UCLA Institute for Technology, Law & Policy

Liability and Preemption in the New Regulatory Framework of Data Driven Healthcare

Akshay Sreekumar & Peter Horton

Liability and Preemption in the New Regulatory Framework of Data Driven Healthcare

Akshay Sreekumar & Peter Horton, UCLA Institute for Technology, Law & Policy

Abstract

Artificial intelligence (AI) and machine learning (ML) algorithms in healthcare are becoming increasingly powerful tools for diagnostic, therapeutic, and operational applications. Although these algorithms have made significant progress on incredibly difficult problems, their inner workings can be difficult to explain, and as such their low interpretability poses challenges for practitioners in clinical settings. Additionally, AI/ML algorithms can continuously improve their performance through the ingestion of new and relevant data, but it can be difficult to bound their performance across successive iterations. The regulation of software as a medical device (SaMD), and in particular AI SaMD, thus requires a shift from the traditional paradigm of medical device regulation to account for the continuous updates that an AI system might receive over the course of its lifetime. The FDA has proposed recent regulatory guidance that would encourage technical innovation while trying to ensure patient safety. This rapidly evolving regulatory framework that aims to keep up with the pace of technology has consequences for the liability of AI/ML device manufacturers. Courts will have to grapple with the novel features of AI/ML medical devices and the regulatory approach the FDA is taking when adjudicating questions of preemption—deciding to what extent the FDA's regulation bars civil lawsuits. If decisions about preemption and products liability close some avenues to recovery, injured patients face an uncertain path, since AI's development in healthcare complicates other liability questions, from the standard of care to the learned intermediary doctrine. Given the uncertainty of using liability to protect patients' interests, it is especially important for the FDA regulatory approach to be comprehensive and effective.

I. Introduction

Recent trends in the availability of large volumes of data combined with increases in computing power have enabled a class of deep data driven algorithms to effectively "learn" highly complex patterns. More specifically, deep neural networks of various architectures have emerged as a powerful class of algorithms capable of tackling previously insurmountable problems. Deep data driven algorithms rely on data to tune and set model parameters, rather than explicitly constructing models from a set of physical first principles. This paradigm is exceedingly useful when making inferences about complex phenomena where it may be difficult to even recognize the underlying fundamental mechanisms at play.

Healthcare has embraced numerous technological advances as part of the digital revolution in medicine, and now stands on the brink of an information revolution enabled by vast quantities of medical data. Deep learning algorithms have the power to diagnose disease, accelerate drug development, and recommend optimal therapies. However, several unique features of data driven algorithms raise concerns regarding liability, regulation, and patient safety.

II. Preemption in Healthcare

Discussions of how liability doctrines will evolve to apply to new technologies sometimes highlight the contrast between modern, rapidly changing technology and ancient common law principles. As a recent UCLA symposium's framing language puts it, "breakneck innovation" is paired with "centuries-old concepts which constitute the foundation of our modern legal system."[1] In some cases though, advances in artificial intelligence (AI) and machine learning (ML) technology are occurring against a backdrop of ongoing debates about liability issues in healthcare. Federal preemption in the field of medical device regulation is one doctrinal area in which the new technology will have to slot into a recently developing legal landscape.

The doctrine of preemption stems from the Supremacy Clause of Article VI of the U.S. Constitution, making federal law the supreme law of the land.[2] When preemption applies, courts invalidate the state or local law found to conflict with federal law. Preemption applies when explicitly invoked by Congress, or when courts infer from the language or the pervasiveness of federal regulation in a given area that state and local law would interfere with the objectives of Congress.[3]

In the 1970s Congress amended the FDA's mandate, the federal Food, Drug, and Cosmetic Act (FDCA), to include regulation of medical devices.[4] The backdrop for the decision to create the Medical Device Amendment of 1976 (MDA) was a wave of state tort actions and an emerging set of state medical device regulations.[5] To ensure that federal law would replace state regulatory systems and curtail excess tort claims, Congress included an express statement of preemption:

---

[1] UCLA JOLT's Special Issue Symposium on Governing the Digital Space, https://law.ucla.edu/events/ucla-jolts-special-issue-symposium-governing-digital-space.
[2] U.S. Const. art. VI.
[3] For an example of express preemption in the medical device context, *see Medtronic, Inc. v. Lohr*, 518 US 470 (1996); for implied preemption, *see Buckman Co. v. Plaintiffs' Legal Committee*, 531 U.S. 341(2001).
[4] 21 U.S.C. § Ch. 9; 21 U.S.C. § 360.
[5] *Riegel v. Medtronic, Inc.*, 552 U.S. 312, 315 (2008).

> Except as provided in subsection (b) of this section, no State or political subdivision of a State may establish or continue in effect with respect to a device intended for human use any requirement -
> (1) which is different from, or in addition to, any requirement applicable under this chapter to the device, and
> (2) which relates to the safety or effectiveness of the device or to any other matter included in a requirement applicable to the device under this chapter.[6]

Courts have wrestled with the language of the preemption clause in aligning the different tiers of FDA review with a wide range of state law claims. Devices that have gone through different pathways of FDA review will have been held to different sets of applicable requirements, which in turn means they will have preemption from regulation by different sorts of state laws.

Regulation of medical devices changes depending on the type of device. The FDA distinguishes medical devices into three classes. Class I devices are considered low risk and are subject to general rules about manufacturing and labeling. Class II devices are considered moderate to high risk and are subject to general rules and "special controls" that differ for different types of device and may include additional performance standards and labeling requirements.[7] Class II approval also includes the 510(k) pathway, where a manufacturer establishes that a device is similar to an already cleared device.[8] Class III devices, often devices intended for implantation, are considered high risk and are subject to the premarket approval process, wherein the FDA reviews their design, labeling, and manufacturing specifications and determines that those specifications provide a reasonable assurance of safety and effectiveness.

