Healthcare AI Regulation

Guidelines for Maintaining Public Safety and Innovation

Kev Coleman

PARAGON
HEALTH INSTITUTE

DECEMBER 2024

The author is grateful to Naomi Lopez, Joe Albanese, and the Paragon team for their exceptional comments and work in review of the paper.

ABOUT THE AUTHOR

Kev Coleman is a Visiting Research Fellow whose policy foci include artificial intelligence, association health plans, and health insurance. Kev is a recognized healthcare leader, having been named one of “The 20 most Creative People in Insurance.” He was also the conceptual architect of the internet’s first private Medicare insurance marketplace in 2006, years before the launch of the government’s Healthcare.gov marketplace. His healthcare research has been cited in top newspapers and media across the country and referenced in congressional health reform discussions.

A veteran executive of multiple technology companies, Mr. Coleman consults within the healthcare market on issues ranging from start-up product evaluation to their business models. His healthcare research has spanned artificial intelligence applications in health care, Medicare plan designs and formularies, the Affordable Care Act, employer-based health plans, dental insurance, and telemedicine.

EXECUTIVE SUMMARY

What Are the Problems Facing Artificial Intelligence (AI) Regulation in Healthcare?

An awareness of AI among policymakers has, at times, substituted for a meaningful understanding of its operations. When coupled with the dystopian AI predictions occasionally in the press, this situation risks misregulation that can not only increase technology costs but reduce the very medical advances policymakers desire from AI. Additional conditions that can produce the same unfortunate results include failures to:

- recognize how the risk for patient harm differs by AI programming subtype as well as its context of use,
- understand the continuities and discontinuities between AI-enabled systems and traditional software applications,
- avoid new rules whose compliance is already required by existing laws and/or regulations,
- accommodate autonomous AI healthcare scenarios where data dictates clinician assistance is not required,
- grasp that a deficit in AI system explainability is not necessarily a deficit in safety,
- consider temporary regulatory sandboxes for innovative AI that does not fit well within the existing regulatory apparatus,
- maintain consistent demographic representation expectations between AI-enabled medical devices and non-AI medical devices, and
- preserve the incentives for AI developers to remediate deficiencies in its AI systems.

How Can These Problems Be Mitigated?

This paper provides guidelines that protect AI's many benefits without increasing risks to public health. Patient safety is an overarching concern throughout, as injury can be caused not only by faulty AI products but also by superior AI products not securing market access in a timely fashion. Closely related to safety concerns is the evaluation of AI risk and its contextualization within individual AI technologies and their specific medical applications.

The paper additionally address standards for accuracy and patient outcomes delivered through AI-enabled care, including AI systems that can function without clinician assistance. This consideration of accuracy and patient outcomes continues in the scrutiny of AI-enabled systems that have deficits in explainability but produce empirically validated benefits.

Previous regulation by the Food and Drug Administration (FDA) of healthcare software has a significant influence on the guidance provided. The agency's basic model for a software's rigor of inspection and evaluation remains unrevised. The guidance assumes the adequacy of the agency's software classifications — such as software as a medical device, software in a medical device, off-the-shelf software, and general health and wellness software products — and their respective approval pathways. The major departures in approach are not from agencies like the FDA but from an emerging sentiment that believes government should implement expansive preemptive rules on AI healthcare, rules that generalize the diverse technologies categorized under the singular term "AI" and ignore the role of existing patient safeguards for medical device safety and health information.

What Are the Benefits Resulting from this Approach?

The guidelines advocated in this paper present an effective and non-disruptive model for crafting AI healthcare regulation. Above all, the guidelines seek to maintain regulatory governance in existing agencies with historical experience in healthcare matters, albeit with recommendations reflecting the new realities specific to AI technologies. Toward that end, the guidelines warn agencies that additional regulation cannot substitute for the personnel, internal AI expertise, and resources necessary to perform their duties properly. This recommendation has the additional virtue of not expanding the federal bureaucracy, which would invite duplicative regulation and chance greater compliance costs for AI developers at a time when the nation's healthcare costs are already exorbitant.[1]

Another merit of the guidelines is that it avoids regulatory capture — that is, a regulatory framework that favors established companies in the AI market. Instead, these guidelines are sensitive to the challenges faced by startup companies and foster an environment where new market entrants can thrive while still complying with appropriate safety expectations. While this latter issue is critical for the United States if it is to continue its healthcare AI leadership, it is even more important for the continued development of AI innovations that will improve the health of Americans.

[1] See Kev Coleman, "Lowering Health Care Costs Through AI: The Possibilities and Barriers," Paragon Health Institute, July 2024, https://paragoninstitute.org/private-health/lowering-health-care-costs-through-ai-the-possibilities-and-barriers/.

SECTION I

Overview

Section I of this paper reviews several complexities regarding artificial intelligence (AI) regulation in healthcare. Against this background, Section II proposes guidelines to balance public protections and innovation. The guidelines rest on principles safeguarding the many benefits of AI while preserving the safety of patients and their personal health information.

Background

The past few years have brought a flurry of governance concerns regarding AI. This is a new development and one not consistent with AI's prior history. The discipline itself traces its beginnings back to the 1950s and most of the subsequent decades been free of significant legislation and regulation.[2] This situation continued into the 1970s as AI had expanded its influence beyond computer science to early applications within the field of medicine. An important example of healthcare AI from this period was INTERNIST-I, a diagnostic aid employing decision tree programming primarily developed by AI pioneer Harry Pople. INTERNIST-I used a "problem-formulation heuristic"[3] that assisted a clinician in making multiple complex diagnoses for a patient.[4] The National Institutes of Health sponsored the first AI in Medicine workshop at Rutgers University in 1975, and that same university went on to develop a causal-associational network model combining statistical pattern recognition with AI for glaucoma consultations in 1978.[5]

In the absence of major regulatory efforts, AI has made enormous strides, including the release of the fourth generation of ChatGPT,[6] a general-purpose AI chatbot system that can (among many other functions) interpret natural language requests and respond with novel outputs in natural language and other formats. This single AI system is reported to be "the fastest-growing consumer application in history."[7] It has become so popular that its parent

[2] Considered by many to be the first instance of AI software, the Logic Theorist was a problem-solving program presented in publicly in 1956. Rockwell Anyoha, "The History of Artificial Intelligence," Harvard University, August 28, 2017, https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

[3] *Problem formulation heuristic* refers to the INTERNIST-I system's techniques for structuring and resolving challenging diagnostic tasks. Peter Szolovits, "Artificial Intelligence and Medicine," in *Artificial Intelligence in Medicine*, ed. Peter Szolovits (Boulder, CO: Westview Press, 1982), https://people.csail.mit.edu/psz/ftp/AIM82/ch0.html.

