



**U.S. Food and Drug Administration
Center for Devices and Radiological Health**

**Digital Health Advisory Committee (DHAC) Meeting on the topic
“Total Product Lifecycle Considerations for Generative-AI Enabled Devices”
Summary Minutes for Day 2 - November 21, 2024**

Introduction:

On November 21, 2024, the Digital Health Advisory Committee convened to discuss regulatory oversight of medical devices incorporating generative artificial intelligence (GenAI). The meeting focused on the unique challenges and considerations for GenAI-enabled devices, particularly in the context of postmarket performance monitoring.

Conflict of Interest Statement

Mr. James Swink, Designated Federal Officer for the DHAC, read the Conflict of Interest Statement and stated that all members and consultants of the Committee are subject to federal conflict of interest laws and regulations, and that the FDA has determined that all of them are in compliance with such laws. No conflict-of-interest waivers were issued.

Day 1 Recap: Premarket Performance Evaluation and Risk Management

An overview of the discussions from day one was presented, summarizing the Committee's discussion of premarket performance evaluation and risk management strategies for GenAI-enabled devices. The recap of day 1 noted that the prior day's discussion included recurring themes on the necessity of trust, transparency, and equity in developing and deploying GenAI technologies.

Open Public Hearing

Dr. Robert Steinbrook of Public Citizen welcomed the FDA's broad review of GenAI devices. He emphasized the importance of requiring transparency whenever GenAI is used, ensuring training datasets represent the populations served, and establishing strong postmarket monitoring. He also stated his position that some GenAI-enabled devices should be presumptively Class III or subject to new, more rigorous approval pathways. Dr. Steinbrook posed that companies concealing harms could face serious legal consequences, and he underscored the need for patient-friendly explanations of GenAI tools, calling for them to not be “black boxes.”

Dr. Frederick Chen of the American Medical Association (AMA) noted that while AI can enhance diagnostics and improve workflows, most physicians remain cautious. He stressed the need for developers to share details on data sources, training methods, and known biases, and he called for robust validation and postmarket surveillance. Dr. Chen underscored that AI should support, rather than replace, clinical judgment and highlighted the critical importance of protecting patient data and privacy.



Dr. Anil Bhatta and Dr. Sandhya Polu of Deloitte Consulting discussed the potential of GenAI but noted its unique risks, including difficulties defining “error,” malicious prompting (“jailbreaking”), and biases or overreliance. They recommended a “trustworthy AI” framework with techniques like domain-specific fine-tuning, prompt shields, and ongoing bias mitigation. They concluded that human-centered design and active oversight across the product lifecycle are crucial to safe and effective GenAI use in healthcare.

Dr. Yujan Shrestha of Innolitics proposed incentivizing developers to perform thorough postmarket surveillance. He suggested systems in which synthetic data and automated ground truthing detect performance issues and model drift, reducing premarket burdens. Dr. Shrestha proposed that aligning surveillance with manufacturers’ goals fosters continuous improvements, benefiting both patients and companies.

Dr. Keith Dreyer of Mass General Brigham explained that radiology’s decade-long experience with AI highlights the need for site-specific validation and ongoing monitoring. He stated that many cleared AI devices still lack sufficient real-world oversight. Dr. Dreyer proposed shifting away from purely comparative studies toward continuous clinical validation, emphasizing that technologies change over time and must be re-verified. He added that registries, such as those from the American College of Radiology, can facilitate consistent assessment of AI performance in clinical settings.

Ms. Kerri Haresign of the Consumer Technology Association (CTA) highlighted CTA’s role in creating technology standards and hosting events like CES. She mentioned the standard CTA-2125, which outlines best practices for AI information disclosure, and introduced CTA’s Health AI Planning Council, currently focused on predictive AI but expanding to generative AI. Ms. Haresign encouraged engagement with CTA’s upcoming conferences and collaborative projects to advance consensus-driven standards for responsible AI deployment in healthcare.

Dr. Bhatt, the Committee Chair, provided a summary of what was heard during this Open Public Hearing, including and the emphasis on the critical role of postmarket monitoring in addressing real-world challenges associated with GenAI-enabled devices. First, she highlighted a call for stronger regulation of consumer health devices that use generative AI, including clear notifications to end users and prompt removal from the market if they pose safety risks. She then summarized the AMA’s focus on four key areas—transparency, validation, workflow integration, and privacy/security—and their support for continuous pre- and postmarket oversight. In addition, she noted that the AMA stressed the importance of allowing patients to opt out of data usage and of maintaining human oversight in higher-risk applications.