Products liability claims operate on three theories: manufacturing defect, design defect, and failure to warn.[9] The FDA's regulation of medical devices can preempt lawsuits on these theories, since "state laws" includes both common law and legislatively enacted law. A federal law with a preemption clause prevents a state or a local government from contradicting it: the legislature of California cannot create a products liability statute that holds all medical devices to their preferred standard of safety, that of Delaware cannot let all device manufacturers get away with the lowest rung of safety compliance. But as the cases discussed below all demonstrate, the preemption of state law also includes common law claims—bodies of law developed by judges through many different cases over time, including products liability claims.

In *Medtronic, Inc. v. Lohr,* the US Supreme Court held that while the premarket approval process involved "rigorous" review, devices cleared under the 510(k) pathway only had to establish similarity, not safety, and did not enjoy the same level of preemption as those with premarket approval.[10] A little over a decade after *Lohr*, *Riegel v. Medtronic* echoed this distinction and emphasized that the preemption clause applies to general tort claims, like the duty to warn, not only tort claims or other state laws that directly refer to medical devices, since "nothing in the

---

[6] 21 U.S.C. § 360(k)a.
[7] 21 U.S.C. § 360(a)(1)(A), § 360(a)(1)(B).
[8] 21 U.S.C. § 360e(b)(1)(B).
[9] RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 1.
[10] *Medtronic, Inc. v. Lohr*, 518 US 470, 477 (1996). The Court's reasoning in *Lohr* was partly due to the relatively recent introduction of the MDA: many Class III devices had been on the market before 1976, and these devices—and by extension those substantially equivalent to them—had not yet been reviewed by the FDA under the PMA process established in 1976. *Id.* at 477-78.

statutory text suggests that the pre-empted state requirement must apply *only* to the relevant device, or only to medical devices and not to all products and all actions in general."[11]

Riegel does not prevent all state common law tort claims. Tort claims that parallel the FDA requirements do not run afoul of the preemption clause. The caveat is that there needs to be some sort of existing tort claim in the given state's common law that the plaintiff can point to—e.g., a medical device or drug manufacturers' duty to report risks to consumers—to mold an action that amounts to essentially a form of private enforcement of the FDA's rules, under color of a tort claim. For example, courts have allowed some warning defect products liability theories as parallel claims.[12]

Although *Riegel* and *Lohr* did not create the same preemption protection for devices cleared under the 510(k) process as for those subject to premarket approval, devices reviewed through a Class II process have still been found to have preemption in some circumstances. In *Buckman Co. v. Plaintiffs Legal Committee*, the Court found implied preemption in a case involving a 510(k) device.[13] Other courts have noted that FDA special controls—requirements specific to different types of devices—in the Class II approval process distinguish the level of regulation further from the level of Class II review considered in *Lohr,* with general rules that applied to a "host of different devices"[14] and can meet a sufficient level of specific consideration for the preemption clause to apply.[15]

Concretely, this means courts need to dive into the FDA's approval process for a Class II approval, including the 510(k) pathway, looking at the requirements and the guidance (which, though nonbinding, indicates what the FDA views as compliant) for that specific kind of device. If the FDA examined and considered an aspect of the device in that process—for example, its labeling—the preemption clause applies and bars tort claims that would impose additional requirements for that aspect of the device.

In sum, the FDA's approval processes, even down to the agency's nonbinding guidance, has a major impact on patients' avenues for recovery after a device-related injury. After the FDA reviews and rubber stamps an aspect of a device's design or labeling, patients can bring a limited set of claims about manufacturers' failure to abide by FDA standards for things like unreviewed changes or unreported incidents, but they cannot use private cases to bring in other experts to second guess the FDA's determinations of safety. Consumers of FDA-approved medical devices simply have to hope that the FDA is fulfilling its mandate diligently.

AI devices add a new layer of complexity to the question of what features of a device the FDA has approved. One of the much-touted promises of AI in healthcare is the potential for algorithms that evolve and improve with exposure to new data. But a changing algorithm might need to seek new FDA approval routinely. In 2017, the FDA released guidance, "Deciding When to Submit a 510(k) for a Software Change to an Existing Device," saying that "if a manufacturer

[11] *Riegel v. Medtronic, Inc.*, 552 U.S. 312, 328 (2008).
[12] *Stengel v. Medtronic Inc.*, 704 F.3d 1224 (9th Cir. 2013).
[13] *Buckman Co. v. Plaintiffs Legal Committee*, 531 U.S. 341 (2001).
[14] *Lohr* 518 US at 498.
[15] *See, e.g., Papike v. Tambrands Inc.*, 107 F.3d 737, 740 (9th Cir. 1997).

modifies their device with the intent to significantly improve the safety or effectiveness of the device (for example, in response to a known risk, adverse events, etc.), a new 510(k) is likely required."[16]

New guidance from the FDA attempts to outline a form of approval that will create a process of sharing updates with the FDA as AI devices change without requiring full blown new applications for approval.

III. FDA Regulatory Approach for AI SaMD

In 2019, the FDA released a proposed regulatory framework that discussed potential modifications to the premarket review process to better accommodate AI SaMD.[17] The FDA recognizes that AI technologies can positively transform healthcare, but wants to ensure that there are still rigorous safeguards for patient wellbeing. To foster responsible innovation, the FDA's regulatory approach to AI SaMD aims to strike an appropriate balance that accounts for the evolving nature of AI SaMD. To this end, the FDA proposed guidance seeks to regulate the total lifecycle of a product.