[4] R. A. Miller, "A History of the INTERNIST-1 and Quick Medical Reference (QMR) Computer-Assisted Diagnosis Projects, with Lessons Learned," *IMIA Yearbook of Medical Informatics*, 2010, https://www.thieme-connect.com/products/ejournals/pdf/10.1055/s-0038-1638702.pdf.

[5] Cedars-Sinai, "AI's Ascendance in Medicine: A Timeline," April 20, 2023, https://www.cedars-sinai.org/discoveries/ai-ascendance-in-medicine.html.

[6] Generative Pre-trained Transformer 4 (GPT-4) was launched on March 14, 2023.

[7] Krystal Hu, "ChatGPT sets record for fastest-growing user base-analyst note," *Reuters*, February 2, 2023, https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

company, OpenAI, announced in 2024 that ChatGPT had over 200 million users per week.[8] This system is now being used for tasks ranging from research and writing original content to assistance with software programming, brainstorming new ideas, language translation, and summarizing video content.

Recent AI advances in healthcare have been equally impressive. AI programs have passed the U.S. Medical Licensing Examination (USMLE)[9] and AI's application within healthcare is spreading from areas such as diagnosis of medical images and new drug development to administrative automation and remote patient monitoring.[10] These new AI systems can synthesize data volumes well beyond the capacity of human retention, and the scale of the systems' statistical and probabilistic analyses are likewise beyond what can be expected for a physician. The resulting medical benefits are many and include instances where cancers have been detected years before the earliest indicators can be identified by a trained radiologist.[11] Additionally, the speed at which the systems operate can compress the time between patient data collection and diagnosis (where diagnosis had formerly been dependent on the manual interpretation of data).

AI in healthcare is blurring the sharp line separating healthcare providers and assistive technologies. AI is being employed in the training of healthcare professionals, thus affecting the care delivered by human clinicians. The New Jersey Institute of Technology is developing an AI software system that, via a video feed, analyzes a medical trainee's surgical exercise and provides real-time feedback.[12] The system is expected to be part of Robert Wood Johnson Medical School curriculum in 2025.[13]

Alongside the achievements of AI have come concerns about the speed of its adoption and its capacity, like other software-based systems, for error. These concerns have elicited responses across the spectrum of policymakers. The Biden administration issued a government-wide executive order[14] promoting federal regulation of AI in late 2023, the full results of which have

[8] Ina Fried, "OpenAI Says ChatGPT Usage Has Doubled Since Last Year," *Axios*, August 29, 2024, https://www.axios.com/2024/08/29/openai-chatgpt-200-million-weekly-active-users.

[9] Michael DePeau-Wilson, "AI Passes U.S. Medical Licensing Exam," *MedPage Today*, January 19, 2023, https://www.medpagetoday.com/special-reports/exclusives/102705.

[10] For a non-technical discussion of recent AI healthcare applications, see Ariel Katz, "10 Ways AI Is Advancing Healthcare," *Forbes*, September 19, 2023, https://www.forbes.com/councils/forbestechcouncil/2023/09/19/10-ways-ai-is-advancing-healthcare/.

[11] Zoe Kleinman, "NHS AI Test Spots Tiny Cancers Missed by Doctors," *BBC*, March 20, 2024, https://www.bbc.com/news/technology-68607059; Berkeley Lovelace Jr. et al., "Promising New AI Can Detect Early Signs of Lung Cancer That Doctors Can't See," *NBC News*, April 11, 2023, https://www.nbcnews.com/health/health-news/promising-new-ai-can-detect-early-signs-lung-cancer-doctors-cant-see-rcna75982.

[12] Michael Giorgio, "AI-Powered Surgical Training Program Provides Real-Time Feedback and Instruction," *NJIT*, February 26, 2024, https://news.njit.edu/ai-powered-surgical-training-program-provides-real-time-feedback-and-instruction.

[13] Giorgio, "AI-Powered Surgical Training Program."

[14] The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," October 30, 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

not been fully realized. At the state level, almost 700 legislative proposals related to AI were considered in 2024, compared to just 191 in 2023.[15] At least 45 states have proposed these AI-related bills in 2024, and 31 enacted such legislation.[16] though the types of AI bills vary considerably.[17] Concerns about patient safety and privacy are among the major themes of these bills. AI safety concerns appear to be largely preemptive as opposed to driven by major events of patient injury, but data privacy is a more complicated matter. AI systems' consumption of large quantities of data is well known. If an AI has access to sensitive information on individuals, this information might be accidentally disclosed. This is particularly true for domains that lack the statutory privacy protections that healthcare has.

The accelerating progress of AI and its attending uncertainties account for the urgency of present discussions of AI regulation. However, an awareness of AI among policymakers has, at times, substituted for a meaningful understanding of the software's operations. When coupled with the dystopian AI predictions, this situation creates an environment conducive for the kind of misregulation that not only increases costs but inhibits the very medical advances policymakers desire from AI. This latter issue is particularly pressing since it directly impairs public health. Perhaps the best current example of this problem is in the State of Colorado.

Colorado was the first state in the nation to pass a comprehensive AI regulation bill, the Colorado Artificial Intelligence Act,[18] which (among other obligations) requires developers of "high-risk" AI systems to prevent algorithmic discrimination when such an AI system's contribution is a "substantial factor" in a "consequential decision." The bill's concepts of "high risk" and "consequential decision" are ambiguous and potentially problematic, while its central preoccupation — unlawful discrimination — are already defined by existing federal and Colorado state anti-discrimination laws.[19] Unsurprisingly, the governor's signing statement expressed concerns about the law,[20] and he subsequently vowed to revise the legislation.[21]

[15] BSA, "2025 State AI Wave Building After 700 Bills in 2024," October 22, 2024, https://www.bsa.org/news-events/news/2025-state-ai-wave-building-after-700-bills-in-2024.

[16] National Conference of State Legislatures, "Artificial Intelligence 2024 Legislation," September 9, 2024, https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation.

[17] Ibid. The many foci of 2024 AI legislation proposals include deepfakes, political advertisements, discrimination in healthcare algorithms, and initiating AI studies or committees to study AI.

[18] Colorado SB24-205, "Consumer Protections for Artificial Intelligence," https://leg.colorado.gov/sites/default/files/documents/2024A/bills/2024a_205_enr.pdf.

[19] Ibid. The Colorado Artificial Intelligence Act defines *algorithmic discrimination* as "any condition in which the use of an artificial intelligence system results in an unlawful differential treatment or impact that disfavors an individual or group of individuals on the basis of their actual or perceived age, color, disability, ethnicity, genetic information, limited proficiency in the English language, national origin, race, religion, reproductive health, sex, veteran status, or other classification protected under Colorado state law or federal law."

