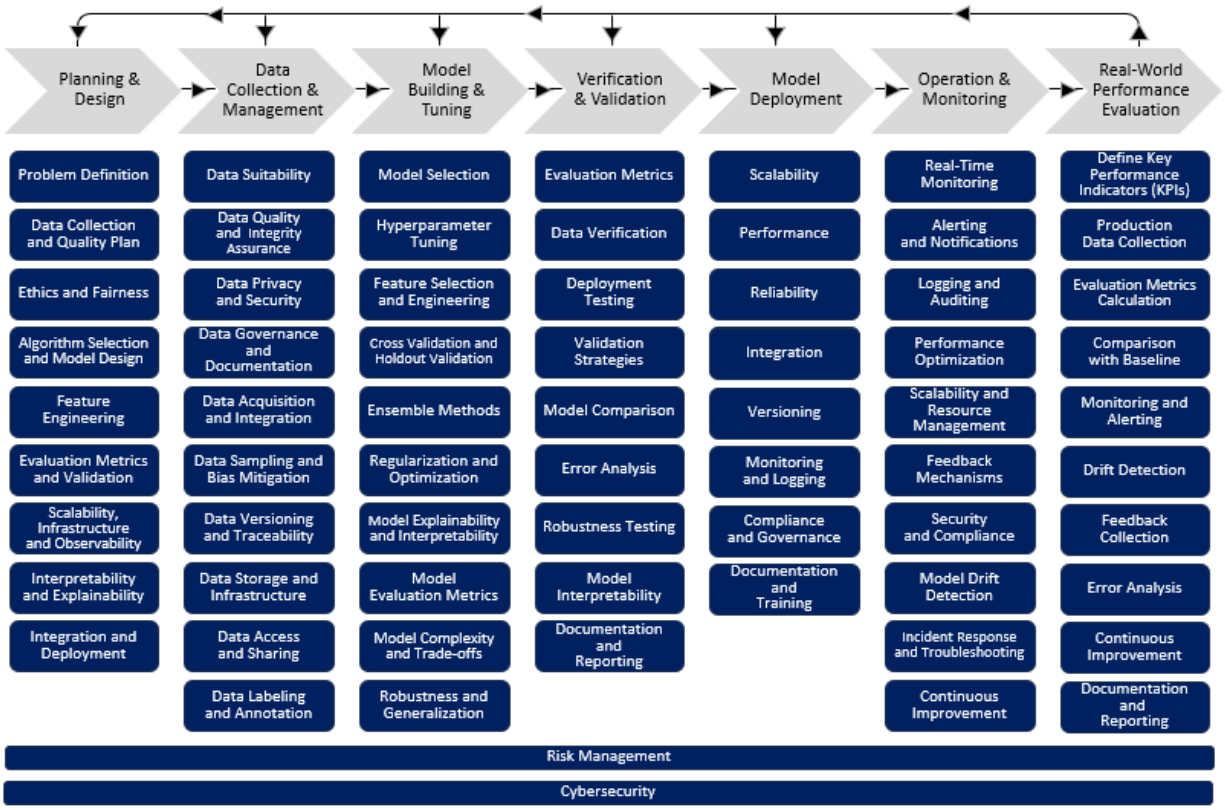

Figure 2: AI Lifecycle Considerations

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

Glossary

TERM	DEFINITION
Bias	Systematic difference in treatment of certain objects, people, or groups in comparison to others. Note 1 to entry: Treatment is any kind of action, including perception, observation, representation, prediction or decision. (ISO/IEC TR 24027:2021) *Source: International Medical Device Regulators Forum. (2022). Machine Learning-enabled Medical Devices: Key Terms and Definitions*
Artificial Intelligence (AI)	A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action. *Source: Executive Order 14110, (October 30, 2023), Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*
Artificial Intelligence Performance Monitoring (AI Performance Monitoring)	Refers to the process of regularly collecting and analyzing data on the use of a deployed AI system to evaluate its performance in accomplishing its intended tasks in real-world settings. The assessment of an AI model's performance involves various performance metrics and criteria depending on the specific application. This monitoring typically aims to assess the performance of these AI systems in practice, detect performance degradation or changes (e.g., due to data drift), identify instances of misuse, and address any safety or usability concerns. *Source: DH/AI Glossary*
Artificial Intelligence System (AI System)	Any data system, software, hardware, application, tool, or utility that operates in whole or in part using AI. *Source: Executive Order 14110, (October 30, 2023), Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*
Continual Machine Learning	The ability of a model to adapt its performance by incorporating new data or experiences over time while retaining prior knowledge/information. The model changes are implemented such that for a given set of inputs, the output may be different before and after the changes are implemented. These changes are typically implemented and validated through a well-defined process that aims at improving performance based on analysis of new data. In contrast to a locked model, a continual machine learning model has a defined learning process to change its behavior. *Source: DH/AI Glossary*

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

Data Drift	Refers to the change in the input data distribution a deployed model receives over time, which can cause the model's performance to degrade. This occurs when the properties of the underlying data change. Data drift can affect the accuracy and reliability of predictive models. For example, medical AI-enabled products can experience data drift due to, statistical differences between the data used for model development and data used in clinical operation due to variations between medical practices or context of use between training and clinical use, and changes in patient demographics, disease trends, and data collection methods over time. *Source: DH/AI Glossary*
Device	An instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is: (A) recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them, (B) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or (C) intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes. The term "device" does not include software functions excluded pursuant to section 520(o). *Source: Section 201(h)(1) of the Federal Food, Drug, and Cosmetic Act*