The key idea proposed by the FDA is the incorporation of a predetermined change control plan that allows manufacturers to anticipate potential modifications and propose a plan for implementing them responsibly at the time of premarket approval. The aforementioned change plan consists of SaMD pre-specifications, and the algorithm change protocol. The SaMD pre-specifications effectively define the space of potential changes that the manufacturers imagine the device will "grow" into over its lifetime of operation. Accordingly, the algorithm change protocol highlights the specific actions that manufacturers will take to safely implement changes anticipated by the SaMD pre-specifications. The FDA intends for scrutiny during the premarket approval process to be effective in managing future risk over the lifecycle of a product, provided that the scope of the changes is limited.

The FDA broadly recognizes three types of potential changes to algorithms in their guidance: changes meant to affect performance, changes to input data types, and changes to intended use. In general, the FDA expects that the risk introduced by such changes can be mitigated by appropriately designed SaMD pre-specifications and algorithm change protocols but acknowledges that this framework is not equipped to handle all changes. In particular, changes that significantly alter the risk or intended use of a device may require a new premarket submission. There are open questions surrounding less significant changes that are still outside the purview of the SaMD pre-specifications and algorithm change protocol, such as when such changes should trigger a "focused review" from the FDA and to what level of scrutiny. In 2021, the FDA announced that it intends to release further draft guidance for comment to define the components

[16] FDA, *Deciding When to Submit a 510(k) for a Change to an Existing Device*, Guidance for Industry (Oct. 25, 2017).

[17] FDA, *Proposed Regulatory Framework For Modifications To Artificial Intelligence/Machine Learning (Ai/Ml)-Based Software As A Medical Device (SaMD)* (2019), https://www.fda.gov/media/122535/download.

of the SaMD pre-specifications, algorithm change protocol, and process of focused review more rigorously.[18]

Critical to the total lifecycle product approach proposed by the FDA is real-world monitoring of performance. The FDA recognizes the need for transparency in reporting to build trust, identify avenues for improvement, and address potential safety concerns. While the 2019 draft guidance contains several suggestions of mechanisms to support real-world monitoring, there are no established standards or guidelines yet that clearly define the required flow of information between manufacturers, the FDA, users, and patients. As of 2021, the FDA, through its action plan, noted that it supports real-world performance monitoring with relevant stakeholders on a voluntary basis. Significant work still remains to synthesize a cohesive framework that defines the role real-world performance monitoring will play in total product lifecycle regulation.

The FDA's total product lifecycle approach to AI SaMD regulation supports innovation in AI that could deliver valuable improvements in patient quality of care. While the proposed guidance by the FDA is in theory comprehensive via the SaMD pre-specifications and the algorithm change protocol, actual patient outcomes are highly dependent on how a predetermined change control plan is implemented, updated, and monitored over the course of a device's operation in the field. Without concrete mandates on performance monitoring, the proposed regulatory framework addresses the continuously changing nature of AI only in part.

At a fundamental technical level, one of the primary barriers to widespread adoption of AI in safety-critical systems is the inability to provide robust bounds on AI error and guarantees of model performance.[19] Data driven algorithms depend significantly on the quality and validity of the underlying data on which they were trained. Notably, the performance of AI systems can degrade as inputs deviate from the original training distribution, leading to challenges in model generalization. For example, such degradation has been observed when the training data has had racial and/or gender biases that did not represent the target population. Studies have found that facial analysis algorithms for gender classification misclassified darker skinned subjects at a significantly higher rate than lighter skinned subjects.[20] The datasets on which such algorithms were trained contained examples of predominantly light skin subjects, which translated into an algorithmic bias against certain groups of people and reduced model accuracy.

Degradation can also occur even with what appears to be a more representative set of training data. Although developers have increasingly tried to correct for bias, omissions, and oversights in training data, AI systems must still adapt to unforeseen changes. In particular, once an AI system has been deployed for use in a real clinical setting, it must deal with the issue of distribution drift. In distribution drift, many potentially subtle changes over time may accumulate such that the input to an algorithm is no longer from the same distribution as the data the algorithm

---

[18] FDA, *Artificial Intelligence/Machine Learning (Ai/Ml)-Based Software As A Medical Device (SaMD) Action Plan* (Jan. 2021), https://www.fda.gov/media/145022/download.

[19] Alessandro Biondi et al., *A Safe, Secure, and Predictable Software Architecture for Deep Learning in Safety-Critical Systems*, 12 IEEE EMBED. SYST. LETT. 78–82 (2020).

[20] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 15.

was trained on. Seemingly small but critical procedures that are idiosyncratic to individuals or hospitals can be difficult to anticipate and design for before the model is released into the wild. In the analysis of slide tissue, for example, samples are often stained with certain chemicals to highlight regions and molecules of interest before further inspection under some kind of microscope. While there do exist standard procedures for different types of stains, it can be difficult to control for slight differences in preparation and technique across hospitals, much like how individuals asked to cook a dish from the same recipe will inevitably produce slightly different meals. In histopathology applications, staining differences for slides were the primary source of variation in similar datasets across different hospitals.[21] Legal scholars have also pointed out that many medical algorithms are being trained in flagship hospitals, with access to a high level of resources, both in technology and in staff and specialists. Such algorithms may experience distributional drift when translated into the context of hospitals with more resource constraints.[22]