[20] Gov. Jared Polis, letter to the Colorado General Assembly, May 17, 2024, https://drive.google.com/file/d/1i2cA3IG93VViNbzXu9LPgbTrZGqhyRgM/view.

[21] Marissa Ventrelli, "Colorado Governor, Attorney General Vow to Revise Recently-Enacted AI Law," *Colorado Politics*, June 17, 2024, https://www.coloradopolitics.com/news/colorado-governor-attorney-general-vow-to-revise-recently-enacted-ai-law/article_53aa96b8-2a7e-11ef-9d99-d791f3cce317.html.

Ultimately, the avoidance of misregulation is neither impossible nor simple. AI, like regulation itself, is complicated and has a spiderweb of dependencies that are not always obvious. Developing regulatory guidelines that preserve safety and privacy while fostering continued innovation requires a grounding in:

- the types of AI and their healthcare contexts,
- the FDA's treatment of medical software,
- AI data dependencies, and
- ongoing AI software improvements.

Types of AI

The path to optimized regulatory efforts begins with AI's proper definition. The term "*artificial intelligence*" does not reference a single technology but rather a category of software programming forms whose operations resemble some dimensions of human learning and reasoning. The learning and reasoning exhibited are often dependent on large datasets to train the software. The programming forms underlying the systems constitute the major subtypes of AI (though further categorizations exist under each subtype). Most notable for healthcare are the following: machine learning, artificial neural networks, generative AI, and large language models.

Machine learning software produces probabilistic predictions (upon which decisions may be made) through a variety of statistical and algorithmic techniques. Its healthcare applications are numerous, ranging from cardiovascular risk prediction[22] and new drug discovery[23] to treatment effectiveness modeling[24] and alerting physicians to potential prescribing errors.[25] However, machine learning is hardly the first implementation of statistical computing. Its history stretches back to the statistical laboratories of the late nineteenth century[26] and industry-leading statistics software SAS (originally Statistical Analysis System) was first

[22] Haya Salah and Sharan Srinivas, "Explainable Machine Learning Framework for Predicting Long-Term Cardiovascular Disease Risk Among Adolescents," *Scientific Reports*, December 19, 2022, https://www.nature.com/articles/s41598-022-25933-5; and Manasvi Singh et al, "Artificial intelligence for Cardiovascular Disease Risk Assessment in Personalised Framework: A Scoping Review," *eClinicalMedicine*, July 2024, https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(24)00239-6/fulltext.

[23] "With advances in the application of ML across diverse therapeutic areas, we posit that fully ML-integrated drug-discovery pipelines will define the future of drug-development programs." Denise B. Catacutan et al, "Machine Learning in Preclinical Drug Discovery," *Nature*, July 19, 2024, https://pubmed.ncbi.nlm.nih.gov/39030362/.

[24] Snigdha Dubey et al., "Using Machine Learning for Healthcare Treatment Planning," *Frontiers*, April 24, 2023, https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1124182/full; and Rui Li et al., "G-Net: A Recurrent Network Approach to G-Computation for Counterfactual Prediction Under a Dynamic Treatment Regime," *Proceedings of Machine Learning Research* 158 (2021): 282-297, https://proceedings.mlr.press/v158/li21a/li21a.pdf.

[25] Ronen Rozenblum et al., "Using a Machine Learning System to Identify and Prevent Medication Prescribing Errors: A Clinical and Cost Analysis Evaluation," *Joint Commission Journal on Quality and Patient Safety* 46, no. 1 (January 2020): p3-10, https://www.jointcommissionjournal.com/article/S1553-7250(19)30396-4/abstract#sec0010.

[26] David Alan Grier, "The Origins of Statistical Computing," *American Statistical Association*, https://ww2.amstat.org/asa175/statcomputing.cfm

developed in the late 1960s.[27] In many ways machine learning continues decades of advances in statistical computing, albeit enhanced by contributions from other disciplines such as computer science and mathematics. One of the principal ways machine learning is differentiated from other statistical computing is in its emphasis on predictive accuracy. Traditional statistics, in contrast, often emphasizes how closely a model depicts what has been previously observed about its subject. What does not separate statistics and machine learning are the underlying types of mathematic equations used to process data.

The same can be said for an artificial neural network. An artificial neural network is a specific form of machine learning. A basic feature of an artificial neural network is a set of interconnected artificial "neurons" known as a layer. An individual layer inside[28] an artificial neural network can be devoted to a variety of different data processings. Each individual neuron performs its own calculation to determine if the neuron will activate and pass its data to other neurons. A bias may be used to affect a neuron's propensity for activation and a weight may be used to modify the influence of an activated neuron[29] for the next neural layer. As was the case for the statistical techniques used in general machine learning, an artificial neural network's calculations (e.g., sigmoid, hyperbolic tangent, binary step, etc.) originate from mathematics outside of AI and are not specific to AI. This origin is extremely important because many of the fears about AI functionality ignore that AI's functions are merely an expression of advanced math and statistics, two areas that do not attract the same calls for heavy regulation.

Artificial neural networks come in different architectures representing additional AI subtypes. A convolutional design, for example, can interpret spatial data such as found in static images. A possible medical application is the detection of melanoma within a photographic image. A recurrent design, in contrast, is better for data that exists in a progression such as video or text. Natural language processing is a classic use for the recurrent design, as the words that are earlier in a textual series (i.e. a sentence) can affect what meanings are most probable for the words used later in the series.

When used in a competitive architecture to generate original content — such as text or images — an artificial neural network configuration is an instance of generative AI.[30] For

[27] https://www.sas.com/en_us/company-information/history.html

[28] "Inside" in this context means existing at a specific depth between the input layer and output layer of an artificial neural network.

[29] More specifically, the value an activated neuron transfers to the next neural layer.

[30] Other forms of generative AI exist such as variational autoencoders and transformers. See Texas A&M University, "What Is Generative AI?," https://geosat.tamu.edu/genai-overview/; and Rebecca U. Shin, "Complete Guide to Five Generative AI Models," Coveo, May 14, 2024, https://www.coveo.com/blog/generative-models/.

example, in a generative adversarial network,[31] one artificial neural network produces a new artifact[32] that is an imitation of a class of objects. The imitation is produced through identifying the patterns and attributes of the class and then reproducing these characteristics within the new artifact. A second network, working in competition, is used to determine if the new artifact is real or an imitation. The second network, known as a discriminator, compares the new artifact to examples in the same class confirmed to be authentic. If the discriminator can distinguish the new content from the confirmed content, the generator must improve the content through subsequent iterations until the discriminator cannot reliably differentiate between the new artifact and the confirmation data.