Explainability	"Refers to a representation of the mechanisms underlying AI systems' operation." (Source: NIST) Explainability may help overcome the opaqueness of black-box systems (i.e., systems where the internal workings and decision-making processes are not transparent or readily understandable). These explanations can take various forms, including free-text explanations, saliency maps, SHapley Additive *exPlanations (*SHAP), or relevant input examples from data. The primary intent is to answer the question "Why" an AI system made a particular decision. Appropriate Explainable AI (XAI) methods may enable the development of more accurate, fair, interpretable, and transparent AI systems to safely augment human decision-making in healthcare**.** *Source: DH/AI Glossary*
Foundation Models	AI models trained using large, typically unlabeled datasets and significant computational resources, that are applicable across a wide range of contexts, including some that the models were not specifically developed

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

	and trained for (i.e., emergent capabilities). These models can serve as a foundation upon which further models can be built and adapted for specific uses through further training (i.e., fine-tuning). These models can perform a range of general tasks, such as text synthesis, image manipulation, and audio generation. These models are based on deep learning architectures like transformers and can use either unimodal or multimodal input data. *Source: DH/AI Glossary*
Generative Artificial Intelligence (Generative AI)	"The class of AI models that emulate 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) This is usually done by approximating the statistical distribution of the input data. For example, in healthcare, generative AI can be used to generate annotations on synthetic medical data (e.g., image features, text labels) to help expand datasets for training algorithms. *Source: DH/AI Glossary*
Confabulation (or Hallucination)	Refers to a phenomenon in which generative AI systems generate and confidently present erroneous or false content to meet the programmed objective of fulfilling a user's prompt. *Source: National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf*
Intended Use	Refers to the objective intent of the persons legally responsible for the labeling of an article (or their representatives). The intent may be shown by such persons' expressions, the design or composition of the article, or by the circumstances surrounding the distribution of the article. This objective intent may, for example, be shown by labeling claims, advertising matter, or oral or written statements by such persons or their representatives. Objective intent may be shown, for example, by circumstances in which the article is, with the knowledge of such persons or their representatives, offered or used for a purpose for which it is neither labeled nor advertised; provided, however, that a firm would not be regarded as intending an unapproved new use for a device approved, cleared, granted marketing authorization, or exempted from premarket notification based solely on that firm's knowledge that such device was being prescribed or used by health care providers for such use. The intended uses of an article may change after it has been introduced into interstate commerce by its manufacturer. If, for example, a packer, distributor, or seller intends an article for different uses than those intended by the person from whom he or she received the article, such packer, distributor, or seller is required to supply adequate labeling in accordance with the new intended uses. *Source: https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-801/subpart-A/section-801.4*