Due to the oftentimes opaque inner workings of such algorithms, humans may also not be able to judge accurately what constitutes minor changes in input a priori. As a consequence, what humans view as small perturbations can have a relatively outsized impact on model performance. Several researchers have explicitly engineered adversarial attacks on traffic signs that cause neural nets to classify them incorrectly as speed limit signs.[23] While these modifications have no bearing on human understanding of the stop sign, they catastrophically affect the model's classification ability. Thus, it is in general difficult to provide concrete guarantees of robustness for AI algorithms. Consequently, AI systems in production are constantly adapting by training on new data to not only improve their performance but necessarily maintain it as well. While this means that AI can improve its own performance in the field by continuously learning, it also implies that AI as a medical device must be viewed in the context of constant change throughout its lifetime. Researchers at MIT found significant degradation over time in the accuracy of a sepsis prediction model developed by Epic.[24] Their study showed that operational changes such as expansion of the hospital system and modifications to disease coding had a substantial negative impact on model accuracy. The adaptive nature of AI can correct for this to some extent through tuning and retraining of the model, but while this may be beneficial in the short term, it does not give assurances of long-term stability.

In the FDA's own language, AI SaMD systems are "unlocked" algorithms in which the input-output mapping of the algorithm is not guaranteed in successive iterations. This raises questions with interesting legal implications. With continuous updates, what defines the scope of how different an AI SaMD can become over time while still remaining the same device from a regulatory standpoint?

---

[21] David Tellez et al., *Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology*, 58 MED. IMAGE ANAL. 101544 (2019).
[22] W. Nicholson Price II, *Medical AI and Contextual Bias*, 33 HARV. J.L. & TECH. 65, 91-94 (2019).
[23] Kevin Eykholt et al., *Robust Physical-World Attacks on Deep Learning Models*, ArXiv170708945 Cs (2018), http://arxiv.org/abs/1707.08945; Nir Morgulis et al., *Fooling a Real Car with Adversarial Traffic Signs*, ArXiv190700374 Cs Eess (2019), http://arxiv.org/abs/1907.00374.
[24] Janice Yang et al., *AI Gone Astray: Technical Supplement*, ArXiv220316452 Cs Stat (2022), http://arxiv.org/abs/2203.16452.

Premarket approval under the new system puts unlocked AI/ML device manufacturers in a more promising position for receiving preemption than under the old system. Under the reasoning outlined by the court in *Kelsey*, the preemption analysis involves looking at what aspects of a device the FDA has reviewed to see what federal requirements have been applied to the device; those federal requirements then displace conflicting state law. If, under the total lifecycle approach with the SPS and the ACP to reviewing these devices, the FDA has considered the process the algorithm follows for its updates, a court can conclude that the changed software, several years after review, is still essentially the device that the FDA reviewed.

III.i. Interactions of New FDA Regulation and Preemption

Must courts defer to the FDA's assertion that a medical device is the same device they approved years before after it has altered the dataset it uses and the recommendations that it makes? Time may tell whether there is a limit to this deference to the FDA. Once a court concludes that preemption applies, preemption prevents second guessing the FDA's determinations. It isn't up to the plaintiff's experts, the jury, or the judge, any more than it would be to the state legislature, to reevaluate the technical determinations and policy decisions the FDA has made.[25] But in the question of whether to apply preemption, courts are the experts on statutory interpretation, and they must still do their own reasoning. Indeed, the Supreme Court emphasized this in *Buckman* when largely dismissing an FDA regulation that interpreted the MDA's preemption clause narrowly. While the FDA is asserting, through the new guidance, that a device with an algorithm approved through a Predetermined Control Plan can be treated as the same device after updates, the courts can make that metaphysical determination on their own.

The MDA refers to "this device"—a device that surely in the minds of the statute's drafters would not change over time. If the total lifecycle review proves to be completely inadequate, and algorithms become completely transformed while still technically approved by the FDA's SaMD AI/ML plan, it would be doctrinally consistent for a court to decline to consider the changed device the same as the approved device, on either a textualist theory or a theory of Congressional intent. This is a power courts should not hesitate to use if the FDA's review of these devices fails to keep pace with their evolution.

There is reason to suspect that courts will defer to the FDA and will consider devices to be the same despite changes over time. There is technological and regulatory precedent for the accumulation of changes in medical devices. The FDA 510(k) pathway for medical device approval is based largely on establishing "substantial equivalence" to previously approved devices, without the need for new clinical testing. It has been observed that decades of 510(k) approvals can result in chains of devices that differ significantly from the original device.[26] In one such case, a series of metal-on-metal hip implants cleared through 510(k) pathways based on prior devices

[25] *See, e.g., Duggan v. Medtronic, Inc.*, 840 F. Supp. 2d 466, 471-72 (D. Mass. 2012). ("The FDA, not litigants, is entrusted with the responsibility to police the sufficiency of the evidence to support a PMA approval.")

[26] Thomas J. Hwang, Aaron S. Kesselheim & Kerstin N. Vokinger, *Lifecycle regulation of artificial intelligence and machine learning–based software devices in medicine*, 322 JAMA 2285 (2019).

that were not shown to be safe or effective were eventually discontinued.[27] While the lifecycle approach proposed by the FDA uses the SPS and ACP as tools to better regulate the evolution of a product, a lack of clear rules defining scope and reporting leaves this question largely unsettled.

Even if courts find credible the FDA's claim that approving the plan for updates makes the updated device the same as the approved device, parallel claims could help consumers ensure that manufactures comply fully with the FDA's plan. Under the current preemption regime, patients have used failure to report adverse events as one form of parallel claim.[28] One part of the FDA's new plan for regulating AI/ML SaMD is real world performance monitoring. If a patient is harmed by an evolving algorithm, they might look into whether the designer is complying with the FDA's (presently vague) RWPM requirements. If the algorithm has been degrading, perhaps due to shifts in data, and the designer has not reported the changes to the FDA, a plaintiff could bring a parallel claim.