Referring to Deloitte’s remarks, Dr. Bhatt recapped the unique complexities of GenAI, such as variable outputs, difficulty defining errors, and the need for substantial computational and human resources. She emphasized their point that broader “trustworthy



AI” frameworks must lead to practical regulatory measures, including monitoring for “jailbreaking” and overreliance. Dr. Bhatt next recalled Dr. Dreyer’s reminder that radiology has long deployed AI without an adequate monitoring mechanism, suggesting today’s discussions could inform broader AI oversight. Finally, she recognized the Consumer Technology Association’s contribution, stressing that many GenAI devices now reach consumers directly, raising new challenges for the FDA and the healthcare community.

Open Committee Discussion Q&A *(Clarification Questions)*

Dr. Thomas Radman asked Dr. Dreyer about the FDA’s role when an AI device performs well at some sites but poorly at others. He questioned whether poor results at one site should prompt a full product recall. Dr. Dreyer likened this to other complex oversight scenarios, suggesting the FDA consider requiring “deployer” responsibilities for local users. He noted that GenAI models work on data rather than directly on patients, thus warranting extended regulatory thinking to include providers and their site-specific validations.

Dr. Chevon Rariy asked Dr. Dreyer about monitoring GenAI systems that evolve post-deployment, pointing out that comparing original drafts to final reports might not capture a model’s continuous learning. Dr. Dreyer responded that monitoring is still needed for each output, even if the device continuously retrains. He suggested a dual strategy of tracking the device’s immediate performance and separately validating any changes in the underlying foundation model.

Dr. Apurv Soni raised concerns about de-prioritizing comparative effectiveness, emphasizing that traditional approaches like randomized controlled trials might be slow, but new methods (e.g., synthetic controls) could assess real-world outcomes. Dr. Dreyer agreed that some form of expanded testing beyond standalone performance is beneficial but argued that rigid comparative effectiveness studies may be inefficient. Dr. Soni suggested leveraging real-world, natural experiments as a middle ground for assessing whether AI tools truly improve care.

Ms. Diana Miller asked about cases where a site’s input data appear invalid or produce poor AI outputs. Dr. Dreyer explained that monitoring typically reveals when changes in scanner protocols or data collection procedures degrade AI performance. He stated that in such scenarios, clinicians either revert to the original data protocol or remove the problematic AI tool from service for that particular use case.

Mr. Steven Posnack proposed the idea of a “prescale” phase between premarket and nationwide deployment where AI devices could be tested at limited scale under controlled conditions. Dr. Dreyer noted that some institutions already do this via “clinical validation” before fully releasing an AI tool into clinical practice, but he agreed that a structured prescale approach could be beneficial. He added that large, shared data repositories might facilitate scaled evaluations without full commercial rollouts.



Dr. Pratik Shah invited Dr. Shrestha and others to comment on site-based validation. Dr. Shrestha proposed a postmarket mechanism where manufacturers continuously verify AI performance using synthetic data and automated ground truthing. He suggested an approach akin to “test-driven development,” potentially incentivized by a flexible FDA clearance process. Participants agreed that site validation could combine with broader “collective intelligence” from multiple sites to prevent narrow assessments.

Dr. Jagdish Khubchandani asked why AI development is concentrated in radiology and cardiology, with limited focus on fields like urology or gynecology. Dr. Shrestha and Mr. Troy Tazbaz explained that radiology data are more standardized (e.g., DICOM) and easier to train, while other specialties face challenges such as inconsistent data formats. Dr. Dreyer emphasized that regulatory hurdles also play a role, potentially limiting innovation in other disciplines.

Dr. Peter Elkin cautioned that relying on a single clinician’s correction of AI output (e.g., final report vs. draft) may not always reflect true “ground truth.” He stressed that multiple experts, follow-up outcomes, or validated outcome measures may be needed for robust evaluation. Dr. Dreyer replied that in radiology, peer review and high case volumes can reduce bias, but acknowledged the need for outcome-based metrics in some areas.

Dr. Thomas Maddox reminded the Panel of real-world constraints, especially in smaller or rural hospitals lacking robust data infrastructure. He noted that heavier regulatory or monitoring requirements could create a “digital divide,” limiting access to AI tools. Dr. Elkin pointed out that emerging health information exchanges and potential federal data initiatives could reduce site burdens by centralizing data, thereby supporting more equitable surveillance.