As is the case for other forms of AI, generative AI has a vast range of applications. For example, it can design new proteins and molecules that can then be tested for clinical efficacy.[33] Perhaps more notably, generative AI has attracted considerable public attention through its large language model (LLM) implementations. Chatbots and other tools using LLMs can interpret natural language requests and produce outputs in the same form. In healthcare, LLMs and related AI technology are being evaluated for producing patient discharge communications, summarizing medical articles, and auto-populating electronic health records.

The discrimination of AI subtypes is one of several critical issues facing policymakers since public safety risks are not uniform in character or incidence across subtypes. Moreover, within public health, harm can arise from an AI-enabled system not only through a software anomaly[34] but also a market access delay for a beneficial AI system, the latter due to AI's staggering capacity for diagnostic, pharmaceutical, and treatment improvements. Some of these improvements can mean the difference between life and death.[35]

The differentiation of individual AI technologies is exceedingly important to policymakers and regulators because:

[31] This is one of several models of generative AI. See Aminu Abdullahi, "Generative AI Models: A Complete Guide." *eWeek*, January 5, 2024, https://www.eweek.com/artificial-intelligence/generative-ai-model/

[32] An artifact in generative AI can be any kind of content from text to an image to a sound or a video.

[33] Xiangru Tang et al., "A Survey of Generative AI for De Novo Drug Design: New Frontiers in Molecule and Protein Generation," *Briefings in Bioinformatics* 25, no. 4 (July 2024), https://academic.oup.com/bib/article/25/4/bbae338/7713723.

[34] The FDA defines *software anomaly* as "any condition that deviates from the expected behavior based on user needs, requirements, specifications, design documents, or standards. Anomalies may be found during the review, test, analysis, compilation, or use of the software (whether before or after release, or whether inside a sponsor's organization or outside it) or at other times." FDA, "Content of Premarket Submissions for Device Software Functions," June 2023, https://www.fda.gov/media/153781/download.

[35] National Heart, Lung, and Blood Institute, "Can Artificial Intelligence Help Save Lives?," February 8, 2024, https://www.nhlbi.nih.gov/news/2024/can-artificial-intelligence-help-save-lives; Pete Arduini and Scott Whitaker, "Health Care AI Is Already Saving Lives," *Star Tribune*, September 20, 2024, https://www.startribune.com/health-care-ai-is-already-saving-lives/601147940; Mount Sinai School of Medicine, "AI Can Help Doctors Make Better Decisions and Save Lives," *ScienceDaily*, June 13, 2024, https://www.sciencedaily.com/releases/2024/06/240613221923.htm.

- The degree and types of consumer risk differ according to the subtype of AI.
    - Some categories of risk are not unique to an AI technology but are inherent to the statistical or probabilistic models on which the technology is based.
- The use of vague or overly broad definitions can produce several undesirable results including the possibility of:
    - An AI-enabled system being marketed as something other than AI to elude AI regulation,
    - Increased compliance costs on an AI subtype for which a regulation does not meaningfully apply,
    - The suboptimization of a future AI subtype regardless of the new subtype's risk profile, and
    - Regulatory capture that benefits the richest vendors in the AI healthcare while impeding innovative startups from entering the market.

With respect to the first point, the differences in risk are best illuminated by so-called AI hallucinations. This phrase, misleading both in its inferences of consciousness as well as uniformity for the associated errors, pertains to several distinct output anomalies that can manifest in outputs from LLMs and generative AI. IBM describes an AI hallucination as "a phenomenon wherein a large language model (LLM) — often a generative AI chatbot or computer vision tool — perceives patterns or objects that are nonexistent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate."[36] These defective outputs include:

- unintelligible word combinations,
- factual errors,
- citations of information resources that do not exist, and
- answers that are factually correct but not relevant to an end user's inquiry.

A bill targeting AI hallucinations would ideally specify the AI technologies to which a new rule would apply, as AI systems can output natural language communications without LLMs or generative AI.[37] Additionally, such a bill might include further specificity since hallucinations are more common in general purpose generative AI/LLMs, where the training datasets are

[36] IBM, "What Are AI Hallucinations?," https://www.ibm.com/topics/ai-hallucinations. For a more detailed treatment see Yujie Sun et al., "AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content," *Humanities and Social Sciences Communications*, September 27, 2024, https://www.nature.com/articles/s41599-024-03811-x.

[37] For example, an AI system could use a simple IF-THEN-ELSE model from traditional software programming to identify when a condition was satisfied that warranted a predetermined message to the system's end user. Such a message would not be at risk for hallucination.

both mammoth and diverse (i.e. a broad range of subjects as opposed to a single subject matter). A purely statistical machine learning system without generative AI or LLM capabilities, in contrast, could be subject to potential anomalies like any other software system, but it would not manifest hallucinations, because hallucinations are a category of errors endemic to generative AI systems and LLMs. Any regulatory intervention would need to avoid (1) disrupting the primacy of the FDA's evaluation process for ensuring medical software safety given the agency's concentration of expertise with medical devices and history dealing with safety issues, and (2) eroding the incentives AI manufacturers have to remediate known technology anomalies.

The context in which generative AI is used also affects the degree of risk associated with the technology. Back-office medical supply purchasing, for example, carries a much lower risk than a patient-facing diagnostic application. Likewise, a generative AI system that develops chemical recipes for new antibiotics has a lower risk severity, because any drug candidates it creates must undergo the FDA's evaluation protocols to ensure drug safety and effectiveness prior to market access.

The context-dependency of risk illustrates why cross-industry rules and a centralized federal AI office are not ideal for mitigating AI risks.

FDA Treatment of Medical Software

AI-enabled systems that are used for medical care are not self-regulated in the United States. As medical devices, they must be approved by the FDA prior to being marketed domestically. The FDA defines medical devices as items that diagnose disease and health conditions, affect either the structure or function of the body, or address disease through prevention, mitigation, treatment, or cure.[38] The FDA considers AI to be a medical device (even in the absence of dedicated hardware to run the software) so long as it is intended for one or more medical purposes.[39] Alternatively, the FDA does not consider an internet search engine a medical device, because its designed purpose is not for medical care, even though people may, of their own volition, use this tool to attempt self-diagnosis of medical conditions. In fact, the FDA does not regulate wellness products despite their health applications if they are only for general wellness use and present a low safety risk.[40] A wearable wellness device, such as an Apple Watch, can provide its user with important alerts, related to the detection of sleep apnea, irregular heart rhythms, and body temperature but, given the low consumer risk

[38] U.S. Food & Drug Administration, "How to Determine if Your Product is a Medical Device," September 29, 2022, https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device.