However, in addition to alleging a failure to meet the FDA's requirements, plaintiffs would still have to find a common law hook. In *In Re Smith*, the court held that a general state common law duty to warn (rather than a specific duty for medical devices) could be the basis for a parallel claim about failure to meet FDA requirements for reporting.[29] A general duty to warn applies just as well to failures to submit real-world performance data. However, courts in a number of states have held that there is no state common law claim for such failures, suggesting there may also be many states that won't allow parallel claims for RWPM requirements.

The impact of the FDA's new guidance on liability for the creators of AI/ML devices will depend on how much preemptive force courts give the new premarket approval pathway. At one extreme, courts might find the pathway preempts all design and labeling products liability claims, as courts have interpreted PMA approval, and might allow only a small universe of parallel claims. At the other extreme, courts might reject preemption entirely if they disagree with the FDA that total life cycle approval makes two distinct pieces of software (at two different points in time) the same medical device in the eyes of the MDA preemption clause.

The weight of how courts decide these preemption questions is also situated in a larger context. Would reading the FDA's guidance as authorization for broad preemption mean reducing designers' liability in one respect while leaving them vulnerable in other ways? Or would it add another layer of shielding to an already thick armor against liability?

We would argue that the latter is true: an expansive view of preemption is less desirable because the developers of AI in healthcare have many other ways to avoid liability.

III.i.a. Learned Intermediary

One relevant barrier to liability for AI devices is the learned intermediary doctrine. Under the doctrine, the manufacturer of a prescription medical drug or device has a duty to warn doctors,

---

[27] Brent M. Ardaugh, Stephen E. Graves & Rita F. Redberg, The 510(k) ancestry of a metal-on-metal hip implant, 368 New England Journal of Medicine 97–100 (2013).
[28] *E.g., In re Smith & Nephew Birmingham Hip Resurfacing (BHR) Hip Implant Prod. Liab. Litig.,* 300 F. Supp. 3d 732 (D. Md. 2018).
[29] *Id.* at 740.

but not patients, of any risks. The doctor is then responsible for conveying whatever information the patient needs to know to make an informed decision about treatment. A failure-to-warn claim against a device manufacturer must allege that the manufacturer failed to adequately warn a patient's physician of the relevant risks. They must also show that the absence of those warnings is what led to the patient's injury.

According to a recent nationwide survey, in 34 jurisdictions either the legislature or the highest court has approved the learned intermediary rule for prescription drugs, and in 9 more an intermediate court has. Fewer states have expressly approved the rule for medical devices: 14 in the legislature or highest appellate court, and 10 in intermediate courts. Federal courts hearing state law claims have also interpreted state common law to anticipate that more states will adopt the doctrine for both devices and drugs.[30]

The same questions of interpretability that could make it hard to preapprove the scope of an AI device could also make it hard for physicians to come to a full understanding of the risks of AI products. However, the low interpretability of AI devices isn't likely to create greater risks of products liability suits from end-users over failures to warn. Courts typically look for three elements: "the existence of [a] physician-patient relationship, (2) the physician's involvement selecting the product, and (3) the physician's superior understanding of the interplay between the product's dangers and the patient's condition."[31] The learned intermediary doctrine does not require that prescribing physicians have a level of expertise in the nature of the device or drug beyond their general medical expertise: a physician does not need to have the structural engineering literacy to evaluate a manufacturer's choice of metal in an implanted device, for example. In this sense, applying the learned intermediary rule to AI devices is not a departure from current practice, and the learned intermediary doctrine would likely apply for manufacturers of AI devices.

III.i.b. Software Products Liability

Another simple barrier to liability is that courts have been reluctant to apply products liability to software at all in any context, not just medicine. A suit against Snapchat – an instant messaging app – for a design defect in their app that encouraged risky behavior framed this question for the Ninth Circuit, but the case was remanded without reaching the issue.[32] A recent lower court case, *Holbrook v Prodomax Automation Ltd*, is an early example of a court making such an extension, but the decision turned on a state products liability statute and may not be generalizable to other states.[33]

The reluctance to extend products liability to software reflects software's intangibility and the fact that software often replaces human activity: operating machinery, driving, analyzing

[30] *The Closing Of The Learned Intermediary Frontier*, Drug & Device Law (June 2, 2011), https://www.druganddevicelawblog.com/2011/06/closing-of-learned-intermediary.html; *Headcount II: The Learned Intermediary Rule And Medical Devices*, Drug & Device Law (Jul 10 2008), https://www.druganddevicelawblog.com/2008/07/headcount-ii-learned-intermediary-rule.html.
[31] *Butler v. Juno Therapeutics, Inc.*, 2021 WL 2156742 (S.D. Tex. 2021).
[32] *Lemmon v. Snap, Inc.*, 995 F.3d 1085 (9th Cir. 2021)
[33] *Holbrook v. Prodomax Automation Ltd.*, No. 1:17-CV-219, 2021 WL 5052101 (W.D. Mich. Oct. 15, 2021).

images, noticing patterns. These activities, carried out by a person, would be subject to liability on a negligence standard: someone becomes liable when they fall below the precautions a reasonable person takes in their activity. Products liability is a form of strict liability, where negligence isn't a component of liability. A human driver is liable for injuries caused in a crash if they were unreasonable, say by speeding or not looking for pedestrians; a software driver could be liable under products liability if a safer design is theoretically possible, without regard to whether the AI driver was taking all the precautions a reasonable human driver would take and more.

AI devices in healthcare are currently not well covered by either realm. There isn't yet a body of precedent establishing products liability for software. But there's also no established path for patients to sue an algorithm that replaced a radiologist for malpractice, a negligence theory. Neither product nor person, AI devices can't neatly be held liable in either category.