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

Large Language Model (LLM)	A type of AI model trained on large text datasets to learn the relationships between words in natural language. These models can apply these learned patterns to predict and generate natural language responses to a wide range of inputs or prompts they receive, to conduct tasks like translation, summarization, and question answering. These models are characterized by a vast number of model parameters (i.e., internal learned variables within a trained model). LLMs build on foundational AI models by developing more comprehensive language understanding beyond basic linguistic patterns. For example, in the context of LLMs, chatbot is a program that enables communication between the LLM and the human through text or voice commands in a way that mimics human-to-human conversation. *Source: DH/AI Glossary*
Locked Model	A model that provides the same output each time the same input is applied to it and does not change with use, as its parameters or configuration cannot be updated. In case of AI-enabled medical products, locked models can help ensure consistent performance. *Source: DH/AI Glossary*
Machine Learning (ML)	A set of techniques that can be used to train AI algorithms to improve performance at a task based on data. *Source: Executive Order 14110, (October 30, 2023), Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*
Machine Learning Model (ML Model)	A mathematical construct that generates an inference or prediction for input data. This model is the result of an ML algorithm learning from data. Models are trained by algorithms, which are step-by-step procedures used to process data and derive results. AI systems (e.g., AI-enabled medical devices) employ one or more models to achieve their intended purpose. *Source: DH/AI Glossary*
Model Weight	A numerical parameter within an AI model that helps determine the model's outputs in response to inputs. *Source: Executive Order 14110, (October 30, 2023), Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*
Neural Network	A computational model inspired by the structure of the human brain. It is composed of interconnected nodes, or "neurons" organized into layers: an input layer that receives data, one or more hidden layers that process and identify patterns in the data, and an output layer that presents the final network output. *Source: DH/AI Glossary*
Performance Metrics	In the context of AI quantitative or qualitative measures that can be used to assess the ability of a model to produce the desired output for a given task. The choice of the metrics depends on the specific task and the model objectives.

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

	Examples of quantitative metrics include accuracy, precision, sensitivity (recall), specificity, F1-score, and Area under the Receiver Operating Characteristic curve (AUC-ROC). Qualitative measures may involve heatmap evaluations or visual interpretations. These metrics enable systematic evaluation, comparison, and refinement of models, and aid in the assessment of whether the model meets its intended objectives. *Source: DH/AI Glossary*
Total Product Lifecycle (TPLC)	An integrated device review, tracking, reporting and compliance scheme employed by FDA. The TPLC approach allows FDA to integrate all regulatory activities from device inception to obsolescence. (Source: Infusion Pump: Glossary | FDA) For purposes of this document, the TPLC approach addresses all phases in the life of a medical device, from the initial conception to final decommissioning and disposal.
Training Data	These data are used by the manufacturer of an AI system in procedures and training algorithms to build an AI model, including to define model weights, connections, and components. *Source: DH/AI Glossary*
Transparency	Describes the degree to which appropriate information about a Machine Learning-Enabled Medical Device (MLMD, including its intended use, development, performance and, when available, logic) is clearly communicated to relevant audiences. *Source: U.S. Food and Drug Administration (2024). Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles. https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles*
Watermarking	The act of embedding information, which is typically difficult to remove, into outputs created by AI—including into outputs such as photos, videos, audio clips, or text—for the purposes of verifying the authenticity of the output or the identity or characteristics of its provenance, modifications, or conveyance. *Source: Executive Order 14110, (October 30, 2023), Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*

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