[39] Ibid.

[40] FDA, "General Wellness: Policy for Low Risk Devices," September 27, 2019, https://www.fda.gov/media/90652/download.

associated with this feedback, the Apple Watch remains classified as a general wellness product.

With respect to FDA review and approval, AI-enabled medical devices have three main pathways:

1. **De novo** for low-to moderate-risk medical devices with no predicates (i.e. referenceable devices that are of the same type as the applicant's product but already approved by the FDA);
2. **Premarket clearance (510(k))**, for moderate-to high-risk medical devices with market predicates and equivalent safety and effectiveness; and
3. **Premarket approval** for high-risk medical devices, is the most demanding pathway, and requires that the clinical investigations in the submission address numerous factors.[41]

The way the FDA accommodates devices without predicates is key aspect of how the agency supports innovation. Alongside the agency's consideration of patient safety pertaining to the safety of a medical devices use, the FDA uses the presence or absence of predicates as another indication of safety. With respect to a determination of predicates for a medical device, the FDA explains that:

> A claim of substantial equivalence does not mean the new and predicate devices needs to be identical. FDA first establishes that the new and predicate devices have the same intended use and any differences in technological characteristics do not raise different questions of safety and effectiveness. FDA then determines whether the device is as safe and effective as the predicate device by reviewing the scientific methods used to evaluate differences in technological characteristics and performance data. This performance data can include clinical data and non-clinical bench performance data, including engineering performance testing, sterility, electromagnetic compatibility, software validation, biocompatibility evaluation, among other data.[42]

Pre-market approval, for high-risk medical devices with no predicates, is the most demanding FDA review pathway.[43] Clinical investigations required in the submission cover "study

[41] These include study protocols, safety and effectiveness data, adverse reactions and complications, device failures and replacements, patient information, patient complaints, tabulations of data from all individual subjects, results of statistical analyses, and any other information from the clinical investigations. FDA, "General Wellness."

[42] FDA, "Premarket Notification 510(k)," August 22, 2024, https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/premarket-notification-510k.

[43] FDA, "Premarket Approval," May 16, 2019, https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/premarket-approval-pma.

protocols, safety and effectiveness data, adverse reactions and complications, device failures and replacements, patient information, patient complaints, tabulations of data from all individual subjects, results of statistical analyses, and any other information from the clinical investigations."[44]

Should a medical device successfully complete one of the three FDA review pathways, the device can be commercially marketed, and it carries with it the FDA's assurance the device has undergone rigorous testing. Central to the FDA's consumer protection for AI-enabled medical devices is internal expertise in AI.[45] Without an adequate grasp of AI, agency regulators cannot make a reliable equivalency determination. Because of its work evaluating the science and data supporting medical devices, the FDA is the preeminent protector of public safety regarding AI-enabled healthcare systems. Additional regulation cannot substitute for the FDA having the personnel, expertise, and resources to perform its reviews properly.

The FDA approval pathways for AI-enabled medical devices provide an important precedent for consumer protection efforts both in its emphasis on the safety of medical-device outcomes and also in its determination of which software is regulated as a medical device. This latter point, while touched upon earlier in the discussion of the FDA's nonregulation of general health and wellness devices, has further nuance relevant to the scrutiny of AI healthcare software.

The FDA regulates software that performs a medical function independently of the hardware on which it operates as "software as a medical device" (SaMD).[46] If the software is embedded or controls the device and is dependent on that device to perform its intended purpose, it is regulated as "software in a medical device" (SiMD).[47] Whether an AI-enabled device is a SaMD or SiMD, they face similar safety standards from the FDA.[48]

The FDA distinguishes SaMD and SiMD from software that is general purpose, even if the general-purpose software is used in a healthcare setting or operates part of a medical

[44] Ibid.

[45] Coleman, "Lowering Health Care Costs Through AI."

[46] FDA, "What Are Examples of Software as a Medical Device?," December 6, 2017, https://www.fda.gov/medical-devices/software-medical-device-samd/what-are-examples-software-medical-device.

[47] FDA, "How to Determine If Your Product Is a Medical Device." Seel also L&T Technology Services, "SIMD/SAMD: Everything You Need to Know," September 2, 2022, https://www.ltts.com/blog/simd-samd.

[48] Stephanie Van Ness, "SaMD vs. SiMD: Do You Know the Difference?," Integrated Computer Solutions, February 7, 2024, https://www.ics.com/blog/samd-vs-simd-do-you-know-difference.

device.[49] Software functions (supplied by the FDA[50]) that do not constitute a medical device cover functions such as:

- data transfer, storage, format conversion, and display (when not controlling or altering a connected medical device's functions);
- patient collection of health status information such as blood glucose, blood pressure, heart rate, and weight to transmit to a heath care professional or electronic health record;
- general office operations automation;
- insurance claim data collection as well as the analysis of insurance claims; and
- exercise activity monitoring.

The FDA further distinguishes off-the-shelf (OTS) software that may be used within a medical device. For example, a SaMD could run on a computer utilizing an OTS operating system. The supplier of the OTS software is not regulated by the FDA but a medical device manufacturer "still bears the responsibility for the continued safe and effective performance of the medical device"[51] that runs on the OTS operating system. In the manufacturer's risk management file provided to the FDA, the manufacturer must include the risks of OTS software as well as their mitigation.[52] The manufacturer must also provide "documentation regarding test plans and test results as part of the verification and validation activities for the OTS software."[53]

Data: Patient Consent, Privacy, and Security

AI is intimately related to data, particularly large data sets. The scope of information used by AI and its volume have raised concerns about AI's access to data, especially with respect to sensitive information on consumers' health. Likewise, even where consumers have clearly consented to their health information being used in research, there are still apprehensions about an AI system's capacity to keep this information private and safe from cybercriminals. Together, patient consent to the use of personal data along with this data's continued privacy and security should frame regulatory debate about AI data within healthcare.

[49] For examples of mobile software applications in a healthcare setting that do not qualify as medical devices see FDA, "Examples of Software Functions That Are NOT Medical Devices," September 29, 2022, https://www.fda.gov/medical-devices/device-software-functions-including-mobile-medical-applications/examples-software-functions-are-not-medical-devices. The FDA also has additional rules for software used for medical device manufacturing and maintenance.

[50] Ibid.

[51] FDA, "Off-the-Shelf Software Use in Medical Devices," August 11, 2023, https://www.fda.gov/media/71794/download.

[52] Ibid.

[53] Ibid.