III.i.c. Malpractice and the Standard of Care

We note one final legal doctrine that is on the periphery of the FDA guidance. Some AI devices with healthcare applications will not be classified as medical devices, and will not be subject to FDA approval. Clinical Decision Support Software (CDS) is a prominent example. CDS devices are already being used in US hospitals, like the supercomputer Watson, which offers diagnostic support for cancer patients and suggests treatments.[34] Because the FDA regulates drugs and devices, not the practice of medicine, CDS has been considered outside of its purview.

As AI tools that provide diagnoses and treatment recommendations become more advanced, it may be hard for doctors to meaningfully reflect on recommendations from the software, and allocation of liability will be complicated. Medical malpractice suits hinge on showing whether a physician's actions were negligent, which courts define as falling below a standard of care. This standard of care is set by the collective actions of the physician's professional peers, based on expert testimony about typical practice and standards set by professional bodies. A minority of U.S. jurisdictions may consider physicians negligent even if they followed national custom, if a ''reasonably prudent'' physician would have followed another treatment that might have averted the patient's injuries. [35]

If an AI recommendation deviated from typical practice, especially when reliance on AI tools is still itself not widespread, a doctor would have reason to worry that they would be liable for malpractice should the patient suffer any injury.[36] This cautionary bias could limit AI's potential to increase discoveries of improved care.

But after a tipping point, where data accumulates that AI recommendations are significantly better than typical practice in some specific context, doctors might be liable for ignoring AI recommendations, even when following the standard of care.[37] In *Burton v. Brooklyn Doctors Hospital*, a course of treatment was held to be malpractice, despite remaining widespread,

---

[34] *IBM Watson Health*, IBM, https://www.ibm.com/watson/health/oncology-and-genomics.
[35] *Helling v. Carey*, 519 P.2d 981 (1974).
[36] *See* Gary E. Marchant & Lucille M. Tournas, *AI Health Care Liability: From Research Trials to Court Trials*, J. HEALTH & LIFE SCI. L. 23, 31 (2019).
[37] *Id.*

because research increasingly showed that it was too dangerous to justify for any prudent physician.[38] Similarly, if AI becomes markedly better than doctors at a task—for example, setting dosages of a drug after a cancer diagnosis—doctors may become liable for not adopting the AI's recommendation, even if it departs from their training and the practice of many of their peers.

Either of these outcomes has some negative implications. If doctors do not consider AI recommendations when they deviate from typical judgment, some potential is lost for new discoveries.[39] But if doctors can be sued for not following AI recommendations, they may over rely on AI, without a meaningful chance to evaluate AI recommendations. The justification for not having the FDA review CDS tools is that they do not directly, physically interact with the patient in the same way that a medical device or a drug does. However, when CDS recommendations come from black box algorithms and do not explain their recommendations or diagnoses to doctors, the doctor may give up agency. If doctors aren't acting as a check on AI tools, the FDA has more reason to look for ways to take over oversight.

For these reasons, it is important for the FDA to give clear rules on when AI in the medical field should be considered a device, and subject to FDA approval, and when it is a support tool. Below, we will discuss the FDA's ability to cooperate with hospitals to create best practices and expand institutional capacity to monitor AI performance. This is a response to the FDA's medical device regulatory approach, but it can also be relevant to helping hospitals manage AI tools that are not directly under the FDA's regulatory authority.

IV. Future Landscape of Safe AI in Healthcare

IV.i. Limitations of Technology Driven Solutions

The field of explainable AI (XAI) has received a significant amount of attention in recent years, particularly for fields like healthcare where professionals have expressed concern over the black-box nature of AI algorithms.[40] XAI aims to peel back the curtain on AI decisions by providing justifications or suggestions as to why particular decisions were made by an AI.[41] While XAI has been touted as a promising avenue to increase trust and adoption of AI systems, there are notable challenges that suggest it is not a panacea for all regulatory issues in AI.

For high dimensional and complex AI models, the primary XAI technique used is post-hoc explainability which aims to make the decision-making process transparent *after* a decision has been made. This is fundamentally different from "inherent explainability" of an AI which provides a clear relationship between how inputs meaningfully map to outputs.

---

[38] *Burton v. Brooklyn Doctors. Hosp.*, 452 N.Y.S.2d 875 (N.Y. 1982).

[39] Of course, patients shouldn't be used as unwitting guinea pigs; doctors would open themselves up to other ethical violations and malpractice liability for prescribing an experimental treatment without getting informed consent. *Shadrick v. Coker*, 963 S.W.2d 726, 732 (Tenn. 1998).

[40] Urja Pawar et al., *Explainable AI in Healthcare*, *in* 2020 INTERNATIONAL CONFERENCE ON CYBER SITUATIONAL AWARENESS, DATA ANALYTICS AND ASSESSMENT (CYBERSA) 1–2 (2020), https://ieeexplore.ieee.org/document/9139655/ (last visited Jul 28, 2022).

[41] Derek Doran, Sarah Schulz & Tarek R Besold, *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives*, 8 (2017).

