PLAYING CATCH-UP: FDA REGULATION OF AI/ML CLINICAL DECISION SUPPORT SOFTWARE

Kirby Ammons[*]

Abstract:

With the significant growth in technological capabilities of Artificial Intelligence (AI) and Machine Learning (ML), there are increased calls for government regulation to ensure safe implementation across numerous industries. The FDA's kludgy attempts to design a workable regulatory framework for AI/ML used by Health Care Providers (HCPs) to make decisions in healthcare over the last three decades are coming to a conclusion. However, the FDA's current approach as published in draft guidance documents fails to expand upon statutory language, leaving significant vagueness regarding what is required of AI/ML manufacturers to obtain FDA approval. Depending on interpretation of this vague language, the current approach risks overregulating or underregulating the industry and ultimately causing injury to patients if more clarification is not provided and enforced by the FDA. This article asserts that the current approach creates a "right to explanation" to patients by way of their HCPs, but that merely requiring an "explanation" is insufficient to adequately protect patients and encourage manufacturers to develop better AI/ML algorithms. Finally, this article offers a straight-forward solution that would provide clarity and reduce risk: requiring counterfactual explanations that provide "if-then" statements to identify influential variables to algorithms to HCPs without inundating them with information to review or failing to provide them pivotal information to review before making health decisions with patients.

[*] JD, Vanderbilt Law School, 2021; BS, United States Military Academy, 2015; Captain, U.S. Army Judge Advocate General Corps. The views presented here are those of the Author and do not reflect the views of the U.S. Army. The Author would also like to thank Professors Owen Jones and Lisa Bressman for their advice and guidance.

TABLE OF CONTENTS

I. INTRODUCTION

Artificial Intelligence (AI) that can detect skin cancer more effectively than board certified dermatologists,[1] determine the most effective anti-depression medicine based on an electroencephalogram (EEG) reading,[2]

[1] Ofer Reiter et al., *Artificial Intelligence in Skin Cancer*, 8 CURR. DERM. REP. Curr. 133, 133–40 (2019) (discussing that AI performs as well as or even better than human raters but adoption is still in early stage).

[2] Wei Wu et al., *An Electroencephalographic Signature Predicts Antidepressant Response in Major Depression*, 38 NATURE BIOTECHNOLOGY 439, 439–47 (2020)

and can identify and track the spread of the 2019 Coronavirus with precision[3] – these are no longer distant ambitions to be sought by Frankenstein-like personalities, but rather near eventualities in everyday healthcare. Over the last decade, researchers have pushed the bounds of AI to new heights, both in terms of applicability and capability of AI.[4] As these examples demonstrate, healthcare in particular has seen significant growth and promise.[5]

With this progress, distrust in the systems has also grown because of increased complexity and opacity.[6] Non-computer scientists do not understand the programming involved in construction and training of AI systems. For example, what most people are talking about when they discuss AI is a subsect of AI called Machine Learning (ML). By design, ML mimics the brain's ability to "think," but does not record the mechanical process of how the machine creates the algorithms. The conclusions generated may also be so complex that they are incomprehensible to humans, thus producing "Black-Box Medicine" or opacity inherent to the process when AI is used in healthcare.[7]

The combination of massive growth and general suspicion regarding the effectiveness of machines has led to increased calls for governmental regulation. In response, the federal government is attempting to manage the

(identifying a neurobiological signature of response to antidepressant treatment as compared to placebo).

[3] *See* John McCormick, *How AI Spotted and Tracked the Coronavirus Outbreak*, WALL STREET J. (Feb. 6, 2020), https://www.wsj.com/articles/how-ai-spotted-and-tracked-the-coronavirus-outbreak-11580985001 (small Toronto-based company BlueDot Inc. used AI to send an alert about the coronavirus outbreak the week before major health agencies issued notifications); *see also* Cory Stieg, *How this Canadian Start-Up Spotted Coronavirus Before Everyone Else Knew About It*, CNBC (Mar. 3, 2020), https://www.cnbc.com/2020/03/03/bluedot-used-artificial-intelligence-to-predict-coronavirus-spread.html (explaining thar BlueDot successfully spotted a virus multiple times in history).

[4] *See, e.g.*, Nat'l Sci. & Tech. Council and Networking & Info. Tech. Rsch & Dev. Subcomm., Exec. Off. of the President, The National Artificial Intelligence Research and Development Strategic Plan 5 n.4 (2016) [hereinafter 2016 National AI Strategic Plan]; Robin Feldman et al., *Artificial Intelligence in the Health Care Space: How We Can Trust What We Cannot Know*, 30 STAN. L. & POL'Y REV. 399 , 399–419 (2019) (discussing the pathways we use to place our trust in medicine provide useful models for learning to trust AI).

[5] *See infra* note 71.

[6] *See, e.g.*, Romain Cadario, Chiara Longoni & Carey K. Morewedge, *Understanding, Explaining, and Utilizing Medical Artificial Intelligence,* 5 NATURE HUMAN BEHAVIOUR 1636, 1636–37 (Oct. 2, 2020) (showing a reluctance to utilize medical algorithms is driven both by the difficulty of understanding algorithms, and an illusory understanding of human decision making).

[7] W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J. L. & TECH. 419, 421–22 (2015).

growth of AI against concerns over opacity and data privacy. In particular, the government is trying to balance the need for cultivating innovation with the need to ensure safe products to consumers; i.e., providing accountability.[8]

The FDA is no stranger to regulation of these kinds of products; in the 1980s, the agency became an unwilling participant in software regulation.[9] But a lot has changed since then. The FDA has begun to enact sweeping changes to the status quo.[10] The FDA regulatory scheme has traditionally been a congressionally-endorsed risk-based model.[11] However, the 21st Century Cures Act and subsequent FDA proposed guidance documents introduced a legally binding, yet vague rule that Health Care Professionals ("HCPs") be able to "independently review" recommendations made by AI systems, referred to as Clinical Decision Support Software ("CDS").[12, 13] This obligation has no stated purpose, but seems to be a shallow attempt at protecting HCP and patient autonomy by forcing manufacturers to provide some amount of reasoning behind the CDS software recommendation.

This article argues that the FDA has not provided clarity in their regulatory efforts because the agency lacks expertise in the area of AI software. Further, this lack of clarity creates the potential for industry to be under-regulated or over-regulated, potentially harming HCPs and patients, as well as CDS software manufacturers. This vagueness limits the manufacturing industry's ability to comply with the regulations, leading to the same issue of under regulation or over regulation, regardless of the FDA's actual, individual regulatory activity. This will have the effect of stifling

[8] 2016 National AI Strategic Plan, *supra* note 4. *See, e.g.*, *Summary of AI Provisions from the National Defense Authorization Act 2021*, STANFORD UNIV. HUMAN-CENTERED ARTIFICIAL INTELLIGENCE, https://hai.stanford.edu/policy/policy-resources/summary-ai-provisions-national-defense-authorization-act-2021 (last accessed Mar. 10, 2022).

[9] NATHAN CORTEZ, *ANALOG AGENCY IN A DIGITAL WORLD*, *IN FDA IN THE TWENTY-FIRST CENTURY: THE CHALLENGES OF REGULATING DRUGS AND NEW TECHNOLOGIES*, 438, 443–44 (Holly Fernandez Lynch & I. Glenn Cohen eds., 2015).

[10] *See* 2017 FDA DRAFT GUIDANCE CLINICAL DECISION SUPPORT AND PATIENT DECISION SUPPORT SOFTWARE, *infra* note 46; 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE, *infra* note 46; 2019 FDA DISCUSSION PAPER, AI/ML-BASED SAMD, *infra* note 139.

[11] *See infra* Section II (discussing FDA's interpretation of section 3060(a) of the 21st Cures Act and CDS).

[12] *Id.*

[13] CDS software is software that "provides clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance... decision making in the clinical workflow... including computerized alerts, . . . clinical guidelines, condition-specific order sets, . . . [etc.]." *Clinical Decision Support*, HEALTHIT.GOV (2019), https://www.healthit.gov/topic/safety/clinical-decision-support. Patient Decision Support ("PDS") software performs the same function, but specifically the user is the patient rather than the HCP. *Id.*

innovation and growth, thus reducing the availability of AI software that could improve patient care.

To prevent this, the FDA needs to articulate a clear standard for manufacturers that will ensure predictability and accountability. This is not to say that the FDA should become wedded to only one methodology; the FDA should remain flexible and willing to work with industry when the baseline standard explanation does not work on a case-by-case basis. That said, a well-articulated standard will give a foundational framework for regulating in this area.

In Part II, this paper outlines the FDA's current statutory and proposed regulatory scheme, highlighting the range of interpretations that the guidance documents can be read to impute on manufacturers of Software as a Medical Device (SaMD).[14] I argue that the legislative and regulatory framework is one borne of fear and misunderstanding of AI systems. This creates vagueness that ultimately serves to undermine manufactures' innovative prerogative and harms patients by presenting unpredictability and an over regulation and under regulation problem. Additionally, as written, the laws and regulations create a legal right to explanation, and this right is exercised by HCPs on behalf of their patients. Whether this should be the case is not evaluated in this paper.[15] Part III discusses several proposed frameworks for balancing safety and innovation and evaluates why they are insufficient under the current statutory framework. It also explores basic guidelines for establishing a workable standard, including why only requiring some explanation without more detail is inadequate to address the over regulation and under regulation problem. Part IV presents my solution of requiring counterfactual explanations as a possible baseline standard when submitting AI algorithms for FDA clearance. The paper compares the positive qualities of counterfactuals and then addresses the concerns regarding this method of explanation. Part V concludes that while a baseline standard is necessary to clear up FDA intentions and while counterfactuals provide a workable standard, the FDA needs to remain flexible, working with individual manufacturers when counterfactuals are inappropriate.

II. THE EXISTING REGULATORY REGIME

The FDA is responsible for protecting and advancing the public health through regulation of a number of products, including drugs, biologics, and

[14] *See generally* 2016 FDA DRAFT GUIDANCE, SOFTWARE AS A MEDICAL DEVICE, *infra* note 30.

[15] *See* Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTEL. 206, 206–08 (2019) (discussing the chasm between explaining black boxes and using inherently interpretable models).

medical devices.[16] What the FDA regulates is best understood by what they do not regulate. The FDA does not regulate physicians' or nurses' practice(s), what HCPs tell their patients, or rating schemes (quality determinations) for regulated medical devices.[17] In other words, the FDA does not regulate the practice of medicine.[18] The dividing line between practice and medical devices is difficult to determine when dealing with SaMD, and CDS software in particular.

When the FDA regulates medical devices, it does so under a risk-based model. The higher the risk a device poses, the higher the need for regulatory oversight of the device.[19] In this way, Class I and II medical devices receive less regulatory scrutiny than Class III devices.[20] New devices are classified through a process of premarket notification that treats any new devices as Class III until the device is reclassified. If the device cannot be reclassified, all new devices must go through a more onerous approval process called Premarket Approval (PMA).[21] This risk-based model makes sense in light of the overall mission of the FDA to enhance innovation and access while ensuring safety and effectiveness. This principle of focus on regulating devices based on risk is not changed when addressing SaMDs,[22] but the FDA struggled to come to this conclusion due to a lack of expertise.[23]

Following the FDA's decision to implement a risk-based model, the

[16] U.S. FOOD & DRUG ADMIN., *What We Do*, FDA.GOV (Mar. 28, 2018), https://www.fda.gov/about-fda/what-we-do. More directly, the FDA Mission Statement in relevant part reads: "The Food and Drug Administration is responsible for protecting the public health by ensuring the *safety, efficacy, and security* of human and veterinary drugs, biological products, and *medical devices* . . . [and for] advancing the public health by helping to *speed innovations* that make medical products *more effective, safer, and more affordable* and by helping the public get the accurate, science-based information they need to use medical products and foods to maintain and improve their health." *Id*. (emphasis added).

[17] U.S. FOOD & DRUG ADMIN., *FDA's Role in Regulating Medical Devices*, FDA.GOV (Mar. 28, 2018), https://www.fda.gov/medical-devices/home-use-devices/fdas-role-regulating-medical-devices.

[18] *See, e.g.,* Efthimios Parasidis, *Clinical Decision Support: Elements of a Sensible Legal Framework*, 20 J. HEALTH CARE L. & POL'Y 183, 191 (2018) (quoting Patricia J. Zettler, *Pharmaceutical Federalism*, 92 IND. L. J. 845, 885–86 (2017)) (discussing the amorphous boundary in the product/practice divide); Anna B. Laakmann, *Customized Medicine and the Limits of Federal Regulatory Power*, 19 VAND. J. ENT. & TECH. 2, 287 (2017); Nicolas Terry, *Of Regulating Healthcare AI and Robots*, 21 YALE J. L. & TECH. (SPECIAL ISSUE) 133, 149 (2019).

[19] 21 U.S.C. § 360c.

[20] U.S. FOOD & DRUG ADMIN., THE 510(K) PROGRAM: EVALUATING SUBSTANTIAL EQUIVALENCE IN PREMARKET NOTIFICATIONS, GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 2 n.1 (2014) [hereinafter 2014 FDA GUIDANCE, 510(k) NOTIFICATION].

[21] *Id*. at 2–3.

[22] *See infra* Section II.A.3.

[23] *See infra* Section II.A.

114th U.S. Congress passed the 21st Century Cures Act ("Cures Act") limiting FDA jurisdiction over CDS software.[24] With little Congressional clarity,[25] the FDA is left with wide discretion to interpret the Cures Act through agency guidance documents. However, the FDA has not offered much clarity to manufacturers, HCPs, or patients; instead, it has relied on language from the Cures Act itself and a few examples in proposed guidance documents.[26] The FDA's initial efforts to regulate CDS SaMD without a complete regulatory framework leads to an over and under regulation problem that burdens manufacturers and harms HCPs and patients.[27]

This article argues that the effect of the Cures Act and FDA guidance documents is the creation of a legal 'right to explanation.'[28] This right is vested in the HCP as the user of the algorithm and reflects standards of healthcare outside of CDS software.

A. Existing Laws and Regulations

The following provides a brief overview of the relevant language and content of the legislative and FDA actions leading up to and following the adoption of the Cures Act. This will provide a basis for analysis of what effect these standards have on manufacturers, HCPs, and patients.

1. Pre-Cures Act: Early Signs

The FDA's actions prior to the Cures Act reveal a collective lack of expertise to regulate emerging technologies. Between 1989 and 2014, the FDA struggled to decide if the risk-based approach would be used for SaMD. It was not until a 2014 statutorily mandated report,[29] which took the joint efforts of the Federal Trade Commission, Federal Communications

[24] 21st Century Cures Act, § 3060(a), Pub. L. 114-255, 130 Stat. 1033, 1130–32 (2016) (codified at 21 U.S.C. § 360j(o)). While it has been argued before that the FDA lacks jurisdiction, the limitation of the FDA's jurisdiction via the Cures Act removed most doubts that the FDA did in fact have jurisdiction to regulate SaMD software. *See* Laakmann *supra* note 18, at 287; Nathan Cortez, *Digital Health and Regulatory Experimentation at the FDA*, 21 YALE J. L. & TECH. (SPECIAL ISSUE) 4, 25 (2019). This article assumes that the FDA does have the jurisdiction to regulate CDS SaMD.

[25] *See infra* Section II.A.2.

[26] *See infra* Section II.A.3.

[27] *See infra* Section II.B.

[28] Barbara Evans & Pilar Ossorio, *The Challenges of Regulating Clinical Decision Support Software after 21st Century Cures*, 44 AM. J. L. MED. 388, 394-395 (2018); s*ee generally* U.S. FOOD & DRUG ADMIN., CLINICAL DECISION SUPPORT SOFTWARE: DRAFT GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019) [hereinafter 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE].

[29] FDASIA HEALTH IT REPORT: PROPOSED STRATEGY AND RECOMMENDATIONS FOR A RISK-BASED FRAMEWORK 3 (2014) [hereinafter, FDASIA HEALTH IT REPORT].

Commission, and Office of the National Coordinator for Health Information over two years to conduct, that the FDA concluded that it should proceed with a risk-based model,[30] partially because a risk-based approach was Congressionally-endorsed.[31] That same report highlighted the need for the FDA to work with industry regarding regulation of CDS software,[32] likely because of the FDA's lack of expertise[33] and the everchanging nature of the problem.[34]

In the same 25 years from 1989 to 2014, the FDA proposed one guidance document relating to regulation of SaMD,[35] which it then rescinded with no actions on the guidance document in 2005.[36] The FDA held numerous working group sessions with key players in the industry dedicated to trying to design a workable regulatory scheme, but with no actionable results.[37]

In a public working group in 1996, the FDA recognized that "increasing complexity and sophistication of current software devices" hampered its efforts to enforce HCP comprehension "sufficiently to know when significant errors have occurred."[38]

2. The Cures Act

Following the FDA's slow response, manufacturers were concerned about the consequences of a future FDA regulatory scheme and efforts were

[30] Presented to the IMDRF for consideration and then adopted as an international standard before being integrated into the FDA draft guidance document, *Software as a Medical Device (SaMD): Clinical Evaluation*; *see generally* INT'L MED. DEVICE REG. F., *"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations* (2014) [herein 2014 IMDRF REPORT]; U.S. FOOD & DRUG ADMIN., SOFTWARE AS A MEDICAL DEVICE (SAMD): CLINICAL EVALUATION, DRAFT GUIDANCE (2016) [hereinafter 2016 FDA DRAFT GUIDANCE, SOFTWARE AS A MEDICAL DEVICE].

[31] FDASIA HEALTH IT REPORT, at 3.

[32] *Id*. at 26.

[33] *Id*. at 12 and 26-27 (directing that the FDA should provide clarity in different areas of Health IT regulation, including CDS software).

[34] *Id*. at 10.

[35] U.S. FOOD & DRUG ADMIN., FDA POLICY FOR THE REGULATION OF COMPUTER PRODUCTS (DRAFT) (1989) [hereinafter 1989 FDA DRAFT POLICY, COMPUTER PRODUCTS] at 1. This FDA move was guided by issues related to a cardio device, the Therac-25, which resulted in deaths in the US and Canada because the device "was plagued with bugs, a confusing interface, incomplete manuals, and repeated malfunctions." Cortez, *supra* note 9, at 442.

[36] Fed. Reg., Vol. 70, No. 3, at 890 (Jan. 5, 2005).

[37] For a more thorough dive into the history of the FDA's regulatory regime from 1989 to 2014, *see generally* Parasidis, *supra* note 18, at 193–203.

[38] Medical Devices; Medical Software Devices; Notice of Public Workshop, 61 Fed. Reg. 36,886, 36,886 (July 15, 1996).

made both to grant and remove jurisdiction from the FDA.[39] Eventually, Congress succeeded in limiting the FDA's jurisdiction over CDS SaMD in Section 3060 of the Cures Act.[40] Section 3060 of the Cures Act limits FDA jurisdiction by excluding certain items from the definition of a medical device.[41] In relevant part, the Cures Act removes from the definition of a medical device (and from FDA jurisdiction) software whose "function is intended. . . [to support or provide] recommendations to a health care professional about prevention, diagnosis, or treatment of a disease or condition," provided that the HCP can "*independently review* the basis for such recommendations. . . so that it is *not the intent that health care professionals rely primarily on. . . such recommendations* to make a clinical diagnosis or treatment decision regarding an individual patient."[42] The FDA is thus limited from regulating CDS software that recommends treatment based on patient-specific data, general medical knowledge,[43] or where the basis of the recommendation is otherwise 'independently reviewable' by the HCP.

The Cures Act's wording, like many laws from Congress, is intentionally vague to allow the agency flexibility when drafting regulatory guidelines.[44] While Congress has made clear in the Cures Act that they do

[39] Parasidis, *supra* note 18, at 198–99. *See also*, Cortez, *supra* note 9, at 442–43.

[40] *See generally* § 3060(a), 130 Stat. at 1130–32. CDS Software industry players lobbied Congress for this limitation, because of the fear that the FDA intended to regulate all CDS Software, not unwarranted based on previous FDA actions and the Clinical Decision Support Coalition's (a leading industry lobbying group) public support for FDA oversight. *See* Evans & Ossorio, *supra* note 28, at 388; CLINICAL DECISION SUPPORT COALITION, CITIZEN PETITION app. at 12 (2016). Ironically, the relevant language of the Cures Act regarding FDA jurisdiction was spawned in large part by manufacturers' lobbying to remove jurisdiction completely at a time when the FDA was not asserting jurisdiction. *See* Evans & Ossorio, *supra* note 28, at 388; *see also* Sydney Lupkin, *Legislation That Would Shape FDA And NIH Triggers Lobbying Frenzy*, NPR NEWS (November 25, 2016), https://www.npr.org/sections/health-shots/2016/11/25/503176370/legislation-that-would-shape-fda-and-nih-triggers-lobbying-frenzy; Evan Sweeney, *After IBM Intensely Lobbied for AI Deregulation in the 21st Century Cures, the FDA Will Determine its Fate*, FIERCEHEALTHCARE (October 5, 2017), https://www.fiercehealthcare.com/analytics/ibm-watson-fda-21st-century-cures-artificial-intelligence-clinical-decision-support.

[41] § 3060(a)(o)(1), 130 Stat. at 1130. The Cures Act also removed from FDA regulatory jurisdiction four other elements: software used (1) for administrative functions "of a health care facility," (2) "for maintaining or encouraging a healthy lifestyle . . . unrelated to the diagnosis, cure, mitigation, prevention or treatment of a disease or condition," (3) to perform EHR functions, and (4) "for transferring, storing, converting formats, or displaying clinical laboratory tests or other device data and results . . . [not] intended to interpret or analyze clinical laboratory test or other device data, results or findings." § 3060(a)(o)(1)(A)-(D), 130 Stat. at 1130–31.

[42] § 3060(a)(o)(1)(E), 130 Stat. at 1131 (emphasis added).

[43] *Id.*; *see also* Efthimios Parasidis, *supra* note 18, at 200.

[44] Matthew C. Stephenson, *Statutory Interpretation by Agencies*, *in Research Handbook*

not wish to expand FDA jurisdiction in regards to SaMD, it is unclear whether this is because the FDA is not well suited to handle design and enforcement of such a framework[45] or because Congress lacks interest in regulating algorithms (beyond what they already do).

The FDA has released a number of non-finalized guidance documents that attempt to provide clarity regarding the Cures Act, but those guidance documents contain the same vague language as the Cures Act[46] and do not provide a workable standard for CDS software manufacturers. Before discussing the concerns about this vagueness, I will first discuss what the guidance documents say and do when taken together.

3. Where We Are Today: FDA Proposed Guidance

While the Cures Act created new limitations on FDA jurisdiction to regulate medical devices,[47] the Cures Act's main focus was not FDA jurisdiction[48] and FDA jurisdiction is not addressed by any of the legislative history. With the lack of Congressional guidance, the FDA wields exceptional discretion over how to interpret the legislation. Despite the non-binding state of guidance documents generally,[49] review of these guidance documents sheds light on how the FDA will interpret the laws moving forward. Additionally, guidance documents are provided significant deference in courts, making them relatively reliable.[50]

in Public Choice and Public Law 285, 286 (Daniel Farber & Anne Joseph O'Connell eds., Edward Elgar Publishing 2010) (2010) (summarizing the reason that Congress does this in three categories: legislative drafting costs, expertise, and political insulation). *See also* Chevron, U.S.A., Inc. v. NRDC, Inc., 467 U.S. at 865-66; Chad Landmon et al., *Open the Floodgates: The Potential Impact on Litigation Against FDA if the Supreme Court Reverses or Curtails* Chevron *Deference*, 74 FOOD & DRUG L. J. 358 (2019).

[45] *See* Nathan G. Cortez et al., *FDA Regulation of Mobile Health Technologies*, 371 NEW ENG. J. MED. 372, 377 (2014).

[46] See *infra* Section II.A.3.; U.S. FOOD & DRUG ADMIN., CLINICAL AND PATIENT DECISION SUPPORT SOFTWARE: DRAFT GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2017), at 7-8 [hereinafter 2017 FDA DRAFT GUIDANCE CLINICAL DECISION SUPPORT AND PATIENT DECISION SUPPORT SOFTWARE]; 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE, at 8, 12.

[47] § 3060(a)(o)(1), 130 Stat. at 1130–31.

[48] 21st Century Cures Act, Pub. L. 114-255 130 Stat. Ironically, the relevant language of the Cures Act regarding FDA jurisdiction was spawned in large part by manufacturers' lobbying to remove jurisdiction completely at a time when the FDA was not asserting jurisdiction. *See* Evans & Ossorio, *supra* note 28, at 388; Lupkin, *supra* note 40; Sweeney, *supra* note 40.

[49] 21 C.F.R. § 10.115(d) (2019).

[50] While the current Supreme Court composition seriously threatens *Auer* deference to agencies, under the current legal regime these guidance documents do have persuasive force when interpreting statutes in the courts. Auer v. Robbins, 519 U.S. 452 (1997). *See also* Stuart Shapiro, *The Role of Guidance Documents in Agency Regulation*, 36 YALE J. ON REG.:

Following the Cures Act, the FDA began a rapid acceleration of FDA regulatory activity. In all likelihood, this was a regulatory response to the growth of AI[51] and general suspicion regarding the accuracy, precision, and privacy concerns[52] of AI systems. But such regulatory changes were still guided by a lack of expertise within the FDA, resulting in vague guidance documents.[53]

First, in 2016, the FDA formally adopted the International Medical Device Regulators' Forum's ("IMDRF") risk-based framework into FDA policy regarding SaMD, generally.[54] Risk is determined by the "significance of the information provided by the SaMD" to the Health Care Provider (HCP) and the "[s]tate of the healthcare situation or condition."[55] The framework identifies four levels of risk based on these two factors indicating that higher risk requires additional scrutiny and regulatory oversight.[56]

Second, in 2017, the FDA published a draft guidance on CDS software,[57] updating that draft guidance in 2019.[58] This draft guidance

NOTICE & COMMENT (May 9, 2019), https://yalejreg.com/nc/the-role-of-guidance-documents-in-agency-regulation-by-stuart-shapiro/. That deference to agency interpretation can have significant impacts on the preemptive effect of federal legislation on state laws. The current executive administration is also relying more heavily on guidance documents to create new standards regulating algorithms across departments and industries. *See supra* note 4.

[51] 2016 National AI Strategic Plan, *supra* note 4.

[52] *See infra* note 162 and accompanying text.

[53] Medical Devices; Medical Software Devices; Notice of Public Workshop, 61 Fed. Reg. 36,886, 36,886 (July 15, 1996); W. Nicholson Price II, Regulating Black-Box Medicine, 116 MICH. L. REV. 421, 452 (citing Nathan Cortez, The Mobile Health Revolution?, 47 U. C. Davis L. Rev. 1173, 1206 (2014)) (FDA recognizes "it lacks technical expertise on mobile technologies"); supra Section II.A.1.; M. Susan Ridgely & Michael D. Greenberg, Too Many Alerts, Too Much Liability: Sorting Through the Malpractice Implications of Drug-Drug Interaction Clinical Decision Support, 5 ST. LOUIS U. J. HEALTH L. & POL'Y. 257, 284 (2012); *cf.* U.S. Food & Drug Admin., Jobs in the Digital Health Center of Excellence, Fda.gov (Nov. 20, 2020), https://www.fda.gov/medical-devices/digital-health-center-excellence/jobs-digital-health-center-excellence.

[54] U.S. FOOD & DRUG ADMIN., SOFTWARE AS A MEDICAL DEVICE (SAMD): CLINICAL EVALUATION, GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2017) [hereinafter 2017 FDA GUIDANCE, SOFTWARE AS A MEDICAL DEVICE]. The FDA maintains a strong interest in producing an internationally applicable standard for efficiency of approval processes, particularly in the European Union.

[55] 2014 IMDRF REPORT, *supra* note 30, at 10; *see also* 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE, *supra* note 28, at 7.

[56] 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE, *supra* note 28, at 7; 2014 IMDRF REPORT, *supra* note 30, at 10.

[57] 2017 FDA DRAFT GUIDANCE CLINICAL DECISION SUPPORT AND PATIENT DECISION SUPPORT SOFTWARE, *supra* note 46. The draft guidance also delineated CDS and PDS but articulated that the FDA would only regulate PDS to the extent that the agency had the authority to regulate CDS. *Id.* at 7.

[58] *Id.*

discusses each prong of Section 3060's exclusion, providing the FDA's interpretation of the law. While the Cures Act was intended to remove jurisdiction from the FDA, it effectively created an outer boundary that the FDA then began to regulate up to. The FDA announced that they interpreted the requirement that the HCP can "*independently review* the basis for such recommendations"[59] to mean that HCPs primarily rely on "their own judgment, to make clinical decisions for individual patients" rather than the software.[60]

To accomplish this, the FDA interprets the Cures Act to require CDS software manufacturers to "describe their software functions in *plain language*" and include "1) The purpose or intended use of the software function; 2) The intended user (e.g., ultrasound technicians, vascular surgeons); 3) The inputs used to generate the recommendation (e.g., patient age and sex); and 4) The *basis for rendering a recommendation*."[61] The "basis for rendering a recommendation" is further described as "e.g., clinical practice guidelines with the date or version, published literature, or information that has been communicated by the CDS developer to the intended user."[62]

The FDA has yet to update their 2019 draft guidance despite original plans to do so in early 2020 and 2021.[63] This delay is likely due in part to the significant demands of the FDA's role in COVID-19 response. Instead, in October 2020, the FDA established the Digital Health Center of Excellence ("DHCOE") whose role includes eventually refining the FDA's draft guidance, but not until the end of 2021 at the earliest.[64]

B. Wanting for Vagueness

While the draft guidance documents provide some clarity, they leave

[59] § 3060(a)(o)(1)(E), 130 Stat. at 1131 (emphasis added).

[60] 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE, *supra* note 28, at 12.

[61] *Id.*

[62] *Id.*

[63] U.S. FOOD & DRUG ADMIN., *CDRH Proposed Guidances for Fiscal Year 2020 (FY 2020)*, FDA.GOV (Oct. 11, 2019), https://web.archive.org/web/20191213125301/https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/cdrh-proposed-guidances-fiscal-year-2020-fy-2020; U.S. FOOD & DRUG ADMIN., *CDRH Proposed Guidances for Fiscal Year 2021 (FY 2021)*, FDA.GOV (Oct. 16, 2020), https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/cdrh-proposed-guidances-fiscal-year-2021-fy-2021.

[64] U.S. FOOD & DRUG ADMIN., *About the Digital Health Center of Excellence*, FDA.GOV (Sept. 22, 2020), https://www.fda.gov/medical-devices/digital-health-center-excellence/about-digital-health-center-excellence.

far more questions unanswered than they resolve. There is no clarity as to what "independently reviewable" means, what details must be supplied to HCPs,[65] what difference should be applied to purely research or partially research CDS software, what the FDA will do with SaMD that are not going to be targeted for regulation for the time being, or how the FDA will handle partial understandings of a system's process.[66] The majority of these questions stem from what must be required when submitting the basis or bases for rendering a recommendation. Ultimately, the question remains: what does the FDA require of manufacturers in order to be approved or cleared?

Regardless of the FDA's intentions, the vague language results in a spectrum along which the FDA guidance can be interpreted. Each end of the spectrum tugs at a dichotomy between innovation and safety, (over regulating or under regulating, respectively) both resulting in harm to patients. Understanding this spectrum allows us to see the range of harms possible and guides a decision for the proper solution or standard that should apply.

In any event, the vague language leaves manufacturers guessing which interpretation is right. They must determine themselves what they think is most appropriate and submit that to the FDA for clearance or approval, at least until the FDA begins to respond to submissions. This process for manufacturers could result in additional costs that will be passed to consumers (patients) rather than avoided upfront.

Arguably, some level of explainability and interpretability are legally required by the FDA. This article argues that in this way, Congress and the FDA has taken a bold step forward and asserted a right to explanation to HCPs. Further, this article argues that it is really about providing a right to explanation to the patient in order to protect patient autonomy.[67]

1. Complete Transparency: Over Regulation

On one end of the spectrum, the FDA documents can be read to require complete transparency from the manufacturer to the HCPs.[68]

[65] Must HCPs receive all information that can be provided (*i.e.*, a whole study) every time they use the software? Further, by focusing on what HCPs do, is the FDA dangerously close to regulating the practice of medicine? *See generally* Parasidis, *supra* note 18 and accompanying text. Certainly, if machines can outperform humans and the federal government should regulate such machines, the line between device and practice becomes an increasingly narrow one.

[66] For more unanswered questions, *see* Evans & Ossorio, *supra* note 28, at 401–02.

[67] *See infra* Section III.C.

[68] This article will use the following definitions:

"Explainability: the level to which a system can provide clarification for the cause of its decisions/outputs.

"Transparency: the level to which a system provides information about its internal

Complete transparency would require manufacturers to provide information about the internal workings or structure of the software in order to enable HCPs to independently review the CDS software's recommendations.

While intuitively appealing, complete transparency goes too far and can result in high costs to manufacturers, both in development costs and time. Producing a complete description for HCPs will take a tremendous number of resources. This cost borne by manufacturers will be passed to HCPs and then to patients in an already expensive health system.[69] Furthermore, the time to develop explanations will be significant. If the time to develop these additional elements does not divert investors and manufacturers from producing CDS software, it could distract manufacturers from other pursuits, reducing technological innovation or efforts to increase model accuracy in exchange for detailed descriptions of inner workings.[70] Also, developing complete transparency will delay the software's entry into the market and to patients who would benefit.

workings or structure, and the data it has been trained with – this is similar to Lipton's definition of transparency. Lipton (2016).

"Interpretability: the level to which an agent gains, and can make use of, both the information embedded within explanations given by the system and the information provided by the system's transparency level." Richard Tomsett et al., *Interpretable to Whom? A Role-Based Model for*

Analyzing Interpretable Machine Learning Systems, International Conference on Machine Learning Workshop on Human Interpretability in Machine Learning 8, 9 (2018).

[69] *See* Margot Sanger-Katz, *Why Transparency on Medical Prices Could Actually Make Them Go Higher*, N.Y. TIMES (Jun. 24, 2019), https://www.nytimes.com/2019/06/24/upshot/transparency-medical-prices-could-backfire.html (explaining how price transparency could lead to raised prices); *see, e.g.*, Margot Sanger-Katz, *In the U.S., an Angioplasty Costs $32,000. Elsewhere? Maybe $6,400.*, N.Y. TIMES (Dec. 27, 2019), https://www.nytimes.com/2019/12/27/upshot/expensive-health-care-world-comparison.html (explaining the large divergence in prices between the U.S. healthcare system and other countries'); Austin Frakt, *The Huge Waste in the U.S. Health System*, N.Y. TIMES (Oct. 7, 2019), https://www.nytimes.com/2019/10/07/upshot/health-care-waste-study.html (discussing waste in the U.S. healthcare system and its relation to high healthcare costs); *but see, e.g.*, Abby Goodnough & Margot Sanger-Katz, *Health Spending Grew Modestly, New Analysis Finds*, N.Y. TIMES (Dec. 5, 2019), https://www.nytimes.com/2019/12/05/health/health-spending-medical-costs.html (illustrating how, though U.S. health care spending is still high, it grew modestly pre-pandemic).

[70] *See* Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation*, BERKMAN CTR. RSCH. 1, 3 (2017) (describing concerns about increasing transparency could lead to reduced technological innovations); *see also* A. Michael Froomkin et al., *When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning*, 61 ARIZ. L. REV. 33, 99 (2019); Ming Yin et al., *Does Stated Accuracy Affect Trust in Machine Learning Algorithms?*, Int'l Conf. on Machine Learning 1 (2018) (describing some instances of diminished trust in machine learning).

Moreover, too much transparency results in HCPs and patients' diminished trust in the AI systems.[71] This can be highly problematic given that CDS software recommendations have been shown to increase physician performance *and* patient outcomes.[72] As these systems continue to advance, it will become increasingly important for HCPs to trust the internal workings of the algorithms underlying the recommendations from CDS software.

Additionally, depending on the method and content of the information relayed to HCPs, complete transparency can result in HCPs who suffer from alert fatigue.[73] While careful selection of when and how alerts are displayed can reduce this fatigue,[74] a full transparency requirement does not allow for these kinds of systematic adjustments. It is highly unlikely that medical professionals will have the time to review all the clinical studies that form the bases of recommendations without significant diminishment in efficiency resulting in less caseload capacity, especially in an era where there shortages of HCPs are projected in the near future.[75]

[71] *See, e.g.*, Jenny de Fine Licht, *Do We Really Want to Know? The Potentially Negative Effect of Transparency in Decision Making on Perceived Legitimacy*, 34 SCANDINAVIAN POL. STUDS. 183, 197 (2011) (concluding that a "we cannot simply assume that a transparent and objectively legitimate procedure will automatically lead to greater public acceptance and trust . . . on a short term, the effect may even be the opposite, especially if the media present critical reports"); David Goad & Uri Gal, *Understanding the Impact of Transparency on Algorithmic Decision Making Legitimacy*, *in* Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology 64, 77 (Ulrike Schultze et al. eds., 2018) (summarizing the study's conclusions that in some cases, transparency could decrease legitimacy).

[72] Mirela Prgomet et al., *Impact of Commercial Computerized Provider Order Entry (CPOE) and Clinical Decision Support Systems (CDSSs) on Medication Errors, Length of Stay, and Mortality in Intensive Care Units: a Systematic Review and Meta-Analysis*, 24 J. OF THE AM. MED. INFORMATICS ASS'N 413, 420–21 (2016) (describing how the implementation of certain computerized systems can decrease medication prescribing errors and mortality risk in ICUs); Tiffani J. Bright, *Effect of Clinical Decision Support*, 157 ANNALS OF INTERNAL MED. 29 (2012). Contrasting these results with results from just a decade prior in 2005 shows the significant growth of AI since then. Amit X. Garg et al., *Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes*, 293 J. OF THE AM. MED. ASS'N 1223, 1236 (2005) (highlighting that CDSS clinical effectiveness still needs to be tested to a greater extent).

[73] *See* Ridgely & Greenberg, *supra* note 53, at 258 (noting that some physicians turn off alerts due to alert fatigues despite records of them turning off alerts are created).

[74] SEE ALLISON B. MCCOY ET AL., A FRAMEWORK FOR EVALUATING THE APPROPRIATENESS OF CLINICAL DECISION SUPPORT ALERTS AND RESPONSES, 19 J. OF THE AM. MED. INFORMATICS ASS'N 346, 351 (2012) (EXPLORING HOW TO REDUCE ALERT FATIGUE THROUGH FILTERING ALERTS).

[75] Patrick Boyle, *U.S. Physician Shortage Growing*, ASS'N OF AM. MED. COLL. (June 26, 2020), https://www.aamc.org/news-insights/us-physician-shortage-growing (discussing the potential shortfall of up to 139,000 physicians by 2033); *see also* Stuart Heiser, *AAMC Report Reinforces Mounting Physician Shortage*, ASS'N OF AM. MED. COLL. (June 11, 2021),

2. Mere Intent: Under Regulation

On the other end of the spectrum, the FDA documents can be read to require little more than the manufacturer's intent that HCPs be able to independently review the CDS software's recommendation.[76] Under this reading, the FDA would require simply that manufacturers make available collections of data that the manufacturer identified as being the basis of a recommendation.[77] This does little to protect patients from harmful recommendations from AI when more can and should be asked behind why the CDS software made the decision it did.[78] Poor oversight is likely to lead to poor-quality devices and poor-quality devices are likely to result in poor outcomes and injuries for physicians and patients.[79]

HCPs will suffer from automation bias,[80] an overreliance on automated decisions as "a trusted final decision."[81] CDS software is not intended to serve as a final decision, but rather as a system that helps HCPs make decisions.[82] This will harm patients by removing HCPs from the equation when there may be legitimate reasons for HCPs to question the decisions of a CDS software.[83]

Additionally, HCPs will suffer from a shift in burden of costs from manufacturers through tort liability.[84] While it has been recommended that tort may be a solution to the problem, these suggestions fail to recognize the

https://www.aamc.org/news-insights/press-releases/aamc-report-reinforces-mounting-physician-shortage (providing data for potential shortages up to 124,000 physicians).

[76] Evans & Ossorio, *supra* note 28, at 398.

[77] *See* 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE *supra* note 46, at 23 (stating that the FDA would not regulate if the HCP could evaluate a ML software's recommendations "because the logic and inputs for the machine-learning algorithm and data inputs used for the algorithm were explained and *available* to the HCP") (emphasis added).

[78] *See* Doshi-Velez et al., *supra* note 70, at 3-5 (describing that "the utility of explanations must be balanced against the cost of generating them," but that "consequential decisions" often provide an appropriate context for requiring an explanation); *supra* Sections III, IV.

[79] Price, *supra* note 53, at 455.

[80] Cortez, *supra* note 24, at 24 (citing Citron, *infra* note 81, at 1271–72).

[81] Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271–72 (2008).

[82] *See supra* note 13 (defining CDS and its purpose as an aid to HCPs).

[83] Josh F. Peterson et al., *Physician Response to Implementation of Genotype-Tailored Antiplatelet Therapy*, 100 CLINICAL PHARMACOLOGY & THERAPEUTICS 67, 71-72 (2016) (indicating that physicians find many legitimate reasons to reject CDS software recommendations).

[84] *See* Price, *supra* note 53, at 467 (illustrating how tort law focuses and regulates providers); *but see*, Terry, *supra* note 18, at 163 (establishing actions against manufacturers as "canaries in the coalmine" within the auto industry).

lurking complications.[85] The shift in liability in turn will create perverse disincentives to HCPs to not modernize methods with technological innovations.[86] Tort law is entirely reactive, so there is no proactive mitigation of risk to patients.[87] And, tort law standards are created incrementally, but AI capabilities are increasing at a rate that will exceed the ability of the judiciary to match in creating those standards.[88]

C. Right to Explanation

The FDA should provide something between either extreme: not a complete transparency requirement, but also legally requiring some minimum level of explanation and interpretability of the algorithms for approval. This is what the Cures Act and the FDA guidance documents should be read to mean in order to avoid overregulating or under regulating CDS software.[89]

Because of this minimal legal requirement of an explanation, the Cures Act and the subsequent FDA guidance documents instill in some users of the algorithms a right to explanation. The concept of a "right to explanation"[90] was generated as a summation of the effects of Article 22 of

[85] *Compare* Jin Yoshikawa, *Sharing the Costs of Artificial Intelligence: Universal No-Fault Social Insurance for Personal Injuries*, 21 VAND. J. ENT. & TECH. 1155 (2020), *with infra* Section III.A.1.

[86] *See* Froomkin et al., *supra* note 70, at 51 (where once "custom" . . . was the starting point for measuring the appropriate standard of care, U.S. courts today are somewhat suspicious of custom-based arguments on the theory that these arguments provide too little incentive to modernize and may favor entrenched modes of service provision at the expense of the victim"); Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competences, and Strategies*, 29 HARV. J. L. & TECH. 353, 392 (2016) (describing "misaligned incentives" that "do not necessarily comport with the need to optimize public risk").

[87] *See* Scherer, *supra* note 86, at 388 (courts are "well equipped to adjudicate cases arising from specific past harms, but not to make general determinations about the risks and benefits associated with emerging technologies such as AI").

[88] *Cf.* 2016 National AI Strategic Plan, *supra* note 4 (AI technology has "supported rapid progress on tasks once believed to be incapable of automation"); *but see* Scherer, *supra* note 86, at 391 ("On the positive side, the incremental nature of the common law provides a mechanism that allows legal rules to develop organically[.]").

[89] *See supra* Section II.B (currently, manufacturers must guess what the regulations mean and lack clear legal standards).

[90] *Compare, e.g.*, Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 76, 80 (2017) providing "several reasons to doubt the existence, scope, and feasibility of a 'right to explanation' of automated decisions"), *with, e.g.*, Gianclaudio Malgieri & Giovanni Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, 7 INT'L DATA PRIV. L. 243, 246 (2017) ("more than a right to explanation, we claim that the GDPR leads to a 'legibility-by-design'

the 2016 EU General Data Protection Regulation ("GDPR").[91] Article 22 granted a right to system users that they would not "be subject to a decision based solely on automated processing" if the decision "produces . . . significant effects on him or her."[92]

On the one hand, this right to explanation was viewed as a mechanical process for ensuring accountability and transparency by providing clarity on the system's functionality or rationale for specific decisions.[93] On the other, it was viewed as forcing function for legible (i.e., interpretable)[94] systems that guaranteed "the autonomous capability of individuals to understand the functioning and impact of the algorithms concerning them,"[95] simplifying the process and reducing liability to data controllers.[96]

While it is contested whether an actual legal right to explanation can exist under the framework of the GDPR,[97] it is evident that under either interpretation a right to explanation does exist under the Cures Act and FDA proposed guidance documents.[98] The Cures Act's requirement that the HCP be able to independently review the CDS system's recommendation[99] identifies the HCP (the "user")[100] and what that user must be able to do (evaluate the CDS software recommendation). This implies that the user must be able to make sense of the recommendation[101] and assess that recommendation against prevailing standards of care.

Further, the FDA proposed guidance documents also define the HCP as the "user" of the explanation and stipulate that the HCP must be able to evaluate the CDS software recommendation by focusing on the HCP's ability

system: what it guarantees is the autonomous capability of individuals to understand the functioning and the impact of algorithms concerning them"), Bryce Goodman & Seth Flaxman, *European Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, 38 AI MAG. 50 (2017), http://arxiv.org/abs/1606.08813 (GDPR "effectively create[s] a 'right to explanation,' whereby a user can ask for an explanation of an algorithmic decision that was made about them), *and* Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, INT'L DATA PRIV. L. 233, 239 (2017) ("We believe that a plain reading of Articles 13(2)(f), 14(2)(g), 15(1)(h), and 22 supports a right to explanation.").

[91] European Parliament and the Council Regulations 2016/679, 2016 O.J. (L 119) 1, [hereinafter GDPR].

[92] GDPR, Art. 22.

[93] *See* Sandra Wachter et al., *supra* note 90, at 4 and 6.

[94] *See generally* Richard Tomsett et al., *supra* note 68.

[95] Malgieri & Comande, *supra* note 90, at 3.

[96] *Id*.

[97] See *supra* note 90.

[98] See *supra* Section II.A.

[99] See *supra* Section II.A.2.

[100] *See* Tomsett et al., *supra* note 68, at 8.

[101] I.e., requires that the explanation is interpretable or legible. *Compare* Tomsett et al., *supra* note 68, at 8 *with* Malgieri & Comande, *supra* note 90, at 3.

to rely primarily on their own judgment.[102] The proposed regulation also suggests that explanations should be provided in plain language.[103] Taken together, this creates a legal requirement that an interpretable explanation be provided to HCPs.

The user (or, the entity using the explanation) is important to clarify because it guides the level of explainability, interpretability, and transparency prudentially (and potentially legally) required by that individual user.[104] Although this right to explanation belongs to the patient (the "decision-subject"),[105] the Cures Act and FDA proposed guidance documents codify that right through the patient's HCP(s) by placing the HCP as the user.[106] The FDA (as the "examiner") will review manufacturers' submissions to ensure that the HCP is able to understand and evaluate the CDS recommendations.[107] This is consistent with standard healthcare practices[108] and avoids requiring full transparency or crossing the line from device into the practice of medicine.[109]

Alternatively, it could also be argued that requiring the "inputs used to generate the recommendation" implies that some amount of transparency into the inner workings of the algorithm is required.[110] However, this goes too far. Knowing inputs (i.e., factors the algorithm considers) is not the same as knowing the specific relations between all those variables and how the algorithm discerned those relations. HCPs need to have the ability to review what factors were considered in order to ensure that the algorithm considered relevant variables, which in turn informs whether they should reject the recommendation as too narrow.[111] Further, the reason inputs need to be saved is in the event that transparency of an algorithm becomes necessary at a future time.[112]

[102] 2019 FDA DRAFT GUIDANCE, CLINICAL DECISION SUPPORT SOFTWARE, at 12.

[103] *Id.*; *but cf.* Malgieri & Comandé, *supra* note 90, at 245 (describing a standard of legibility based on making data and analytics algorithms both transparent about their commercial use and comprehensible in how they function).

[104] *See* Tomsett et al., *supra* note 68 (envisioning six classes of users based on the role in which they interact with the platform).

[105] Tomsett et al., *supra* note 68, at 12.

[106] It is the HCP who must receive the explanation and be able to understand it.

[107] *See* Tomsett et al., *supra* note 68, at 10, 13.

[108] Gregory E. Pence, Medical Ethics: Accounts of Groundbreaking Cases 16-18 (7th ed. 2015).

[109] *See supra* note 18 and accompanying text.

[110] *See generally* U.S. FOOD & DRUG ADMIN., CLINICAL DECISION SUPPORT SOFTWARE: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2019), at 12.

[111] *See* Peterson et al., *supra* note 83.

[112] *See* Doshi-Velez et al., *supra* note 70, at 9 (identifying litigation as a scenario where the need to review inputs *ex post* may occur).

III. ARE EXPLANATIONS THE RIGHT ANSWER?

Unlike the GDPR, the Cures Act and (when they are finalized) the FDA proposed guidance documents are unquestionably legally binding.[113] Regardless of concerns about the efficacy of agency guidance documents,[114] the Cures Act creates a legal obligation for requiring manufacturers to provide local explanations[115] to HCPs when their CDS software is used; additionally, the FDA will be the entity to implement that legislation.

However, several alternatives to a regulatory fix should be assessed.[116] Regardless of how compelling some of the solutions seem, this article argues that they all result in underregulating or are otherwise not viable at this time.[117]

Additionally, merely requiring an explanation is vague and provides too much variability in an agency lacking expertise.[118] What is prudentially required then is a narrower standard, a specific type of explanation that manufacturers should provide or work directly with the FDA on when an explanation cannot be provided.[119]

This article will then argue that there are practical and ethical guidelines that any proposed standard must meet. These are not discussed in great depth, but provide enough to evaluate a workable standard.

A. Alternative Proposed Solutions

Alternative solutions have been proposed, but fall short for various reasons. These shortfalls collapse into three broad categories of concern. The first is under regulation, such as tort law or FDA post market surveillance.[120] The second is nonviability, such as educating HCPs, hiring more experienced regulators at the FDA, or creating a new agency mirroring the FDA to regulate AI algorithms.[121] The third is undue risk, such as completely

[113] *See generally* Sandra Wachter et al., *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J. L. AND TECH. 841, 861-862 (2018).

[114] *See supra* note 44; *see also* Rudin, *supra* note 15.

[115] *See* Doshi-Velez et al., *supra* note 70, at 7 (defining local explanations as an explanation for a "specific decision, rather than an explanation of the system's behavior overall"). Also referred to as "specific decisions." Sandra Wachter et al., *supra* note 113, at 6. Local explanations are required rather than total transparency because the expectation is that the HCP is the one who must assess the explanation. *See infra* Section III.C.

[116] See *infra* Section III.A.

[117] *Id.*

[118] See *infra* Section III.B.

[119] *Id.*

[120] *See infra* Section III.A.1.

[121] *See infra* Sections III.A.2, III.A.3.

overhauling the FDA approach.[122] This is not meant to be a comprehensive review of all literature, but to provide a quick understanding of the hurdles these suggestions are not able to surpass.

1. Tort Law

Tort law results in under regulation of the field in several ways, including that it is entirely reactive.[123] This results in patient harm that is addressed ex-post through litigation to establish new standards. Litigation takes time and results in delayed standards. Meanwhile AI will continue to advance, ever increasingly.[124] Even with changing standards, not all HCPs will conform, which means patients will still be at significant risk of harm. Further, these standards are not fueled by expertise in the field, but rather by judge and jury at whatever level of expertise they have as laymen, which introduces chances of less-than-ideal standards.[125]

One glaring example of why tort is insufficient is the original reason for the FDA's first guidance document in 1989.[126] The FDA's initiation of regulation was due to tort law's inability to adequately mitigate medical devices' risks lacking FDA procedures.[127]

Another issue of relying on tort law is inconsistencies across the national landscape. The manufacturers that produce the algorithms in a tort system would need to conform to each states' standards. Federal preemption is desirable to avoid local and ex post liability.[128] Going further, there is a need for an international standard, or at least interoperability of developed technologies, in order to encourage economic competition and American leadership in the area of AI altogether, which cannot be achieved through tort law.[129]

2. HCP Education and Licensure

Another solution would be to have state licensing boards establish

[122] *See infra* Section III.A.4.

[123] *See supra* Section II.B.2.

[124] *See generally* 2016 National AI Strategic Plan, *supra* note 4; Scherer, *supra* note 86, at n.143 (citing Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* 63–66 (2014)); Laakmann, *supra* note 18, at 309; *but see* Scherer, note 86, at 391.

[125] *See* Price, *supra* note 53, at 462; Scherer, *supra* note 86, at 390.

[126] 1989 FDA DRAFT POLICY, COMPUTER PRODUCTS, *supra* note 35.

[127] Cortez, *supra* note 9.

[128] Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 91 (2017).

[129] *See generally* 2016 National AI Strategic Plan, *supra* note 4; Keith Bradsher & Katrin Bennhold, *World Leaders at Davos Call for Global Rules on Tech*, N.Y. TIMES (Jan. 23, 2019), https://www.nytimes.com/2019/01/23/technology/world-economic-forum-data-controls.html.

standards of use, ensuring a level of competence for use of AI medical devices.[130] In this way, the heads of the profession would self-regulate. In fact, hospitals are already creating user and developer guidelines.[131] However, sustaining this practice as the norm would cause the shift in tort liability discussed above and is unviable for numerous other reasons.

Presumably this would lead to state boards regulating AI use, which would result in the same problem of varying standards as in tort liability,[132] but with the added undue influence of certain professional bodies, such as the AMA.[133] It would be exorbitantly expensive to train physicians and surgeons. Even if financially feasible, physicians and surgeons are typically advanced in age[134] and the vast majority have little professional training in the area of computer science, generally, much less algorithms.[135] Relying on doctors would also increase the cost of healthcare by reducing the time physicians spend with their patients and increasing the time spent examining code.[136] Regardless, too much industry self-regulation can result in harm nonetheless, hence the original impetus for federal regulation.[137]

3. FDA Hiring

Another solution would be for the FDA to increase hiring of computer scientists with AI/ML expertise. This would increase the FDA's ability to conduct thorough PMA and post market monitoring while managing a variety of different approaches from different manufacturers. However, this is infeasible at this time because of the shortage of skill in the area of AI.[138] The FDA would be unlikely to capture an effective portion of the market to

[130] *See generally* Terry, *supra* note 18, at 153; *see also* Claudia E. Haupt, *Governing A.I.'s Professional Advice*, 64 MCGILL L. J. 665, 680 (2019).

[131] *See* Cortez, *supra* note 24, at 13.

[132] *See generally* Terry, *supra* note 18, at 153.

[133] *See, e.g.*, *id.* at 155–56. And an inappropriate influence as the AMA has biased interests when it comes to treatment of AI in research settings.

[134] The average age of all physicians and surgeons in the United States of America is 46.8 years in 2017. *Physicians and Surgeons: Demographics*, DATAUSA (2017), https://datausa.io/profile/soc/physicians-surgeons#demographics.

[135] In the simplest terms, around 7.63% of currently practicing physicians and surgeons have a bachelor in computer science in 2017. *Physicians and Surgeons: Education*, DATAUSA (2017), https://datausa.io/profile/soc/physicians-surgeons#education.

[136] *See, e.g.*, Doshi-Velez et al., *supra* note 70, at 3; *see generally supra* note 69 and accompanying text.

[137] *See* Cortez, *supra* note 24, at 22; *see also* Price, *supra* note 53, at 455.

[138] *See, e.g.*, Takla S. Perry, *Intel Execs Address the AI Talent Shortage, AI Education and the "Cool" Factor*, IEEE SPECTRUM (Sept. 11, 2018), https://spectrum.ieee.org/intel-execs-address-the-ai-talent-shortage-ai-education-and-the-cool-factor. It should be noted not to confuse the growth of education in computer science with capabilities in AI as the two are mutually exclusive.

adequately review such a wide range of submissions. The DHCOE efforts to capture some expertise is still ideal in working with the different stakeholders, other agencies, and manufactures on individual bases when they cannot provide explanations consistent with this paper's recommendation, but this is not a one-size-fits-all solution.

4. Regulatory Reform

Because of the recognized lack of expertise in AI as it develops and improves, the FDA is on the precipice of reforming their regulatory model into a more holistic model that relies on industry norms.[139] In a 2019 discussion paper, the FDA described the model as one that fosters a culture of quality by enforcing Good Machine Learning Practices (GMLP).[140] The FDA is currently in the process of trying to define GMLP with industry input.[141] Once a manufacture satisfies the requirements for a culture of GMLP and are rated,[142] they will receive no review or streamlined clearance procedures similar to the 510(k) approval pathway.[143]

The 510(k) pathway is considered a "comparative" process where the

[139] The FDA was not designed for the "adaptive AI/ML technologies." U.S. FOOD & DRUG ADMIN., PROPOSED REGULATORY FRAMEWORK FOR MODIFICATIONS TO ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE DISCUSSION PAPER AND REQUEST FOR FEEDBACK 3 (2019) [hereinafter 2019 FDA DISCUSSION PAPER, AI/ML-BASED SAMD]; *see generally* U.S. FOOD & DRUG ADMIN., DEVELOPING A SOFTWARE PRECERTIFICATION PROGRAM V 1.0 (2019) [hereinafter 2019 FDA PRECERTIFICATION PROGRAM V1.0].

[140] 2019 FDA DISCUSSION PAPER, AI/ML-BASED SAMD, *supra* note 139, at 9–10.

[141] *See* U.S. FOOD & DRUG ADMIN., SOFTWARE PRECERTIFICATION PROGRAM: 2019 TEST PLAN (2019) [hereinafter 2019 FDA PRECERTIFICATION PROGRAM TEST PLAN].

[142] Rated in a way that seems eerily like the FDA is making a determination of the company's ability to produce an effective and safe sound product rather than making a determination as to whether the product is in fact effective and safe. This creates risks that investors will use these ratings inappropriately, that the FDA will begin to regulate outside their intended realm, and that well-meaning companies' medical devices could be approved despite safety concerns that could have been discovered efficiently. *See* U.S. FOOD & DRUG ADMIN., *FDA's Role in Regulating Medical Devices*, *supra* note 17 and accompanying text.

[143] *See* 2019 FDA PRECERTIFICATION PROGRAM V1.0, *supra* note 139, at 29; U.S. FOOD & DRUG ADMIN., THE 510(K) PROGRAM: EVALUATING SUBSTANTIAL EQUIVALENCE IN PREMARKET NOTIFICATIONS, GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF 2 (2014) [hereinafter 2014 FDA GUIDANCE, 510(k) NOTIFICATION]. The 510(k) substantial equivalent notification pathway considers similarities demonstrated by a manufacturer of a product's substantial equivalence to a device that has previously been approved (the "predicate"). *See, e.g.*, Parasidis, *supra* note 18, at 192; Diana M. Zuckerman et al., *Medical Device Recalls and the FDA Approval Process*, 171 ARCHIVES INTERNAL MED. 1006, 1007 (2011) (asserting that, "[t]he 510(k) process was specifically intended for devices with less need for scientific scrutiny, such as surgical gloves and hearing aids.").

PMA requires "an independent demonstration of safety and effectiveness."[144] The 510(k) pathway has been heavily scrutinized because of the loophole it creates whereby devices are cleared by the FDA despite the fact that the predicates are no longer considered safe or are off the market.[145] These criticisms could apply equally to a new, untested process by the FDA to expedite review of new SaMD. GMLP will work as a "comparative" process creating loopholes for devices because of the 'culture' of an institution.

Additionally, those devices which are approved through the PMA process receive federal preemption from state tort liability.[146] However, those devices that are approved through the 510(k) process are not preempted.[147] If the FDA approves devices through this holistic model rather than the PMA, manufacturers will not have the freedom to innovate because of susceptibility to state tort law with a wide range of different standards again resulting in harm to patients.[148]

5. FDA for Algorithms

Some, such as Andrew Tutt, have argued that a new agency empowered with the ability to review algorithms and set generally applicable standards for all government agencies would provide consistency and governmental cost-efficiency.[149] While this suggestion is a highly desirable end-state that would resolve many of the expertise issues that the FDA currently faces, the shift could be costly and excessively time consuming to develop when regulation is desperately needed (including the tasks of garnering legislative action, executive buy-in, and establishing and integrating the agency itself). Additionally, such an agency may be overwhelmed by requests for review and standards development, which would prevent this new agency from effectively adapting to the quickly changing international regulatory and technological landscape.

Absent these challenges, the agency would still need to assess the needs of different industries to develop appropriate standards, resulting in the status quo. There would still be no existing standards, applicable legislation would still govern, and there would still be an expertise gap in the new agency when assessing AI in healthcare. While an FDA for algorithms may be an

[144] 2014 FDA GUIDANCE, 510(k) NOTIFICATION, *supra* note 143, at 6.

[145] *See, e.g.*, Cortez et al., *FDA Regulation of Mobile Health Technologies*, 371 NEW ENG. J. MED. 372, 375 (2014); Last Week Tonight with John Oliver, *Medical Devices: Last Week Tonight with John Oliver*, YOUTUBE (June 3, 2019), https://youtu.be/-tIdzNlExrw.

[146] Riegel v. Medtronic, Inc., 552 U.S. 312, 317–18 (2008) (relying on 21 U.S.C. § 360k(a)).

[147] Medtronic v. Lohr, 518 U.S. 470, 492 (1996).

[148] *See supra* Section III.A.1.

[149] Tutt, *supra* note 128, at 113.

aspirational goal, it is not currently a solution.

B. Merely 'Explanation' Isn't Enough

A right to explanation is viewed as a broad legal duty. The basic tenets of the duty are elusive with no clear obligations in itself.[150] Because this legal obligation is so vague, manufacturers cannot rely on the FDA proposed guidance documents. Current FDA guidance documents provide some explanations.[151] However, they do so to varying degrees of depth and usefulness to the HCPs as users of the explanations.[152] As a result, any of a wide variety of explanations could be required.

The result of this variability in options available to manufacturers is the same under regulation and over regulation problem discussed above.[153] If manufacturers attempt to reveal the entire inner workings of their algorithms, innovation will suffer, as will patients.[154] If manufacturers produce only the simplest explanations, there will likely be poor quality that could result in substantial harm to patients.[155]

To correct the FDA's current direction, they need to provide a narrower standard that manufacturers can use as a baseline. This will provide left and right limits to the healthcare industry so as to prevent over regulation or under regulation and protect patient safety. However, as stated by Doshi-Velez and colleagues, any system must be a fluid one because of the growing nature of the AI field.[156] Technological changes, methodological changes, and redistribution of resources all have the power to upset the status quo, and the FDA must be flexible enough to adapt to these changes. However, these are not reasons to refrain from creating a workable industry standard as a baseline.

C. What Standards Should Apply?

There are several practical concerns that play into the ultimate

[150] *See generally* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18, 35 (2017).

[151] *See generally* CHRISTOPH MOLNAR, INTERPRETABLE MACHINE LEARNING: A GUIDE FOR MAKING BLACK BOX MODELS EXPLAINABLE (2019).

[152] *Id.*

[153] *See supra* Section II.B.

[154] *See supra* Section II.B.1.

[155] *See supra* Section II.B.2.

[156] *See* Doshi-Velez et al., *supra* note 70, at 11.

viability of the standard chosen, such as cost-effectiveness,[157] efficacy,[158] flexibility,[159] etc.[160] Further, any standard that should be pursued will need to "be based on generally accepted ethical and moral frameworks."[161] Several readily apparent standards should apply such as data privacy[162] and avoid racial discrimination.[163]

While Professor Terry does not synthesize sources, he does identify transparency, avoidance of bias, equity, cost-effectiveness, and data protection as key generally applicable ethical and moral obligations.[164] He articulates that this is not necessarily a complete list, but serves as a starting point.[165]

The Cures Act recognizes another general ethical obligation: protecting patient autonomy.[166] An HCP acting under the principle of autonomy must respect the competent patient's right to self-determination regarding their treatment, and therefore has a duty to provide the patient with

[157] *E.g.*, does the standard applied cost too much to maintain? *See, e.g.*, *supra* Section II.B.

[158] *E.g.*, does the standard applied actually do its job?

[159] Preferably model-agnostic so as to allow the standard to stand both varying models today and new models in the future. *See* Molnar, *supra* note 151, at 8.2.

[160] *See, e.g.*, Terry, *supra* note 18, at 165 (discussing "William Kissick's healthcare iron triangle (access, quality, and cost containment)"); William Kissick, Medicine's Dilemmas: Infinite Needs Versus Finite Resources (1994).

[161] Terry, *supra* note 18, at 164.

[162] *See, e.g.*, Roger Allen Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOM. & TECH. L. REV. 1 (2016). The subject of data privacy will not be addressed in great detail, but is an important element of any standards, especially as private companies begin amassing more health data. Natasha Singer & Daisuke Wakabayashi, *Google to Store and Analyze Millions of Health Records*, N.Y. TIMES (Nov. 11, 2019), https://www.nytimes.com/2019/11/11/business/google-ascension-health-data.html; Natasha Singer, *New Data Rules Could Empower Patients but Undermine Their Privacy*, N. Y. TIMES (Mar. 9, 2020), https://www.nytimes.com/2020/03/09/technology/medical-app-patients-data-privacy.html; *see also*, *infra* Section IV.A.

[163] *See, e.g.*, Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023 (2017). Particularly when dealing with things like 'race,' which is a social construct, but also informative to HCPs regarding vulnerabilities a person may have relating to their ancestry. Take for example the increased risk of sickle cell disease in populations around malaria because of malaria's lessened effectiveness in those with sickle cell disease. *Cf.* Michael Aidoo et al., *Protective Effects of the Sickle Cell Gene Against Malaria Morbidity and Mortality*, 359 THE LANCET 1311 (2002).

[164] Terry, *supra* note 18, at 168.

[165] For a comprehensive review of medical publications, *see* Jessica Morley et al, *The Ethics of AI in Health Care: A Mapping Review*, SOC. SCI. MED. (2021).

[166] While there is no singular code of conduct among medical professionals, there are four bioethical principles that are highly regarded: beneficence, nonmaleficence, justice, and autonomy. Pence, *supra* note 108, at 16.

the level of knowledge necessary to make a rational, informed decision.[167] The Cures Act codifies this obligation by ensuring the HCP is informed enough to offer adequate advice and information to the patient in making a decision.[168]

In order to be able to provide the patient with that level of knowledge, an HCP need not understand the internal workings of the algorithm, but must be able to evaluate the decision of the algorithm.[169] Further, there must be some value gained that is worth the cost of producing the evaluation.[170] Evaluation can be summarized in three distinct purposes of the explanation: "(1) to inform and help the [HCP] understand why a particular decision was reached, (2) to provide grounds to contest...decisions, and (3) to understand what could be changed to receive a desired result in the future."[171]

The level of explainability, transparency, and interpretability required by the Cures Act and FDA guidance documents is defined then by the ability of HCPs to understand, use (including to contest CDS software recommendations), and alter future decisions. Since the explanations being used will be for the HCP to understand, contest, and alter local decisions, there is a need for high explainability and interpretability, but not transparency.

IV. COUNTERFACTUAL EXPLANATIONS

In the midst of the debates about the 'right to explanation' in the GDPR, Sandra Wachter and colleagues designed a method for constructing an explanation, called a counterfactual, from black-boxes without opening the black-box, a euphemism for not producing full transparency.[172] Their method avoids the costly and inefficient tactics that involve producing full transparency.[173] The result is if-then style statements that do not need significant background knowledge and expertise to be understood.[174]

By changing the original variable (x_i) to the closest possible synthetic data point (x'), the user can see if or what impact it has on the

[167] *Id.* at 16–17, 33-34, 289–90.

[168] *See supra* Section II.A.2.

[169] *See* Richard Tomsett et al., *supra* note 68, at 12; *see also*, Wachter et al., *supra* note 113, at 843.

[170] *See* Doshi-Velez et al., *supra* note 70, at 11.

[171] *See* Wachter et al., *supra* note 113, at 843.

[172] *See generally* Wachter et al., *supra* note 113, at 854–59. This method has been incorporated into machine learning textbooks. *See* Molnar, *supra* note 151.

[173] *See* Wachter et al., *supra* note 113, at 854, 860 (referring to the computations as "easy"). Indeed, Wachter's formulas and examples span a mere seven pages. *Id.* at 854–60.

[174] *See id.* at 871 (explaining that counterfactuals provide information in an easily understandable form and do not require knowledge about the algorithm to understand).

recommendation.[175] For example, if the original value represents age, that number is changed by one year, and the result is a different recommendation, then the user will know that age is highly consequential to the algorithm's recommendation.

Knowing what inputs influence the algorithm model(s), to what extent, and in what direction(s) are more important than fully comprehending the underlying calculus. Knowing the inputs and relative influence on the outcome provide what HCPs need to know regarding CDS software predictions and recommendations, such as efficacy of the model and potential sources of bias. When dealing with a biased algorithm, this process can reveal those biases by demonstrating that a change in race results in different algorithmic recommendations.[176] This allows algorithm users to identify these biases and alter future decisions based on identified biases. Bias can exist and yet exist in proxy with other variables.

Using the smallest possible change is critical though because there may be significant volatility across a dataset and the smallest possible change reduces the impact of that volatility and produces the least artificial, most realistic explanation.[177]

This article argues that there are significant advantages[178] and limited concerns[179] associated with establishing a baseline standard of requiring manufacturers to provide counterfactual explanations to HCPs using CDS software. The advantages align with what would be desired from an FDA standard and provide additional returns.[180]

A. Advantages of Counterfactual Explanations

The first key advantage is the greatly reduced regulatory and financial burden of producing full transparency. Full transparency comes at a significant cost[181] that will be passed to consumers (patients) in an already expensive healthcare system.[182] Once the FDA establishes a standard, it is likely that an industry will begin to standardize production and integrate

[175] *See* Wachter et al., *supra* note 113, at 855 (describing how a synthetic point can vary a feature but remain close to the original variable).

[176] *Id.* at 856–58.

[177] *Id.* at 855.

[178] *See infra* Section IV.A (discussing advantages of counterfactuals).

[179] *See infra* Section IV.B (addressing concerns of counterfactuals).

[180] *Compare supra* Section III.C. (reviewing what standards should apply and what HCPs need to understand), *with infra* Section IV.A (explaining how counterfactuals reduce cost and increase HCP effectiveness).

[181] *See supra* Section II.B.1 (explaining how complete transparency results in higher costs for manufacturers and patients).

[182] *See supra* note 69 (discussing the high cost of healthcare and how complete transparency could increase these costs).

explanations into software.[183] As explanation systems[184] become more standardized, cost of production will decrease, reducing costs to patients and increasing equitable access to CDS software.

A counterfactual requirement by the FDA would also be highly adaptable to changing technologies because the method of constructing the counterfactual is not tied to a certain kind of algorithmic model, be it machine learning or otherwise. Counterfactual explanations are thus model-agnostic[185] explanation systems and applicable to even cutting-edge technologies,[186] despite the pace of growth.[187] This flexibility is of the nature that the FDA needs to apply to allow manufacturers to have the standard explanation system. Further, this flexible standard would keep manufacturers able to adapt their explanation systems to comply with international regulatory norms, once developed.[188]

FDA endorsement of this standard to satisfy "independently reviewable" would produce highly human-comprehensible,[189] local explanations.[190] Human comprehensibility is essential for understanding CDS recommendations because of the level of expertise HCPs have in the

[183] *See* Doshi-Velez et al., *supra* note 70, at 7 (arguing that by distinguishing the AI explanation system from the AI system, regulators will create the environment for an industry built around the explanation system).

[184] *See id.* at 8 (explaining the role and importance of explanation systems); *see also* Matt Turek, *Explainable Artificial Intelligence (XAI)*, DARPA, Figure 2, https://www.darpa.mil/program/explainable-artificial-intelligence (depicting how explainable AI systems work).

[185] *See* Molnar, *supra* note 151 (defining and explaining model-agnostic methods); Wachter et al., *supra* note 113, at 852 (explaining the techniques of generating counterfactual explanations).

[186] *See* Wachter et al., *supra* note 113, at 852 (explaining how counterfactuals benefit cutting-edge architectures).

[187] *See* 2016 National AI Strategic Plan, *infra* note 124 (explaining the pace of technological advancements); Scherer, *supra* note 86, at n.143 (citing Nick Bostrom, Superintelligence: Paths, Dangers, Strategies 63-66 (2014)); Laakmann, *supra* note 18, at 309 (explaining the FDA's response to technological advances related to human-source material); *but see*, Scherer, *supra* note 86, at 391 (explaining how the legal system is slow to respond to technological advancements).

[188] United States leadership in this area is critical to ensuring adequate safety and effectiveness globally through standardization. *See supra* note 129 and accompanying text (reporting on world leaders' calls for global technology regulation).

[189] *See* Wachter et al., *supra* note 113, at 861 (explaining how counterfactuals provide comprehensible information). Being easily human-interpretable without specialized skills in computer science is vitally important for explanations as there is an extreme scarcity of trained personnel in the field. *See, e.g.*, Perry, *supra* note 138 (describing the shortage of AI engineers); *supra* Section III.A.2 (explaining the cost and challenges of training physicians and surgeons).

[190] *See generally* Doshi-Velez et al., *supra* note 70, at 7.

area of computer science.[191]

Further, without understanding the provided recommendation, HCPs will generally be unable to enable the patient to make informed, autonomous decisions. HCPs will be unable to contest recommendations with valid reasons to reject recommendations.[192] Under a normal standard of care, HCPs do not describe minute details leading to their recommendations to each and every patient: doctors do not have the time to do this, and many patients would likely not understand it if they did. The same standard should apply with CDS software recommendations. They must explain in adequate detail the basis of the decision, but do not need to be so transparent of the inner workings of the algorithm that they become inefficient, costly, confusing, or untrustworthy.[193]

The counterfactual explanations will increase HCP and patient trust in the systems[194] while giving them enough information to make informed, autonomous decisions. Too much transparency would hinder trust, whereas no explanation prevents autonomous decision-making.[195] While some concern exists over automation bias and deskilling of physicians over time, the current uses of CDS software have shown improvements to physician performance and patient outcomes.[196] Because of the potential growth of concern in this area, monitoring should continue with increased incorporation.

Lastly, the counterfactual model preserves intellectual property rights because understanding the counterfactual does not require access to the underlying data.[197] This important element of counterfactual explanations increases value to manufacturers and assures them that by providing an explanation, they are not waiving their intellectual property rights. By reducing risk to manufacturers, there is increased potential to incentivize high-value growth and innovation, benefiting patients long-term. This also increases privacy of patients' information used to create the model.[198]

[191] *See supra* note 135 and accompanying text. It is not cost-effective to train HCPs as an alternative approach either.

[192] *See, e.g.*, Peterson et al., *supra* note 83.

[193] *See supra* Section III.B.

[194] In fact, physician *overreliance* is a more prevalent worry, at first because they know that their decisions and inputs affect the models, automation bias sets in, and then causes deskilling of physicians over time. *See, e.g.*, Froomkin et al., *supra* note 70, at 99; Ming Yin et al., *Does Stated Accuracy Affect Trust in Machine Learning Algorithms?*, International Conference on Machine Learning, (2018); Cortez, *infra* note 24, at 24 (citing Citron, *infra* note 81, at 1271-72).

[195] *See supra* note 71 and accompanying text.

[196] *Supra* note 72 and accompanying text.

[197] Molnar, *supra* note 151.

[198] Manufacturers do not need to worry about providing transparency in order to meet

B. Addressing Concerns

Counterfactuals have a number of distinct disadvantages. One in particular is that there is no guarantee that a counterfactual will be able to be produced for the algorithm.[199] Of course, the opposite is usually true. There are usually *multiple* counterfactual explanations for each instance, known as the Rashomon Effect.[200] From a list of counterfactuals, HCPs will have the option to choose which ones they find most applicable based on their experiences and knowledge.[201] This is a reason that HCPs remain a part of the decision-making process, to evaluate CDS recommendations and understand why certain factors are affecting the ultimate decision.

Besides HCPs being selective with purchase and use of a manufacturer's algorithm on the basis of data governance and integrity,[202] HCPs may have an additional role as gatekeepers on the basis of the user interface. An effective interface at the healthcare facility will be crucial to minimize alert fatigue.[203] And the FDA will not micromanage the manufacturers in this regard, because such FDA activity treads too heavily into regulating the practice of medicine.[204]

While costs can be significantly cheaper than producing transparency to the internal workings of the algorithm, the reality is that producing the counterfactuals will still cost organizations money. This can be mitigated through FDA flexibility with manufacturers, but ultimately the FDA should only approve those algorithms that can honor the principle of autonomy, giving patients and HCPs meaningful choice. While there may be some upfront costs, over time, an industry built around producing explanations may come into existence as regulation and research require explanations for safe and effective deployment of algorithms.[205]

The FDA would need to require that not only would the training data have to be submitted with the algorithm, but so would descriptions on how the explanations were produced so that they could be validated. Without these

Health Insurance Portability and Accountability Act ("HIPAA") Privacy Rules as some have argued, because the Privacy Rules provide an exception for physician access to treat patients. Evans & Ossorio, *supra* note 28, at 399. There are privacy concerns for PDS that may warrant additional regulation, but that is outside the scope of this article. *See, e.g.*, Bradsher & Bennhold, *supra* note 129.

[199] Molnar, *supra* note 151.

[200] *Id.*

[201] *See id.*

[202] Dina B. Ross & Campbell Tucker, *Artificial Intelligence and Healthcare Regulation*, The Law of Artificial Intelligence and Smart Machines, 59-60 (Theodore F. Claypoole ed., 2019).

[203] *See, e.g.*, McCoy et al., *supra* note 74.

[204] *See supra* Section II.

[205] Doshi-Velez et al., *supra* note 70, at 7.

two steps, the FDA would be unable to thoroughly review the explanations from manufacturers.

For these reasons, the FDA will have to stay flexible with different forms of explanation presented by manufacturers, but there needs to be a starting point that is less vague than the current guidance documents suggest. Other options that can produce explanations exist and many are quite similar to counterfactuals. These alternative methods can be substituted in place of counterfactuals by the FDA on a case-by-case basis when the manufacturer gives a compelling reason why they should have their algorithm assessed by a different standard.

Some of the alternative methods of producing explanations are common algorithmic interrogation methods among researchers. This is significant as research institutes and commercial businesses[206] that publish their research will be able to offer alternative, already produced work when algorithms are submitted for approval. For example, Permutation Feature Importance (PFI) offers one advanced means of identifying *why* an algorithm made a particular decision. PFI works by randomizing one variable at a time. By shuffling an *important* variable, the model will have a much higher error rate. By measuring the difference in error rates based on which variable is shuffled, a picture of the model emerges.[207] Thus, other models exist, but may not work well in every scenario or be easily human-interpretable, such as counterfactuals. These alternatives should not be excluded by the FDA when regulating, but the FDA should disfavor them in favor of explanations that are more interpretable to a wider base of potential users.

V. CONCLUSION

The FDA's current regulatory scheme is based on more than three decades of back-and-forth work in the area of SaMD regulation with little expertise. As such the FDA's guidance documents are wanting for clarity in several fashions, most striking of which is the ambiguous language of "independently reviewable." The work "independently reviewable" does to the legal obligations of CDS software manufacturers is minimal and results in a potential over- or under-regulation problem. By ensuring independent

[206] In 2014, Microsoft adopted an open access policy for publications that allowed employees to publish research. *E.g.*, Jim Pinkelman and Alex Wade, *Microsoft Research Adopts Open Access Policy For Publications*, Microsoft Research Blog (January 20, 2014). In 2016, Apple, Inc. followed Microsoft and announced that they would break with their current operating procedures and begin publishing AI research. *E.g.*, Mike Wuerthele, *Apple AI Researchers Gagged No More, Now Allowed to Publish and Confer with Colleagues*, appleinsider (December 6, 2016).

[207] Molnar, *supra* note 151; *e.g.*, Andre Altmann et al., *Permutation Importance: A Corrected Feature Importance Measure*, 26 BIOINFORMATICS 1340 (2010).

reviewability, the FDA is trying to ensure safe and effective SaMDs. However, the vague language hampers understanding of what manufacturers of SaMD need to produce for approval. The FDA thus should produce a workable standard as a baseline for manufacturers to achieve. That standard need not apply in every situation and the FDA needs to be flexible with industry when that baseline standard is not workable in their case, but at the very least both parties should have a shared starting point. I propose one solution, counterfactual explanations, as a promising option for the FDA to explore as it will satisfy the needs of the FDA, is simple and malleable, and meets the baseline theoretical needs of the HCPs and patients who will use the CDS software to make decisions. There are a few potential disadvantages of this as a baseline, but so long as the FDA treats the standard as a *baseline*, then most of those disadvantages are negligible.