In closing the Open Committee Discussion Q&A, Dr. Bhatt underscored several themes for continued discussion, including balancing local validation with broader real-world or pragmatic trials, identifying high-risk vs. low-risk errors and tailoring oversight accordingly, recognizing the need for both premarket rigor and flexible postmarket monitoring, and ensuring equitable adoption of AI, especially in resource-limited settings.

Sub-Topic: Postmarket Performance Monitoring

FDA Perspective – Approaches for Managing Changes for GenAI-enabled Devices

Jessica Paulsen, Associate Director for Digital Health in the FDA CDRH Office of Product Evaluation and Quality (OPEQ), outlined the agency’s regulatory approach to AI-enabled devices, emphasizing a Total Product Lifecycle (TPLC) approach for rapidly evolving technologies like GenAI. She noted that while 950 AI-enabled devices have been FDA-authorized, none are generative or adaptive AI devices.



Highlighting an authorized sepsis device software as an example, Paulsen stressed the importance of special controls, such as postmarket performance plans, to address risks like bias and poor data quality. For GenAI devices, the FDA may create new device types under the De Novo process with tailored controls and monitoring requirements.

She described Predetermined Change Control Plans (PCCPs), which allow manufacturers to make pre-specified updates without additional submissions, provided they include detailed modification descriptions and validation plans. Paulsen emphasized transparency, clear modification guidelines, and robust postmarket monitoring to ensure safety and effectiveness as AI devices evolve.

Stakeholder Perspective – Supporting Health AI for Impact

Dr. Christopher Longhurst, Chief Clinical Innovation Officer at UC San Diego, discussed regulatory oversight for GenAI in healthcare, proposing three key recommendations. His first recommendation was to consider the adequacy of FDA’s 510(k) pathway for AI algorithms, including whether increased authority to require transparent outcomes data may be needed. His second recommendation was to expand FDA authority for postmarket surveillance of AI devices, and his third recommendation was to create a regulatory framework with shared responsibility for healthcare delivery, including hospitals and health systems bearing responsibility for local testing and monitoring to aid surveillance.

Dr. Longhurst presented three case studies:

1. Chest X-rays for COVID-19 Detection: An AI tool at UC San Diego Health initially influenced clinical decisions in 20% of cases but lost relevance as testing improved, leading to its discontinuation in 2020.
2. AI in Sepsis Prediction: A deep learning model reduced sepsis mortality at UC San Diego, benefiting underserved populations.
3. GenAI in EHR: GenAI at UC San Diego streamlined patient messaging while disclosing AI involvement, inspiring California’s Assembly Bill 3030, which mandates AI disclosure in healthcare communications.

Dr. Longhurst also emphasized outcome-driven evaluations and called for a network of AI implementation science centers to assess real-world impact.

Stakeholder Perspective – Technological Innovation and Considerations for Postmarket Performance Monitoring within Radiology

Dr. Nina Kottler, Associate Chief Medical Officer for Clinical Artificial Intelligence at Radiology Partners, shared insights on postmarket performance monitoring in radiology. Radiology Partners has used AI extensively since 2016 to interpret images, including in practices in urban and rural settings.



Dr. Kottler highlighted the use of text-based AI outputs and some computer vision applications in radiology. She emphasized the importance of continuous validation and monitoring, noting that universal best practices are still evolving.

She discussed a study evaluating a GenAI system used in a significant percentage of radiology practices. The study compared AI-generated reports to final radiologist-edited reports to assess clinical accuracy and relevance. A panel of radiologist experts reviewed 3,000 reports to identify clinically significant errors.

The study revealed a 4.8% clinically significant error rate (one in 21 reports) in AI-generated impressions that could impact patient care. After radiologist review and edits, the error rate dropped to 1%. Dr. Kottler stressed the role of radiologist oversight in ensuring AI-generated reports meet clinical standards and minimize patient risk.

Dr. Kottler underscored the impact of integrating an expert human-in-the-loop approach into GenAI workflows in radiology, which she considers key to reducing error rates and improving patient care. This collaborative approach she described has achieved a 90% improvement, decreasing error rates from 4.8% to 0.5%. Dr. Kottler highlighted that addressing errors proactively, before GenAI deployment or during monitoring, is key for ensuring reliability and safety.

Dr. Kottler categorized common GenAI errors into: numerosity, laterality / body site, recommendation, missed finding, and other hallucination. She explained how errors in numbers, laterality, or irrelevant recommendations can lead to clinically significant consequences if not identified and corrected, which demonstrates the importance of precise language and education to empower radiologists to anticipate and address GenAI's limitations.