In medical imaging, attention maps are one of the most commonly implemented forms of post-hoc explainability XAI. Attention maps work by highlighting regions in an image which a model's decision depended on significantly, effectively showing which parts of an image were salient to the AI decision making process.[42] However, the resultant attention maps themselves need to be interpreted by a human user, which adds a complicating dimension to the process of understanding the underlying model. Humans are inclined to ascribe meaning to an attention map that aligns with their given understanding of the world. For example, in diagnosing whether a sample has a tumor from a tissue slide, an attention map highlighting a particular cluster of cells may make intuitive sense with a physician's medical understanding of pathology and physiology. In reality, it is in fact unknowable whether the *model's* reasons for using those regions are rooted in those same medical factors. The emphasis on a particular region could be the result of anomalous pixel values, image acquisition artifacts, or staining discrepancies that are opaque from the human perspective and non-medical in nature. Because attention maps only offer a glimpse into the model's behavior, their interpretation can be compromised by human biases that reinforce a false sense of trust in a model's predictions. A study from Siemens showed that images that have been adversarially modified to produce an incorrect prediction by a model will still effectively yield the same attention map when queried for explainability.[43] Researchers at MIT and Harvard argue that the focus on XAI as a tool for practitioners and patients is misguided.[44] The interpretation of XAI in local decision-making settings is susceptible to confirmation bias, but post-hoc explainability methods can be valuable tools for model developers. Studying attention maps, for example, provide a global view of how a model behaves in various settings and can be used to justify model tuning to ensure that the model's focus is broadly aligned with expectations.[45] Thus, XAI may in fact be better used as a backend tool for the development of better AI rather than as a forward deployed mechanism for fostering trust between users and AI systems.[46]

In the context of FDA regulation of medical devices, it is not the FDA's duty to encourage the use of AI in the practice of medicine by building trust between AI tools and practitioners. Rather, the FDA's mission is "assuring the safety, efficacy, and security…of medical devices", from which trust arises as a natural byproduct of effective regulation.[47]As such, any XAI requirements imposed by the FDA are tangential to its true regulatory mandate. Scholars have argued that the traditional practice of medicine has relied heavily on building confidence in safety not necessarily through explainability but through rigorous validation and testing.[48] Notably, the biological mechanisms of acetaminophen are only partially understood but it has been widely

[42] Randy Goebel et al., *Explainable AI: The New 42?*, 11015 *in* MACHINE LEARNING AND KNOWLEDGE EXTRACTION 295–303 (Andreas Holzinger et al. eds., 2018).
[43] Jindong Gu & Volker Tresp, *Saliency Methods for Explaining Adversarial Attacks* (2019), https://arxiv.org/abs/1908.08413.
[44] Marzyeh Ghassemi, Luke Oakden-Rayner & Andrew L Beam, *The false hope of current approaches to explainable artificial intelligence in health care*, 3 THE LANCET DIGITAL HEALTH e745–e750 (2021).
[45] *Id.*
[46] *Id.*
[47] Mission Possible: How FDA Can Move at the Speed of Science, 38.
[48] Ghassemi, Oakden-Rayner, and Beam, *supra* note 44.

accepted because of a mountain of empirical evidence supporting its safety and efficacy.[49] In this sense, the FDA already has extensive experience in regulating technologies that are "black-box" by requiring careful scientific evidence to back up claims. Of course, a defining feature of AI that separates it from other drugs or devices is its inevitable evolution. To handle this, the FDA needs to establish reporting, transparency, and monitoring criteria that are reviewed periodically and consistent with a total product lifecycle approach to regulation.

IV.ii. FDA Regulatory Reporting Recommendations

In its 2021 action plan, the FDA responded to stakeholder feedback regarding its original proposed guidance released in 2019. Among these topics, it is critical to clarify how real-world performance (RWP) monitoring would work in practice to curb the ramifications of a persistently evolving AI system. The FDA is currently working with manufacturers to pilot RWP monitoring on a voluntary basis, with the intention of gathering data to establish a framework for RWP parameter, threshold, and evaluation methodologies.[50] As of now, this process does not define any enforceable mandates for RWP monitoring to be included as part of the review process. It is imperative that the FDA codify RWP monitoring best practices as requirements in the approval process itself so as to mitigate any liability impacts during the lifecycle of a product. There are several pertinent questions that need to be answered to define the scope of RWP monitoring in practice. Consider the following from the 2021 action plan: (i) What type of reference data are appropriate to utilize in measuring the performance of AI/ML software devices in the field? (ii) How much of the oversight should be performed by each stakeholder? (iii) How much data should be provided to the Agency, and how often?

With regard to (i) above, the FDA should at the time of premarket review establish with manufacturers some form of reference data against which to measure in field performance. Of course, the specific nature of this data is highly application dependent, but it should be aligned with the targets established in the ACP. A cohesive plan should consider (ii) and (iii) simultaneously to define the nature of the manufacturer's relationship with the FDA during the lifecycle of the product.

Based on the specifics of the SPS and ACP, the FDA should establish mandatory review requirements for manufacturers at regular intervals to ensure that there is not an undesirable accumulation of changes that significantly alter a device's behavior. These reviews should consider comparisons to reference data to assess efficacy, drift monitoring statistics to check for distributional shift, and validation of previously made changes to ensure that the ACP is being properly followed. A group of scholars at Harvard suggests that this could be taken even further by limiting the lifetime of a predetermined change control plan based approval to have a limit of five years, effectively guaranteeing that there is a hard cap on the number of accumulated changes. However, they note that a time-limited regulatory authorization pathway would likely require new

[49] K. Toussaint et al., *What do we (not) know about how paracetamol (acetaminophen) works?: Paracetamol's analgesic mechanism?*, 35 J. CLIN. PHARM. THER. 617–638 (2010).
[50] FDA, *supra* note 17.

legislation.[51] Furthermore, the FDA's actions to date have indicated that they are interested in managing risk over a device's lifetime through implementing regulatory safeguards that allow for safe evolution rather than strictly limiting the product's life.