The function of data often differentiates AI software from traditional software. AI software relies on large datasets to train the system to produce desired outputs. In traditional software, the application is programmed with predetermined set of commands to process input and produce outputs, including the allowable conditions for the use of alternative instructions. These instructions define the set of possible outputs. AI, in contrast, begins with an architecture that allows the system to learn and improve based on the analysis of data. While the specifics of data training vary according to AI category, generally an AI system is trained on a large set of examples. Once the system can replicate the outcomes corresponding to the training examples, it is tested on a new set of data to confirm its determinations (or predictions) are accurate and repeatable.[54]

AI training data may be unlabeled (i.e. lacking descriptions or associations connecting different categories of data) to maximize the possibility of detecting unrecognized relationships and patterns. In other cases, data is labeled — that is, annotated by specialists with domain expertise to describe or associate categories of data. In some cases, the system will receive feedback as it interacts with data, creating a reinforcement mechanism for desired outcomes.

The integrity of the training data is essential for the success of the AI system. If data is incomplete, biased, or otherwise flawed,[55] predictions are impaired. Data that is unrepresentative of real-world scenarios will likely produce output errors or inadequacies. Moreover, unrepresentative data can also introduce bias into the system if the training data is not adequately representative of the population the system will serve. These data dependencies are not unique to AI and apply equally to any medical research used to develop a medical device. The history of medical device development has already established multiple challenges around medical devices producing identical outcomes across all population cohorts or having perfectly representative subjects within the medical studies supporting a medical device's development. Some of the more salient challenges can be summarized as follows:

- There are medical conditions that affect demographic segments who are unrepresentative of the general population. Examples include:
    - Illnesses that are more common for specific age groups (e.g., pediatric cancers such as neuroblastoma),
    - Sex-specific cancers (e.g., prostate cancer, ovarian cancer), and

[54] There are a variety of testing methodologies, including cross-validation, where an AI-system's available data is organized into multiple arrangements of training and testing data so that the testing data in one validation cycle might be part of the training data in another cycle.

[55] Google, "What Are AI Hallucinations?," https://cloud.google.com/discover/what-are-ai-hallucinations.

  - Diseases that have a disproportionate representation among particular racial and ethnic groups (e.g., sickle cell disease and Tay-Sachs).
- Historical data resources maintained for a disease (e.g., a medical image library for lung cancer patients) are typically maintained for archive purposes without consideration for demographic representation.
- Strict inclusion and exclusion criteria necessary for patient participation in a randomized clinical trial may create a trial population that is not representative of the general population.[56]
- Rare and ultra-rare disease groups that affect a very small unrepresentative segment of the population.

Even the most manual of medical studies must contend with whether their subjects are sufficiently representative for broadly applicable medical conclusions. AI, as a healthcare technology, has demographic representation concerns that are common to healthcare technology as a whole. Consequently, attempts to require unique demographic expectations for AI data will need to demonstrate why such a standard should be isolated to AI-enabled devices and not the research funding all medical device development.

AI data challenges go beyond issues of demographic representation. As is the case for statistical models used by non-AI medical software, the data used by AI systems carry the risk of underspecification.[57] Underspecification occurs when system training and validation produce multiple statistical models of equivalent efficacy leaving the system unable to predict which model will be more effective in the real-world. This can happen even when using very large data sets. This is a serious challenge not only for AI but for any medical device employing statistical modeling. For both AI and non-AI statistical models, the risk of underspecification may require customized stress testing and other interventions to satisfy the FDA's efficiency and safety requirements.

A more common data-related concern about AI (arising more often outside of healthcare) is its ability consume other entities' content and imitate it without attribution or compensation. Healthcare advocates, on the other hand, are more worried about sensitive patient information used by AI-enabled systems and the attending security and privacy considerations. Any computer system involved in the collection, storage, processing, or transmission of sensitive health information raise the question of that information's privacy and security. With respect to these concerns, AI-enabled medical devices used by healthcare providers must comply

[56] This issue is discussed in the context of complementing randomized clinical trials with "real world data" in Lawrence Blonde et al, "Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician," *Advances in Therapy*, October 24, 2018, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6223979/.

[57] Alexander D'Amour et al., "Underspecification Presents Challenges for Credibility in Modern Machine Learning." *Journal of Machine Learning Research* 23 (2022): 1-61, https://jmlr.org/papers/volume23/20-1335/20-1335.pdf.

with the Health Insurance Portability and Accountability Act (HIPAA) and the and the Health Information Technology for Economic and Clinical Health (HITECH) Act if they receive, store/maintain, or transmit protected health information (PHI).[58] HIPAA defines PHI as individually identifiable health information. This definition includes physical status or mental health status (past and present), healthcare services provided, and associated payment details that can be reasonably associated with a person (e.g., through an identifier such as a name or Social Security number).[59]

To comply with HIPAA, a manufacturer of a medical device must maintain a patient's privacy rights and protect the PHI from disclosure without the patient's consent. These HIPAA-defined privacy rights include the need for a patient's written consent before PHI may be used or shared.

The Security Rules of HIPAA are closely related to HIPAA's Privacy Rules. The security rules[60] obligate medical devices to ensure PHI's:

- confidentiality, integrity, and availability;
- protection against reasonable anticipated security threats;
- protection against reasonably anticipated, impermissible uses or disclosures; and
- handling by the manufacturer's workforce is compliant with the security rules.

The possibility of AI enabling self-service medical care delivery (without a clinician) raises the issue of whether HIPAA's definition of *covered entity*[61] needs to be updated to ensure such devices handling PHI remain under the same security and privacy rules as a human healthcare provider.

With AI-enabled medical devices, there is a considerable dependency on the proper enforcement of privacy and security regulations. Vigilance in HIPAA enforcement around patient consent and data ownership will continue to be an essential tool in consumer protection as hospital systems and AI manufacturers collaborate on new AI systems. On this

[58] Daniel Lopez, "What Are the HIPAA Rules for Medical Devices?," *NetSec*, November 15, 2022, https://www.netsec.news/hipaa-rules-medical-devices/.

[59] U.S. Department of Health and Human Services, "Summary of the HIPAA Privacy Rule," October 19, 2022, https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[60] U.S. Department of Health and Human Services, "Summary of the HIPAA Security Rule," October 19, 2022, https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.

[61] Covered entities are the people and organizations to which HIPAA rules apply.

front, policymakers can craft violation fines that strongly encourage AI developers to embrace PHI de-identification (or other methods of anonymization) in design when conditions permit.