She advocated for robust pre-validation processes and ongoing monitoring to continuously evaluate GenAI performance. She suggested that methods such as ensembling—using multiple models for quality checks—and assessing whether data inputs align with training data are important for improving outcomes. She also underscored the value of stress testing models and collecting user feedback, while cautioning against over-reliance on any single method.

She concluded that it may not be appropriate for GenAI to operate autonomously in radiology. Instead, she proposed it may be more appropriate to use the technology in conjunction with experts who are equipped with the right tools and knowledge to enhance its effectiveness. This partnership as she described ensures that the technology supports high-quality, patient-centered care while maintaining rigorous safety standards.

Stakeholder Perspective – In Real Life: The Patient Health Information Journey and Generative AI-Enabled Tech Impact

Dr. Grace Cordovano emphasized the importance of a patient-centric approach to postmarket performance monitoring of GenAI-enabled medical devices. Drawing on her



experiences as a Board-Certified Patient Advocate, care partner, and patient, she highlighted the need for transparency, trust, and active patient involvement in the oversight of these technologies. She identified four key areas for advancing patient engagement and ensuring robust monitoring systems.

First, she identified partnership and co-creation with patients. Dr. Cordovano called for intentional collaboration with patients and care partners throughout the product lifecycle. She emphasized that it is important for oversight of GenAI to be democratic, with patients included as stakeholders alongside industry. She stated that this co-creation approach would help ensure diverse voices shape the design, deployment, and monitoring of GenAI-enabled devices. She clarified that patients have demanded transparency about where and how AI is used, advocating for continuous involvement to build trust and accountability.

Second, she identified access to AI outputs. According to Dr. Cordovano, patients have a right to access GenAI-generated outputs when they inform their care decisions, as part of their medical records. She stated that transparency in these outputs enables informed decision-making and shared decision-making. Sharing a personal story, Dr. Cordovano highlighted the challenges patients face when seeking detailed GenAI-related information, from false positives to discrepancies in follow-up recommendations. She explained that standardized access protocols and clearer communication from healthcare providers are essential for building trust.

Third, she identified addressing patient concerns. Dr. Cordovano underscored the pressing need to answer patient questions about GenAI-enabled devices, including their efficacy, biases, and safeguards. She proposed creating patient task forces or workgroups to provide structured feedback and address these concerns systematically.

Fourth, she identified patients as humans-in-the-loop. Dr. Cordovano argued that patients can be recognized as active participants in the monitoring and feedback loop for GenAI-enabled tools. She stressed the importance of a structured mechanism for patients to report issues or errors related to GenAI outputs, similar to escalation paths for other healthcare challenges.

Dr. Cordovano concluded by advocating for a dedicated infrastructure to address patient feedback and concerns, whether through local representatives, a national office, or a hybrid approach. She suggested that this approach would ensure patients and families have clear points of contact for resolving issues related to AI-enabled medical devices.

Stakeholder Perspective – Industry View on Effective Evaluation Methods and Postmarket Performance for Health Generative AI Products

Dr. Dale Webster, Health AI Research Director at Google, discussed the industry's perspective on effective evaluation methods and postmarket performance monitoring for GenAI-enabled health products. His presentation highlighted the evolution of GenAI, its unique challenges compared to predictive AI, and the importance of robust evaluation frameworks to ensure safe and effective deployment.



Dr. Webster contrasted GenAI with earlier predictive AI systems, emphasizing its reliance on large language models (LLMs) that generate text-based outputs rather than classifying or analyzing fixed inputs like images. He explained how GenAI models are fundamentally different from predictive AI, as they generate sequences of text, making their outputs effectively infinite in scope. He compared this rapid evolution to the emergence of convolutional neural networks in 2016, noting the current focus on improving cost, speed, and generalizability for real-world applications.

While Dr. Webster said that the lifecycle steps are similar for predictive and generative AI, he identified key differences in real-world performance evaluation and verification, and validation. He described how predictive AI relies on discrete, well-established metrics. In contrast, he noted how GenAI outputs, such as textual answers to questions, lack standardized evaluation metrics, complicating the task of assessing correctness and relevance.

Dr. Webster described postmarket monitoring as important for tracking GenAI performance in real-world settings. He said that in predictive AI, automated systems can efficiently evaluate discrete outputs like image classifications. However, he clarified that generative AI's complex and nuanced outputs require human judgment, making large-scale monitoring costly, slow, and challenging to automate. He emphasized the opportunity for effective dashboards and sampling methods to detect performance shifts and guide updates.