A systems view of AI enabled healthcare emphasizes the interactions between the devices, algorithms, and interpretation of results that comprise the decision-making chain. Biases in any of these links can compromise the efficacy and reliability of an AI tool deployed in medicine. While the existing total product lifecycle management framework along with sufficient RWP requirements is well positioned to regulate the devices and algorithms themselves, a more local regulatory architecture is needed to eliminate bias in interpretation and practice.

Because AI systems are highly context dependent and can evolve differently in different hospital settings, there should be mechanisms in place to monitor this variability and trigger additional scrutiny outside of regularly scheduled reviews. The FDA is in a position to enable this by requiring manufacturers to cooperate with and support the implementation of AI systems in hospitals. Researchers at Harvard argue that although the FDA can neither regulate the practice of medicine nor the training of medical professionals, it can require manufacturers to support hospitals through a variety of operational measures. Specifically, they suggest that manufacturers provide monitoring, retraining, and inspections to ensure AI systems are functioning and being used properly by practitioners.[52] Additionally, manufacturers can review aggregate usage statistics in conjunction with the hospital to identify possible drifts and discrepancies. This relationship could also empower hospitals to demand additional model validation and robustness guarantees via re-training with different subsets of data and adversarial perturbations.[53] This allows for a more local regulatory architecture where manufacturers are beholden to the constraints of specific deployment settings that would otherwise be difficult for the FDA to broadly capture through global regulation.

There is regulatory precedent for the FDA mandating specific criteria like those enumerated above. IDx-DR is an FDA approved software that uses AI to screen for diabetic retinopathy by analyzing retinal images taken with a specific retinal camera. The FDA mandated a training program that included instructions on image acquisition, tuning image quality, and submitting images for analysis.[54] Given the paradigm shift to the total product lifecycle approach to handle constant evolution of AI systems, it is natural for the FDA to promote more local regulation to ensure that sustained and periodic validation of AI systems can occur throughout the product life cycle. The FDA will still act as a central authority for the collection and final review of information to regulate manufacturer behavior and devices.

---

[51] Thomas J. Hwang, Aaron S. Kesselheim & Kerstin N. Vokinger, *Lifecycle regulation of artificial intelligence and machine learning–based software devices in medicine*, 322 JAMA 2285 (2019).
[52] Sara Gerke et al., *The need for a system view to regulate artificial intelligence/machine learning-based software as medical device*, 3 NPJ DIGIT. MED. 53 (2020).
[53] *Id.*
[54] FDA, *De Novo Classification Request for IDx-DR*, (2018), https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180001.pdf.

The above model of mandating rigorous reporting and review requirements addresses several closely linked challenges associated with AI. Broadly, this approach addresses the numerous algorithmic transparency concerns that surround the use of AI systems. The issue of AI explainability (fundamentally an issue of decision-making transparency) has already been considered in great depth. There, the solution was not necessarily more explainable AI but thorough and careful validation. Validating the end-to-end results of AI frequently and meticulously inherently captures other transparency issues like training set bias as well. While it would be ideal for the FDA to screen training sets for bias, manufacturers may be unwilling to divulge proprietary information publicly and it may be infeasible for the FDA to accurately understand when there is a bias present.[55] Rather, scrutinizing the system through validation, data, and testing can reveal deficiencies in an AI that can then be corrected appropriately. While these aspects of technical transparency might be better handled through empirical testing, it is important that this slew of data and observations are made transparent to patients and the public. The FDA should require clinical and experimental data supporting changes covered by the ACP to be published by manufacturers as well as making public the data from validation processes carried out in conjunction with hospitals. The best way to effectively manage data driven algorithms is through data driven regulation. The FDA can use its regulatory mandate to generate vast quantities of clinical, operational, and scientific data that will be analyzed and responded to in a more symbiotic framework by regulators, manufacturers, and hospitals.

V. Conclusion

The coming years will only see a rise in the prevalence of AI/ML algorithms in healthcare. As these technologies continue to grow, patients will undoubtedly benefit from the advanced diagnostic and therapeutic capabilities unlocked by AI—from detecting various cancers earlier to the recommendation of optimal immunotherapies to fight disease. The safe adoption and integration of these technologies into the practice of medicine, however, will depend largely on the regulatory structures in place to nurture responsible innovation. The FDA proposed guidance in recent years has laid the groundwork for a new paradigm that recognizes some of the unique features of AI that make it both a powerful technology but tricky to regulate. The total product lifecycle approach to AI SaMD regulation from the FDA requires further clarification on real-world performance monitoring, reporting, and transparency between stakeholders. As the FDA continues to refine and push forward this regulatory framework, courts will have to contend with issues of preemption and to what degree the new regulations limit civil claims.

This comment has explored these issues by considering fundamental features of AI/ML as it currently exists, and discussing how they complicate the balance of innovation and safety in

[55] One avenue for further exploration is whether the proprietary information important to AI SaMD manufacturers could be presented to the FDA but kept from public disclosure through the FDA's existing procedures. 21 C.F.R. §§ 20.21, 20.61(c). See also Douglas Nemec et al., *Protecting Trade Secrets Disclosed To The FDA*, PORTFOLIO MEDIA, INC. (Feb. 13, 2018), https://www.skadden.com/-/media/files/publications/2018/02/protecting_trade_secrets_disclosed_to_the_fda.pdf.

regulation. Constant algorithmic evolution coupled with models that are difficult to bound mean that any regulation must be flexible enough for change but robust enough to ensure safety and transparency for clinicians and patients. Courts will have to look for a balance that incorporates positive aspects of the tort system as a protection for consumers and a fallback in case of under-regulation, while also maintaining preemption's function as a way to prevent the unpredictability and cost of having multiple state regulatory schemes that conflict with federal oversight.