There is also the issue of countries of concern gaining access to sensitive PHI through business partnerships or company acquisitions. This concern goes well beyond AI manufacturers. It applies equally to credit bureaus, pharmaceutical companies, cell phone service providers, financial services, and many more industries who capture sensitive information (health or otherwise) on Americans. Given the scope of the issue, its proper address may be more in the realm of a statute that has gone through legislative debate than agency regulation.[62]

Ongoing AI Software Improvements

Effective regulation must preserve industry incentives for improving deficiencies in AI-enabled systems. If, in contrast, a regulation targeting a specific deficiency was issued that added a compliance obligation regardless of whether the issue was remedied, then the industry has little incentive to correct the deficiency. Specifically, a regulatory obligation (e.g., a supplemental clinical evaluation) addressing a known AI deficiency should no longer apply to an AI system that can satisfactorily demonstrate that the issue has been successfully remediated. It should be noted that in the absence of explicit regulation on hallucinations, the AI field has nevertheless evidenced progress on the matter in both commercial and academic contexts. Researchers at the University of Oxford revealed this year a method that estimates a question's degree of uncertainty and its likelihood to produce a LLM hallucination.[63] Retrieval Augmented Generation (RAG) systems are being developed to perform intra-system fact validations on LLM outputs using external data sources such as peer-reviewed research papers,[64] and some such systems are being further enhanced by knowledge graphs that structure relationships among semantic entities (things, ideas, events, etc.) drawn from multiple sources.[65]

In the FDA approval paths for medical AI systems, risk plays a central role in the efforts of AI developers to improve their systems. An AI-enabled system's risk profile for patient injury affects what pathway is used for FDA approval as well as the extensiveness of the data and

[62] It should be noted in connection to this issue that the Biden Administration has issued an executive order directing multiple governmental entities to protect Americans' sensitive personal data from foreign exploitation. The White House, "Fact Sheet: President Biden Issues Executive Order to Protect Americans' Sensitive Personal Data," February 28, 2024, https://www.whitehouse.gov/briefing-room/statements-releases/2024/02/28/fact-sheet-president-biden-issues-sweeping-executive-order-to-protect-americans-sensitive-personal-data/.

[63] University of Oxford, "Major Research into 'Hallucinating' Generative Models Advances Reliability of Artificial Intelligence," June 20, 2024, https://www.ox.ac.uk/news/2024-06-20-major-research-hallucinating-generative-models-advances-reliability-artificial.

[64] Tufts University, "AI Literacy in the Biomedical Sciences," https://researchguides.library.tufts.edu/c.php?g=1398045&p=10352472.

[65] Vinay K. Chaudhri et al., "An Introduction to Knowledge Graphs," Stanford University, May 10, 2021, https://ai.stanford.edu/blog/introduction-to-knowledge-graphs/. Each entity in the knowledge graph (also known as a semantic network) acts as an individual node. The association between one node and another, known as an edge, provides relationship context for information queries.

science review associated with the system. As a consequence, unresolved issues that pose a significant patient safety risk will fail FDA review, but a minor software defect that does not pose such a risk may be permitted. The FDA classifies these latter issues as unresolved software anomalies and defines one as "a defect that still resides in the software because a sponsor deemed it appropriate not to correct or fix the anomaly, according to a risk-based rationale about its impact to the device's safety and effectiveness."[66]

Ongoing software improvements, of course, are not limited to software defects. New system functionality requires new regulatory approval by agencies such as the FDA. There are also improvement scenarios that pertain to neither a defect nor a new function. For example, the degree to which an AI-enabled system can function without the oversight of a clinician may grow over time. Thus, regulators should provide an economical pathway for AI developers to re-apply for approval where functionality remains the same but system autonomy increases. Related to this matter is the prospect that AI systems may eventually provide oversight for human clinicians, such as nurse practitioners and physician assistants. State laws mandating the supervision of one or both roles by a licensed physician may need to be revised for scenarios where an AI system could perform the same supervision at less cost but with the same quality.

Returning to the FDA, their historic work in medical device oversight provides several lessons for future rules on healthcare AI improvements. First and foremost, the agency's approach does not demand perfection from medical devices but does enforce patient safety as its preeminent priority. Risk is considered in terms of both probability of occurrence and severity of harm. Conversely, the FDA also considers a medical device's benefits alongside risk, producing a nuanced strategy for dealing with medical device improvements.

SECTION II

Overview

Having outlined major considerations grounding the discussion of AI regulation in healthcare, this section proposes a series of guidelines that discourage misregulation of AI in healthcare. *Misregulation* refers to regulation that fails to:

- scope its requirements to the technologies associated with the regulation;
- avoid duplication of rules already issued;
- enhance patient protections (e.g., safety, data privacy) while increasing administrative burden for AI developers; and
- foster technology improvement and innovation.

[66] FDA, "Content of Premarket Submissions for Device Software Functions."

Guidelines for the Regulation of Healthcare AI

1. Identifying AI Type
    a. Any regulation of healthcare AI must explicitly identify the AI technology (or technologies) to which the rule applies.
    b. Any healthcare AI specified for regulation must include the technology's operational definition, and this definition should reflect the categorical granularity appropriate to the issue(s) or risks being addressed by the rule.
        i. For example, if a regulation targets a problem distinctive to recurrent artificial neural networks, it should designate this technology explicitly and not use the more general "machine learning" category, for which a recurrent artificial neural network is a subtype.
        ii. This granular approach provides a technical foundation for regulators to determine if a rule should be extended for a de novo AI innovation such as a newly developed AI programming form.
    c. Policymakers must specify criteria to differentiate AI-enabled systems from non-AI systems in law and regulation.
        i. This approach prevents traditional non-AI software from unnecessary rules given the overlap between the functionality of traditional software and the AI-enabled systems.
        ii. This approach also prevents AI software evading appropriate regulation by marketing itself as something other than AI.

2. Identifying AI Context
    a. Any regulation of healthcare AI must identify the use context to which the rule applies. This context does not require detail on clinical indications but does require a description of the AI's environment and function.
        i. The same AI technology in different use contexts can have very different risk severity.
        ii. Medical facilities may have AI in a variety of areas, including activity with no risk to patient safety (e.g., email spam filters, human resources systems, word processors, medical claims processing).
    b. Policymakers should consider whether off-label uses be permitted in scenarios where scientific evidence or expert medical opinion argues for the practice.

   c. Non-commercial use of AI-enabled systems, such as AI as an investigational medical device, is not contemplated within the scope of these guidelines.