Dr. Webster also outlined the complexities of evaluating GenAI, first characterizing the how GenAI's outputs can vary widely, requiring multidimensional evaluation frameworks. Then he described subjectivity, explaining that both clinical correctness and user intent are important considerations during evaluation, which introduces variability in assessments. He also identified automation gaps, clarifying that current systems lack the ability to automate evaluation effectively for large-scale deployment.

Lastly, Dr. Webster predicted rapid progress toward a small set of widely accepted metrics and evaluation methods for GenAI, particularly in high-impact use cases. However, he noted comprehensive evaluation frameworks for all applications, including multimodal GenAI that integrates text, images, and video, will take much longer to develop. He stressed the importance of sustained investment and research to ensure responsible and scalable deployment.

Dr. Webster concluded by emphasizing the industry's responsibility to refine evaluation methodologies and monitoring systems to support the safe, effective, and equitable use of GenAI in healthcare. He indicated that these efforts will be important to unlock the potential of these transformative technologies.

Open Committee Discussion Q&A *(Clarifying Questions)*

During the clarifying questions session, panel members addressed key aspects of GenAI transparency, error rates, and outcomes in healthcare, with a focus on how GenAI tools integrate into clinical workflows.



Dr. Elkin raised concerns about the transparency of GenAI use in patient care, particularly when GenAI may outperform traditional clinicians. Dr. Cordovano highlighted the lack of transparency in knowing which healthcare facilities use specific AI tools and emphasized the potential benefit of a publicly accessible resource that maps approved technologies and their usage. She suggested that patients could access this information to make informed decisions.

Dr. Shah asked Dr. Kottler about the variability in error rates (type one and type two errors) for AI-assisted and clinician-only diagnoses, seeking a baseline comparison. Dr. Kottler explained that error rates range widely in the literature (5%-30%) and noted her findings showed a 1.7% error rate attributable to clinicians mirrored by AI. She emphasized the bi-directional improvement when pairing educated physicians with AI, citing a study where AI improved radiologist sensitivity by 24% and radiologists improved AI accuracy by 34%. Regarding outcomes, Dr. Kottler highlighted the challenges of large-scale measurement without access to comprehensive EHR data but identified key metrics, including clinical efficiency, mental burnout reduction, quality improvements, and return on investment. She noted that while generative AI could provide broader benefits in the future, current AI tools mostly enhance efficiency and decision-making within radiology practices.

Dr. Maddox addressed challenges in under-resourced medical centers, asking Dr. Kottler about acceptable error rates in primary care or emergency departments where AI may be used more autonomously. Dr. Kottler stressed that while some errors are acceptable, those that affect patient outcomes significantly, such as incorrect stenosis measurements, are not. She discussed how AI could help alleviate staffing shortages and improve efficiency in radiology, especially in less-resourced settings.

Dr. Soni inquired about the shared responsibility between electronic medical record companies, medical device manufacturers and health care practitioners in enabling generative AI. Dr. Sonja Fulmer clarified that while the device manufacturers are the ones within FDA's purview, the FDA is interested in learning about the shared responsibility to help identify what types of controls may be applied. She encouraged collaboration and provided examples including a predetermined change control plan and other mechanisms that may be part of a special control for collecting postmarket monitoring data.

Dr. Rita Kukafka asked Dr. Cordovano about patient training in using GenAI tools. Dr. Cordovano emphasized the importance of formal training programs that teach patients to use these tools responsibly and contribute to the AI ecosystem by identifying errors. She also highlighted the role of caregivers in supporting patients, noting that family members could actively engage in improving the safety of medical devices.

Dr. Elkin provided an example of highly sensitive D-dimer tests used in diagnosing VTE, suggesting that AI could similarly help rule out serious diagnoses, leading to cost savings and improved efficiency in clinical decision-making. Dr. Melissa Clarkson added that patients want to understand why decisions are made, and if AI is overridden, the rationale can be documented in the medical record for transparency.



Dr. Clarkson also shared a personal story about her father’s inadequate care in a rural hospital, illustrating the opportunity for AI tools to assist in triaging patients and determining when to transfer them to specialized facilities. She argued that AI, though not perfect, could play a crucial role in improving patient care in rural settings by facilitating timely decisions and transfers.

The session highlighted the importance of AI transparency, clear documentation, and patient engagement in healthcare, especially in underserved areas where access to expert care may be limited. Panelists emphasized the role of AI in improving triage decisions and ensuring better patient outcomes in resource-constrained environments.

Committee Discussion of the FDA’s Questions (*Deliberation and response to FDA*)

Members of the Committee were asked to respond to three FDA questions related to postmarket performance monitoring of GenAI-enabled devices in which they mentioned some key aspects of postmarket monitoring and evaluation that may be considered critical to maintaining the safety and effectiveness of these devices.

Question 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.*

Question 3a: *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?*

As posted in the meeting 24 hour Summary, available at: <https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-20-21-2024-digital-health-advisory-committee-meeting-announcement-11202024#event-materials>

The Committee emphasized the importance of postmarket performance monitoring of Gen AI-enabled devices, including automated and scalable approaches to capture how the product is being used, data drift, detection of hallucinations and resulting adverse events. The Committee discussed methodologies including the opportunity for an interim deployment phase before large-scale deployment. The Committee also felt there is a need for specific monitoring capabilities to monitor the postmarket effectiveness of the human-AI interactions. The Committee recognized current FDA postmarket surveillance models and change management approaches as a baseline and encouraged building on existing resources such as Predetermined Change Control Plans (PCCPs) as a useful framework for monitoring devices locally and across multiple sites. The Committee encouraged real world evidence trials to support FDA’s goals to enhance monitoring and evaluation. The Committee also proposed the use of synthetic data as a potential tool for performance evaluation, especially in scenarios with limited data availability. The Committee emphasized the importance of establishing and leveraging standards specific to GenAI, as well as centralized information



sharing infrastructure to facilitate reporting back to the manufacturer and broadly to the ecosystem. This may include providing tools to support transparency such as automated user feedback and watermarking to clearly identify the use of GenAI. The Committee discussed including a broad spectrum of reporting considerations, such as error reporting, and implementation failure, as well as clinical outcomes. The Committee called for the need to educate users on postmarket concepts including data drift and to provide clear definition of failure modes and mitigation strategies..

Question 3b: *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?*

As posted in the meeting 24 hour Summary, available at: <https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-20-21-2024-digital-health-advisory-committee-meeting-announcement-11202024#event-materials>

The Committee recommended comparing distributions of local data against the training dataset to identify drifts over time. The Committee emphasized the need for automated auditing processes capable of measuring data drift, assessing errors, and identifying corrective actions. The committee was supportive of using approaches such as ensemble methods combining outputs from multiple models to improve robustness, and quality assurance checks embedded within the device. They discussed long-term monitoring of patient outcomes and tracking changes in clinical practices over time. The Committee discussed monitoring for health care professional's correction of errors in device output.

Question 3c: *What methods and metrics can be utilized to effectively monitor and evaluate the postmarket 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?*

As posted in the meeting 24 hour Summary, available at: <https://www.fda.gov/advisory-committees/advisory-committee-calendar/november-20-21-2024-digital-health-advisory-committee-meeting-announcement-11202024#event-materials>

The Committee acknowledged the challenge of evaluating GenAI devices that utilize foundation models for which limited information may be available. The Committee encouraged providing information to understand what is in the foundation model, and where that is not possible, noted that it is important to find ways to mitigate the resulting uncertainty. The Committee proposed the need for new tools and frameworks to evaluate and characterize foundation models and their impact or use on a GenAI-enabled device. They emphasized the importance of assessing the representativeness of the training data of the foundation model. The Committee proposed that manufacturers should clearly define performance metrics for each subgroup of the intended use population of the device and monitor outcomes across demographics. The Committee stressed the need for metrics tailored to specific use cases. Beyond model development, the Committee emphasized the need for mechanisms to bring performance data and insights back to community hospitals and medical centers, such as through registries or nonprofit frameworks.



Closing Remarks and Adjournment

Gratitude was expressed to the Committee for their participation, with emphasis placed on the significance of their collective efforts. The meeting was officially adjourned with appreciation for everyone's contributions.

Contact Information:

Artair S. Mallett
Management Analyst
Center for Devices Radiological and Health
Office of Management
U.S. Food and Drug Administration
Tel: 301-796-9638
Mobile: (301) 538-4714
Artair.Mallett@fda.hhs.gov



I approve the minutes of the meeting as recorded in this summary.

__ *Ami B Bhatt, MD* signed electronically 2-5-2025 __

Ami Bhatt, M.D.

Chairperson

I certify that I attended this meeting on November 21, 2024
and that these minutes accurately
reflect what transpired.

James P. Swink
Designated Federal Officer