3. Identifying AI Risks
   a. The relationship between AI type and use context must inform policymakers' estimation of risk, as these factors affect what problems may manifest, the likelihood of their occurrence, and the possibilities of patient harm.
   b. The estimation of risk should reflect the existing risk classification scheme of the regulating government agency.
      i. If further nuance of risk is necessary, it should be situated ideally within one of the agency's existing risk classifications.
   c. Given context dependencies in risk estimation, policymakers must use existing regulatory bodies to govern healthcare AI.
      i. Existing regulatory bodies have experience with both healthcare contexts and larger healthcare industry considerations in regulation.
      ii. A centralized cross-industry regulation of AI increases the chances for duplicating rules already issued by other agencies and decreases sensitivity to the risks specific to the healthcare industry.

4. Preserving Safety Expectations
   a. AI-enabled systems must, as dictated by each system's purpose and context, conform to regulators' existing safety standards for medical diagnosis, treatment, and monitoring, including the monitoring of all adverse events and other problems.
   b. Policymakers must acknowledge scenario-specific limits of diagnostic and therapeutic science when setting expectations for the accuracy of healthcare AI findings or for patient outcomes for AI-enabled care.
      i. Current research for a given medical scenario may be neither perfect nor definitive.
      ii. Some obstacles to accuracy, such as underspecification[67] scenarios, are associated with statistical models and are not unique to AI.

[67] Underspecification, as discussed in Section I, occurs when the procedure for system training and validation produces multiple models of equivalent efficacy according to the test data and the system cannot predict which model will be more effective in real-world interactions.

    iii. Accuracy rates observed for competing, and similarly purposed, systems (or, in their absence, human clinicians) should be referenced when assessing risks versus benefits for an AI-enabled healthcare system where outcome perfection is not possible or expected.

c. In cases where healthcare AI does not require additional assistance from a clinician, the AI-enabled system must empirically demonstrate the fulfillment (to the satisfaction of the regulating agency) of three criteria:

    i. Accuracy levels (or patient outcomes depending on the nature of the system) equal to or exceeding the average rate observed for clinicians performing the same function for a similar cohort,

        1. Regulators will need to formally articulate criteria for what data acceptably establishes clinician accuracy levels.

    ii. No amplification of health risks as compared to when the same function is performed by a clinician, and

    iii. Output communications that are comprehensible and actionable for the patient without the additional explanation from a clinician.

d. Deficits in AI explainability do not, in and of itself, invalidate the system's outputs.

    i. While the top-level aspects of an AI system's operations are easily explained by system developers, the scale of data employed for system training and/or the number of artificial neurons used for pattern detection may result in a developer not being able to fully account for a system output (i.e. a diagnosis, disease prediction, treatment efficacy probability, etc.).

    ii. Limitations in the ability to fully explain an AI system output does not invalidate its output's empirical validity. It is widely accepted that there are FDA-approved medications where the mechanisms for the drugs' outcomes are not understood. For the FDA, the central concerns connected to a drug's lack of explainability are the drug's efficacy and safety, and these overriding principles apply equally to the regulation of AI-enabled systems.

    iii. In situations where an AI-enabled medical device has empirically validated benefits but explainability is suboptimal, regulators should consider requiring the AI developer to provide a disclosure characterizing the data used to train the AI system and the method for quality assurance testing.

e. In situations where policymakers believe an innovative category of AI-enabled technology is extremely promising but may not fit within an existing regulatory framework, they should consider creating a "regulatory sandbox" temporarily waiving specific (not all) regulatory obligations for a specified class of products or services.[68] A regulatory sandbox has a defined, short-term duration (e.g., one-year) that allows developers to demonstrate a product or service on a small scale and policymakers to evaluate the results in consideration of future evidence-based rulemaking for the category.[69]

5. Addressing Data Issues
   a. New rulemaking to safeguard sensitive patient information must not duplicate existing regulation, such as those enacted under HIPAA and the HITECH Act.
   b. Any rulemaking pertaining to data demographic expectations for healthcare AI must resemble those for non-AI technology performing a similar function. The method of programming in a healthcare software, whether AI or non-AI, must not be the determining factor regarding data demographic expectations.
   c. Regulators should recognize the differences between systems that use closed data sets for training and systems that can adapt based on new data inputs after their market release. Systems that improve their accuracy over time may require new testing protocols (and possible post-marketing data collection) at regulating agencies to prevent such systems for having to constantly re-apply for FDA approval and delay market access for system accuracy enhancements.

6. Preventing Discrimination
   a. Rulemaking to prevent discrimination must first establish that the intended protections are not already embodied in existing laws such as Title VI of the Civil Rights Act of 1964, the Age Discrimination Act of 1975, the Americans with Disabilities Act, and Section 1557 of the Patient Protection and Affordable Care Act.

[68] State Policy Network, "Everything You Need to Know About Regulatory Sandboxes," October 12, 2021, https://spn.org/articles/what-is-a-regulatory-sandbox/.

[69] West Virginia University, "What Are Regulatory Sandboxes?," November 2023, https://scitechpolicy.wvu.edu/files/d/210c95fc-4be2-4f4a-a865-5ef3f441c912/what-are-regulatory-sandboxes-3.pdf. See also an AI-specific discussion of regulatory sandboxes within Rea S. Hederman Jr. and Logan Kolas, "A Healthcare World Reimagined: How Big Government Threatens Healthcare AI and What to Do About It," Buckeye Institute, April 1, 2024, https://www.buckeyeinstitute.org/library/docLib/2024-04-01-A-Healthcare-World-Reimagined-How-Big-Government-Threatens-Healthcare-AI-and-What-to-Do-About-It-policy-report.pdf.

b. Regulations to combat demographic bias must recognize that AI-enabled healthcare systems may face limitations related to the data of the medical conditions they address. Some medical conditions may affect a very small, unrepresentative population, or cluster around a particular demographic factor such as age (e.g., pediatric cancers such as neuroblastoma), sex (e.g., prostate or ovarian cancers), or race and ethnicity (e.g., sickle cell disease and Tay-Sachs).

c. Rules requiring AI-enabled systems to reveal basic demographic traits associated with the system's training dataset must be exempted where demographic data does not apply (e.g., in generative AI drug molecule design) or where disclosure risks patient identification through reverse engineering the data.

d. Regulators must not dictate outputs — demographic or otherwise — determined by considerations irrespective of medical research and empirical data.

7. Improving AI Software

a. Regulation must protect the incentives for software improvement, including but not limited to feature enhancements and the remediation of known software anomalies that do not impair the system's safety or effectiveness.

b. Regulators should provide an economical pathway for innovators to re-apply for FDA approval on their devices where the functionality remains the same, but system autonomy increases over time. On this front, the FDA could leverage work already performed by the U.S. Department of Transportation for self-driving vehicles with differing levels of system autonomy (e.g., driver-assistance vs self-driving).[70]

[70] National Highway Traffic Safety Administration, "Standing General Order on Crash Reporting," https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting.