Houston Journal of Health Law & Policy

INTRODUCING A FAIRNESS CHECKPOINT FOR DATA QUALITY AND EVIDENCE DURING REGULATORY REVIEW OF AI/ML-ENABLED MEDICAL DEVICES

Mahsa Shabani, PhD, LLM* and Vasiliki Rahimzadeh, PhD**

* Associate Professor in Health Privacy Law and Innovation at the University of Amsterdam.

** Assistant Professor in the Center for Medical Ethics and Health Policy at Baylor College of Medicine.

ABSTRACT

Artificial Intelligence (AI) in healthcare and medical research is here. Large, well-characterized and representative datasets are the foundations of safe and effective AI systems, including AI-enabled medical devices. Fairness in the ways that clinical data are collected, analyzed and shared to train AI models used. In medical devices is consequential for the safety and efficacy of those devices. Regulators, however, do not explicitly consider issues of fairness in evaluating the rigor of clinical evidence used to substantiate device safety and efficacy as part of the regulatory approval process for medical devices. Other ethics and compliance oversight bodies—including institutional review boards (IRB) and data access committees (DAC)—work upstream to ensure ethical data collection and use practices during device development and validation. However, IRB and DAC reviews are rarely, if ever, made available to regulators during pre-market approval. In this paper, we argue for why regulatory approval bodies should be concerned with fairness at the level of training data supporting AI-enabled devices, and how they could integrate fairness assessment into their regulatory decisions. We discuss the opportunities and operational barriers of three possible models for a fairness "checkpoint" in the regulatory approval process for AI-enabled medical devices. These models build on the extant literature in fair AI and which regulatory bodies could feasibly integrate into existing device application review and approval processes.

I. INTRODUCTION

In this paper, we consider whether regulatory approval bodies should be concerned with issues of fairness in evaluating the quality of evidence that underpins new AI/ML-enabled medical devices. And if so, why and how should fairness be considered to help regulatory agencies fulfill their primary duty to ensure device safety and efficacy?

The use of AI in healthcare is already a reality, and with it promises to dramatically improve healthcare outcomes, transform clinical practice, and improve the overall care experience for patients. According to a 2021 World Health Organization report, AI is expected to "augment the ability of health-care providers to improve patient care, provide accurate diagnoses, optimize treatment plans, support pandemic preparedness and response, inform the decisions of health policy-makers, and help allocate resources within health systems."[1] During the COVID-19 pandemic, tools developed using Machine Learning (ML) methods, a subset of AI that uses algorithms trained on large datasets to automate certain tasks, were used to assist healthcare providers with diagnoses and predict criticality, mortality, and hospitals stays of COVID-19 patients.[2] Most recently, AI researchers have highlighted the capabilities of generative AI, AI models capable of generating new media by interpreting human provided prompts, in the medical context. For instance, Large Language Models (LLM), generative AI that analyzes and synthesizes output based on natural language, like OpenAI's ChatGPT and Google's Bard could be used to draft progress plans for patients or directly answer patients' questions via automated chat response, profoundly influencing medical practice in areas that were once the sole preserve of human care providers.[3]

[1] WORLD HEALTH ORG., ETHICS AND GOVERNANCE OF ARTIFICIAL INTEL. FOR HEALTH 11 (June 28, 2021), https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1.

[2] *See* Nikolas Blomberg & Katharina B. Lauer, *Connecting data, tools and people across Europe: ELIXIR's response to the COVID-19 pandemic,* 28 EUR. J. HUM. GENET. 719, 719-20 (2020).

[3] Timo Minssen et al., *The Challenges for Regulating Medical Use of ChatGPT and Other Large Language Models*, 330 JAMA 315, 315–16 (2023); Jesutofunmi A. Omiye et al., *Large language models propagate race-based medicine*, 6(195) NPJ DIGIT MED. 1, 1 (2023).

It is expected that use of AI/ML-enabled medical devices will have a significant impact on disease diagnosis, patient monitoring, medical robotics, and genome and image analysis.[4] Many of these devices are already subject to specific rules under medical devices regulations,[5] which require their validation through clinical trials or other approaches to generating clinical evidence of the disease indications and in specific patient populations.[6]

Nonetheless, AI/ML-enabled medical devices are granted regulatory pre-market approval in an environment already beset by concerns related to data privacy, transparency and fairness that affect the equitable access to AI/ML-enabled services and device products worldwide.[7] These disparities manifest in diverse ways and are shaped by an array of vulnerabilities linked to specific demographic categories, encompassing variables such as age, gender, sexual orientation, disability, ethnicity/race, socio-economic status, migration background, and geographical location. For example, the existing lack of racial/ethnicity representativeness in genomic datasets can perpetuate biases when used to train genomic data interpretation algorithms.[8] Similarly, algorithms designed to allocate healthcare resources may discriminate against migrants, LQBTQ+, or people belonging to lower socio-economic groups, because they unintentionally rely on improper proxies for health needs and fail to account for their lived experiences

[4] PwC, Eur. Comm'n, Studies on eHealth, Interoperability of Health Data and Artificial Intel. for Health and Care in the EU 86–88 (2019); Kevin B. Johnson et al., *Precision Medicine, AI, and the Future of Personalized Health Care*, 14 Clinical and Translational Sci. 86, 86–88 (2021).

[5] Minseen, *supra* note 3, at 315.

[6] *See* PwC, *supra* note 4, at 45.

[7] Ritika Manik and Gelareh Sadigh, *Diversity and Inclusion in Radiology: A Necessity for Improving the Field*, 94 Brit. J. Radiology 1126, 1126 (2021); Dena R. Matalon et al., *Clinical, Technical, and Environmental Biases Influencing Equitable Access to Clinical Genetics/Genomics Testing: A Points to Consider Statement of the American College of Medical Genetics and Genomics (ACMG)*, 25 Genetics in Med. 1, 6 (2023); David Leslie et al., *Does "AI" stand for augmenting inequality in the era of covid-19 healthcare?*, 372 Brit. Med. J. 1, 1–2 (2021); Annabel Kupke et al., *Pulse Oximeters and Violation of Federal Antidiscrimination Law*, 329 JAMA 365, 365–66 (2023).

[8] Kevin B. Johnson et al., *Precision Medicine, AI, and the Future of Personalized Health Care*, 14 Clinical Translational Sci. 86, 90 (2021); Raquel Dias & Ali Torkamani, *Artificial intelligence in clinical and genomic diagnostics*, 11 Genome Med. 1, 9 (2019); Jose Florez et al., *Addressing Diversity and Inclusion in Human Genetics Research* 175 Cell Press, 303, 303–05 (2018).

(algorithmic bias).[9] In terms of clinical implementation, inequitable opportunities to share the benefits of AI/ML-enabled medical devices due to a lack of digitally mature hospitals (privilege bias) or population level mistrust in using such tools, may also exacerbate unfairness in the use AI/ML-enabled medical devices.[10]

To address existing health inequities as well as prevent new disparities from emerging, we argue that fairness must assume a central role throughout the entire lifecycle of AI/ML-enabled medical devices—from design, testing, and development to clinical implementation. However, a closer examination reveals that fairness remains a fragmented theoretical concept, often disconnected from its varied interpretations and ambiguously applied to help guide responsible technology innovation and use.[11] Historically, fairness has been intertwined with concepts of justice.[12] Theories have been proposed that seek to define fairness in relation to the processes (procedural fairness) and/or outcomes (substantive fairness).[13] Theories of fairness also aim to strike a balance between individual rights and the collective welfare.[14] There is limited consensus among various philosophical traditions as to what "fairness" normatively entails in the development and implementation of emerging technologies, such as health AI.[15] Similarly, there is no uniformly applied definition of fairness in the context of AI-enabled medical device development, validation or implementation.[16] As van Nood and Yeomans eloquently claim, "Philosophical

[9] *See* Eduard Fosch-Villaronga et al., *Accounting for diversity in AI for medicine*, 47 COMPUT. L. & SEC. REV. 1, 2 (2022); Justyna Stypinska, *AI Ageism: A Critical Roadmap for Studying Age Discrimination and Exclusion in Digitalized Societies*, 38 AI & SOC'Y 665, 665–67 (2023); Renate Baumgartner et al., *Fair and Equitable AI in Biomedical Research and Healthcare: Social Science Perspectives*, 144 A.I. MED. 1, 2 (2023).

[10] Auna Lorena Ruano et al., *Understanding inequities in health and health systems in Latin America and the Caribbean: a thematic series*, 20 INT'L. J. EQUITY HEALTH 1, 1–3 (2021).

[11] *See* John Rawls, *Justice as Fairness*, 67 PHIL. REV., 164, 164 (1958); *see also* Alan Ryan, *Fairness and Philosophy*, 73 SOC. RSCH. 597, 597 (2006).

[12] Rawls, *supra* note 11.

[13] *See* Rawls, *supra* note 11.

[14] *See* Rawls, *supra* note 11, at 165–67.

[15] Alan Ryan, *Fairness and Philosophy*, 73 SOC. RSCH. 597, 597 (2006).

[16] *See* Haytham Siala & Yichuan Wang, *SHIFTing Artificial Intelligence to be Responsible in Healthcare: A Systematic Review*, 296 SOC. SCI. & MED. 1, 2 (2022) ("… there is no universally accepted ethical framework….").

interest in fairness in the context of AI is, in part, a response to the systemic limitations of algorithmic tools, limitations which do not necessarily constrain everyday exercises of fairness and which therefore present novel challenges to ethical design."[17]

Different experts, organizations, and regulators have proposed diverse metrics, methodologies, and ethical frameworks to assess and ensure fairness in AI/ML-enabled tools.[18] From a technical perspective, fairness has been often perceived as debiasing AI by addressing statistical disparities.[19] The technical interventions therefore emphasize demographic parity, striving for equal outcomes across different patient groups.[20] In contrast, fairness in healthcare delivery compels greater accounting for genuine biological and/or socio-economic differences that differentially impact optimal levels of individual and population health (i.e. structural and social determinants of health).[21] In other words, from a clinical perspective, fairness has most often been construed as equity in the delivery of care at both the individual and group level.[22]

When it comes to regulatory oversight for AI/ML-enabled medical devices, fairness is ambiguous. The existing legal scholarship on algorithmic fairness unspecific to healthcare, has predominantly focused on invoking non-discrimination laws to prevent deployment of

[17] Ryan V. Nood & Christopher Yeomans, *Fairness as Equal Concession: Critical Remarks on Fair AI*, 27 SCI. & ENG'G ETHICS 73, 76 (2021).

[18] Suvodeep Majumder et al, *Fair Enough: Searching for Sufficient Measures of Fairness*, 32 ACM TRANSACTIONS SOFTWARE ENG'G & METHODOLOGY 1, 2 (2022) ("recent research has proposed a plethora of new fairness metrics.…").

[19] *See* Amarachi Mbakwe et al., *Fairness Metrics for Health AI: We Have a Long Way to Go*, 90 EBIOMEDICINE 1, 1 (2023); SAHIL VERMA & JULIA RUBIN, FAIRNESS DEFINITIONS EXPLAINED 3 (2018); Ninareh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning*, 54 ACM COMPUTING SURV. 1, 1 (2021); Alessa Angerschmid et al., *Fairness and Explanation in AI-Informed Decision Making*, 4 MACH. LEARNING & KNOWLEDGE EXTRACTION 556, 557 (2022).

[20] Mehrabi et al., *supra* note 19.

[21] *See* Daiju Ueda et al., *Fairness of Artificial Intelligence in Healthcare: Review and Recommendations*, 42 JAPANESE J. OF RADIOLOGY 3, 6 (2024) ("By incorporating data from various patient populations, age groups, disease stages, cultural and socioeconomic backgrounds, and health-care settings, AI can learn to recognize, diagnose, and treat a broad spectrum of patient conditions with greater precision and contextual understanding.").

[22] Mingxuan Liu et al., *Towards Clinical AI fairness: A Translational Perspective*, 6 NPJ DIGIT. MED. 1, 3 (2023).

unfair AI tools.[23] Yet the standards used to detect and prove direct or indirect discrimination differ among many national, subnational, and supranational jurisdictions. For example, non-discrimination laws in the EU prohibit the treatment of people or groups differently based on sensitive characteristics, including race, gender, sexual preference, political or religious convictions, etc.[24] Despite these protections, critics have pointed out that proving the necessary elements of agency, intentionality, harm exceedingly difficult, and moreso demonstrating systemic injustice against protected sub-groups, thereby rendering many regulatory protections vis-à-vis nondiscrimination ineffective.[25]

In other cases, fairness safeguards have been employed in legal systems to target the negative ramifications of developing and deploying unfair AI tools in terms of data protection, accuracy, safety, and transparency.[26] While some of the regulatory frameworks are not specific to AI but medical devices generally, government agencies in the United States[27] and Europe[28] have joined a chorus of academic

[23] *See, e.g.*, Sandra Wachter et al., *Why fairness cannot be automated: Bridging the gap between the EU non-discrimination law and AI*, 41 COMPUT. L. SEC. REV. 1, 1 (2021) (analyzing EU non-discrimination law and algorithmic and automated fairness).

[24] *See generally id.* (discussing EU non-discrimination law).

[25] *See id.* at 37.

[26] Laura Sikstrom et al., *Conceptualizing Fairness: Three Pillars for Medical Algorithms and Health Equity*, 29 BRIT. MED. J. HEALTH CARE INFORM 1, 1 (2021); Carmel Shachar & Sara Gerke, *Prevention of Bias and Discrimination in Clinical Practice Algorithms*, 329 JAMA 283, 284 (2023); Marvin van Bekkum & Frederik Borgesius, *Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?*, 48 COMPUT. L. SEC. REV. 1, 1 (2023); Marieke Bak et al., *You Can't Have AI Both Ways: Balancing Health Data Privacy and Access Fairly*, 13 FRONTIER GENETICS 1, 1 (2022).

[27] Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED'L TRADE COMM. (Apr. 8, 2020), https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms; U.S. DEP'T OF HEALTH & HUM. SERV., *Considerations for IRB Review of Research Involving Artificial Intelligence* (July 21, 2022), https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-e-july-25-2022-letter/index.html.

[28] *Artificial Intelligence and Algorithmic Fairness Initiative*, U.S. EQUAL EMP. OPPORTUNITY COMM'N, https://www.eeoc.gov/ai (last visited Jan. 7, 2024); U.S. DEP'T OF COMM., NAT'L INST. OF STANDARDS AND TECH., U.S. LEADERSHIP IN AI: A PLAN FOR FEDERAL ENGAGEMENT IN DEVELOPING TECHNICAL STANDARDS AND RELATED TOOLS 3 , (2019), https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf; *EU AI Act: First Regulation on Artificial Intelligence*, EUR. PARL. (updated Dec. 19, 2023), https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

scholars,[29] international bodies,[30] nonprofit organizations,[31] and advocates, among others, in proposing frameworks for safe and trustworthy AI and directing resources to support responsible governance of AI systems.

In Europe, the European Commission has proposed the legislative draft for AI Act in April 2021, to regulate AI systems on various sectors, including healthcare.[32] On July 12, 2024, the European Union's Artificial Intelligence Act, Regulation (EU) 2024/1689 ("EU AI Act") was published in the EU Official Journal, making it the first comprehensive horizontal legal framework for the regulation of AI systems across the EU.[33] The EU AI Act enters into force across all 27 EU Member States on August 1, 2024, and the enforcement of the majority of its provisions will commence on August 2, 2026.[34] This regulation aims to ensure that fundamental rights, democracy, the rule of law and environmental sustainability are protected from high-risk AI, while boosting innovation and making Europe a leader in the field.[35] The rules establish obligations for AI based on its potential risks and level of impact.[36]

Similarly, in the US, President Biden signed an executive order on the *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* with important implications for healthcare payers, patients, and providers in October 2023[37] The executive order will soon require federal

[29] Siala & Wang, *supra* note 16; Floridi Luciano, *Establishing the Rules for Building Trustworthy AI*, 1 NATURE MACH. INTEL. 261, 261 (2023).

[30] WORLD HEALTH ORG., *supra* note 1.

[31] Lara Groves, *Algorithmic Impact Assessment: A Case Study in Healthcare*, ADA LOVELACE INSTITUTE (Feb. 8, 2022), https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/; Emanuel Moss et al., *Assembling Accountability: Algorithmic Impact Assessment for the Public Interest*, DATA & SOCIETY, https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability.pdf (last visited Mar. 6, 2024).

[32] *EU AI Act: First Regulation on Artificial Intelligence*, *supra* note 28.

[33] *AI Act,* EUR. COMM'N (Oct. 14, 2024), https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai .

[34] *Id.*

[35] *Id.*

[36] *Id.*

[37] Michelle M. Mello et al., *President Biden's Executive Order on Artificial Intelligence—Implications for Health Care Organizations*, 331 JAMA 17, 17 (2024).

agencies to develop standards for safe and trustworthy AI systems across sectors, as well as to evaluate and monitor AI products based on these standards over time.[38] The executive order specifically directs the U.S. Department of Health and Human Services (DHHS) to "develop a strategic plan that includes policies and frameworks—possibly including regulatory action, as appropriate—on responsible deployment and use of AI and AI-enabled technologies in the health and human services sector (including research and discovery, drug and device safety, healthcare delivery and financing, and public health)."[39]

In section G, Executive Order 14110 specifically calls on the Secretary of the DHHS, in collaboration with other federal agencies, to provide resources for the "development of AI assurance policy—to evaluate important aspects of the performance of AI-enabled healthcare tools—and infrastructure needs for enabling pre-market assessment and post-market oversight of AI-enabled healthcare-technology algorithmic system performance against real-world data."[40]

Our proposal that regulators should implement a fairness checkpoint in their review for AI/ML-enabled devices is advanced by the executive order's explicit focus on building infrastructures and evaluation programs for real world performance at the pre- and post-market phases of the AI innovation life cycle.[41]

It further aligns with the Coalition for Health AI[42] (CHAI)'s next steps for institutionalizing trustworthy AI evaluation and monitoring by naming federal regulators as having specific oversight responsibilities in this regard:

> To ensure that the AI tools used by health systems possess these elements, an opportunity exists to specify who tests and when they test. Therefore, in addition to assurance standards, there may be a need for adjudicating bodies, and such tests may represent something that is certifiable, thus promoting confidence in such tools. The result is ongoing

[38] *Id.*

[39] Exec. Order No. 14110, 88 Fed. Reg. 75191 at 75214.

[40] *Id.* at 75215.

[41] *Id.* at 75215.

[42] CHAI is a volunteer group of subject matter experts in medical AI from diverse institutions representing healthcare systems, academia, government, and industry. More information on their activities, consensus reports and recommendations are available online. *See* COALITION FOR HEALTH AI, https://www.coalitionforhealthai.org/.

> monitoring to ensure continued trustworthy AI, facilitated by testing, evaluation, and/or assurance bodies.[43]

In this paper we discuss the relevance of assessing fairness in the review of clinical evidence used to substantiate safety and efficacy of novel AI/ML-enabled medical tools for regulatory approval. In Section II, we first explain what elements are relevant for fairness across the AI innovation continuum (e.g. from development, to validation, to implementation and evaluation/monitoring). We then comment in Section III on the need for, and specific responsibilities to, assess fairness at the level of data among major regulatory bodies in the US and Europe, namely FDA and EU's medical devices regulation. Other oversight bodies have complementary roles and responsibilities with respect to ensuring data fairness. We elaborate on how specific duties of data access committees (DAC) and institutional review boards (IRB) relate to ensuring fairness upstream of regulatory review and approval. In Section IV, we introduce the concept of a fairness "checkpoint" in the pre-market approval process for new AI/ML-enabled medical devices. We conclude in Section V with describing the opportunities and challenges of three potential checkpoint models for assessing fairness in clinical evidence used to substantiate AI-enabled device safety and efficacy at the pre-market authorization stage of regulatory approval.

II. FAIR MEDICAL AI ACROSS THE DEVELOPMENT CONTINUUM

Our ideas of fairness necessarily emerge out of concept of justice. That is, what we judge to be fair—fair treatment, or fair allocation of resources for instance—can be a helpful proxy for our values and perspectives about justice. Justice has deep roots in traditions of moral and political philosophy and ethics and is a central tenet of biomedical ethics. The Romans proposed a legal definition for justice in the *Institutes of Justinian*: "the constant and perpetual will to render to each his due."[44] Legal scholars have since distilled four core concepts from this

[43] COAL. HEALTH ARTIFICIAL INTELLIGENCE, BLUEPRINT FOR TRUSTWORTHY AI IMPLEMENTATION GUIDANCE AND ASSURANCE FOR HEALTHCARE 19 (2022), https://www.coalitionforhealthai.org/papers/Blueprint%20for%20Trustworthy%20AI.pdf.

[44] David Miller, *Justice*, 2023 STAN. ENCYCLOPEDIA PHIL. 1, 2 (2023), https://plato.stanford.edu/archives/fall2023/entries/justice/.

Roman definition.[45] Namely, justice compels attention to "how individual people are treated ('to *each* his due') . . . is a matter of claims that can be rightfully made against the agent dispensing justice, whether a person or an institution . . . as general rules impartially applied over time . . . [and] requires an agent whose will alters the circumstances of its objects."[46] Justice, or more appropriately distributive justice, is one anchoring principle in research ethics that promotes just allocation of research benefits and burdens; no one group should bear the sole burdens of research to benefit a few, and vice versa.[47] Distributive justice is also consequential for advancing social justice (of AI),[48] which broadly aims to promote access to goods and services among individuals and groups irrespective of social constructions like race or gender.[49] Systems that allow individuals, groups, or communities to be disproportionately harmed—just as those that allow only a few to singularly benefit—might be considered socially unjust.[50]

Fairness as applied to the development and deployment of AI frustrates each of the core justice pillars. AI models and systems themselves do not "treat" people. Rather, AI systems operate entirely with data; that is, digital representations of specific characteristics or attributes of actual people.[51] The common rights claims and interests owed to people by AI systems can be ambiguously defined and even more difficult to evidence under current AI law and regulation.[52] AI models

[45] *Id.* at 1.

[46] *Id.* at 3–4.

[47] *Id.* at 12.

[48] Matthias Kuppler et al., *From Fair Predictions to Just Decisions? Conceptualizing Algorithmic Fairness and Distributive Justice in the Context of Data-Driven Decision-Making*, 7 FRONTIERS SOCIO. 1, 1 (2022); Iason Gabriel, *Toward a Theory of Justice for Artificial Intelligence*, 151 DAEDALUS 218, 221 (2022).

[49] Miller, *supra* note 44 at 19.

[50] Miller, *supra* note 44 at 7.

[51] Kuppler, *supra* note 48, at 2.

[52] These ambiguities are fueled in part by the lack of a human agent that can be proven to act on those rights positively or negatively. Scholars in the medical liability space have begun to reconcile tort legal theory and practice with AI accountability. However, the central question of agency that underlies tort continues to plague the courts in medical AI liability cases. Agency is particularly elusive for generative AI in which humans are involved only in the earliest stages of development but that can learn unsupervised over time, or for artificial general intelligence that operate completely independent of humans.

are also not static. AI systems by default are iterative, dynamic, and responsive to new data inputs. Therefore, expecting that rules meant to govern AI development and deployment be static both limits AI's capabilities and could underestimate new or dynamic instances of bias as technology advances.

Ethical frameworks for development and deployment of AI/ML enabled tools abound.[53] Floridi and Cowls, for example, meta-theorizes across leading international frameworks and distil five overarching principles of ethical AI that build on foundational principles in bioethics: autonomy, beneficence, justice, non-maleficence and explicability.[54] Jobin and colleagues propose a slightly modified list of principles that include transparency, justice and fairness, non-maleficence, responsibility and privacy.[55] While frameworks coalesce around a common set of principles, scholars disagree on how actors should translate principle into practice. Other organizations have therefore focused their efforts on responsible AI development, validation and implementation.[56] The National Institute for Standards and Technology (NIST) framework for AI risk management, for example, is organized around four key functions (map, measure, manage, and govern).[57]

The authors of these frameworks make evident that principles such as justice, equity and fairness are consequential also for safety and

[53] Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 NATURE MACH. INTEL. 501, 501 (2019) (Middlaestadt claims at least 84 organizations have released statements "describing high-level ethical principles, tenets, values, or other abstract requirements for AI development and deployment." He goes on to argue that consensus principles for ethical AI is insufficient on its own to motivate ethical AI praxis because AI development lacks common aims and fiduciary duties, a professional history and set of ethical norms, proven methods to translate principles into practice, and robust legal and professional accountability mechanisms).

[54] Luciano Floridi & Josh Cowls, *A Unified Framework of Five Principles for AI in Society*, 1.1 HARV. DATA SCI. R. 1, 5–17.

[55] Anna Jobin et al., *The Global Landscape of AI Ethics Guidelines*, 1 NATURE MACH. INTEL. 389, 389 (2019).

[56] *See e.g.*, COAL. HEALTH ARTIFICIAL INTEL., *supra* note 43, at 3.

[57] *Id.* (The Coalition for Healthcare AI (CHAI) summarizes the NIST's functional approach to managing AI risks as follows: "MAP establishes the context for framing risks related to an AI system. *Measure* employs quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts. *Manage* function entails allocating risk resources to mapped and measured risks on a regular basis and as defined by *Govern*, which is a cross-cutting function infused throughout AI risk management that enables the other functions of the process").

efficacy of AI systems, and that they manifest differently across the AI development pipeline. Our own field of genetics/genomics is a case in point. In the section that follows, we use genomics as a case exemplar to illustrate how issues of algorithmic fairness impact safety and efficacy of AI-enabled medical devices in ways that should fall under the purview of federal regulators when considering device approvals.

A. Model Training

We believe fairness in AI development begins with a socially just purpose for applying the technology to address an existing problem.[58] There should be intent to develop an AI model or system that does not intentionally aim to discriminate against an individual or group.[59] We advance the notion that fairness in AI development is contingent on several factors.[60] AI models and systems are not only as good as the quality of the data upon which they are trained[61] , but also on the methods for procuring this data to begin with.

Delimiting the lawful bases for controlling and processing personal data, as well as clarifying what is or is not considered personal data are among the core protections under the GDPR.[62] The means through which personal data are collected matters little for determining the type of protections that are required. All personal data in the EU, whether collected in the hospital, at a grocery store or elsewhere, are subject to protections under the GDPR.[63] This is not so in the United States.[64]

[58] *Id.*

[59] Kadija Ferryman & Mikaela Pitcan, *Fairness in Precision Medicine*, DATA SOC'Y (Feb. 2018), https://datasociety.net/wp-content/uploads/2018/02/DataSociety_Fairness_In_Precision_Medicine_Feb2018.pdf

[60] Michael Madaio et al., *Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support,* 6 PROCEEDINGS ACM HUMAN-COMPUTER INTERACTION 1, 1–26.

[61] *Id.* at 7.

[62] Masha Shabani, *Collection and sharing of genomic and health data for research purposes: Going beyond data collection in traditional research settings*, N.1S BIOLAW J. 251, 252 (2021).

[63] *Id.* at 252.

[64] *See* Thorin Klosowski, *The State of Consumer Data Privacy Laws in the US (And Why It Matters),* N.Y. TIMES (Sept. 6, 2021), https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/ ("[T]here's no single, comprehensive federal law regulating how most companies collect, store, or share customer data.").

In the United States, different data types are regulated differently, and different U.S. agencies are responsible for data privacy regulation. There are three broad data types and three primary regulatory agencies responsible for their protection.[65] Health data that is considered protected health information is subject to protections outlined under HIPAA.[66] HIPAA regulates the sharing and disclosure of protected health information from covered entities–consisting of providers, payers, healthcare clearinghouses, and business associates.[67]

Yet health-related data can—and frequently are—collected, processed, stored and shared by non-covered entities.[68] In the United States, de-identified data[69] from electronic health records are one of the principal sources of training data for AI models.[70] The barriers to access de-identified data for AI training purposes are relatively low for bona fide researchers who are subject to data protection requirements under human research regulations and even lower for industry developers who work at companies that are business associates of

[65] These three data types include health data, finance data, and education data. The Health Insurance Portability and Accountability Act outlines federal data privacy rules that are enforced by the Department of Health and Human Services. Financial data are protected by the Federal Trade Commission, while education data is subject to federal regulations outlined in the Federal Education Rights and Privacy Act (FERPA) under the U.S. Department of Education's jurisdiction. *See* Health Insurance Portability & Accountability Act of 1996 (HIPAA), 42 U.S.C. §§ 1320d-d-8; *see* Gramm-Leach Bliley Act of 1999 (GLBA), 15 U.S.C. § 6821; *see* Federal Education Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g

[66] *Summary of the HIPAA Security Rule*, U.S. DEP'T OF HEALTH AND HUM. SERVS., https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.

[67] *Id.*

[68] *See AHIMA Policy Statement on Health Information held by HIPAA non-covered entities*, AM. HEALTH INFO. MGMT. ASS'N, https://www.ahima.org/media/jial0h2q/hipaa-nce-policy-statement-final.pdf.

[69] The Health Insurance Portability and Accountability Act (HIPAA) governs the use, sharing and disclosure of protected health information in the United States. Data that has been stripped of 18 unique identifiers, or adequately de-identified as per expert determination are exempt from HIPAA. *See* 45 C.F.R. § 164.514(b) (2024).

[70] *See e.g.,* David Raths, *Truveta Trains Large-Language Model on EHR Data*, HEALTHCARE INNOVATION (Apr. 12, 2023), https://www.hcinnovationgroup.com/analytics-ai/artifical-intelligence-machine-learning/news/53057113/truveta-trains-large-language-model-on-ehr-data.

hospitals.[71] Importantly, most patients whose de-identified data are used for research, including to train AI models, are unaware of these uses.[72]

There is expansive literature on the importance of representation and diversity in datasets used to train AI/ML enabled tools.[73] We argue this issue of representation is closely associated with fairness in that it matters who bears the burdens and benefits of data contribution.

B. Development

The concept of algorithmic fairness is perhaps most relevant during the validation stages, where the effects of input biases can be exposed.[74] Underrepresentation of patients and groups in datasets used to train AI models poses an early threat to the internal validity, and therefore efficacy, of AI models.[75] There are other ways that such threats can also arise in the validation phase of AI development. In one study of FDA approvals, for instance, investigators found that retrospective or historical data were more often provided to substantiate efficacy.[76] That is, the evidence applicants provided was about how well an AI model likely *would have* performed by using data from historical cases instead of how well they performed when tested with *actual* patients.[77]

Much of the extant computer science scholarship frames algorithmic fairness as a remedial issue that can be adequately addressed with

[71] *See* Kayte Spector-Bagdady, *Governing Secondary Research Use of Health Data and Specimens: The Inequitable Distribution of Regulatory Burden between Federally Funded and Industry Research*, 8 J. L. BIOSCIENCES 1, 14 (2021) (Spector-Bagdady previously exposed this regulatory double standard for secondary use of health data for research and commercial purposes).

[72] *Id.* at 27–28.

[73] *See* Richard J. Chen et al., *Algorithmic Fairness in Artificial Intelligence for Med. and Healthcare*, 7 NAT. BIOMEDICAL ENG'G 719, 719–20 (2023).

[74] *See* Mark MacCarthy, *Fairness in Algorithmic Decision-Making*, BROOKINGS (Dec. 6, 2019), https://www.brookings.edu/articles/fairness-in-algorithmic-decision-making/.

[75] *Id.*

[76] Eric Wu et al., *How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals*, 27 NATURE MED. 582, 582 (2021).

[77] *See id.* at 582–83.

technical solutions.[78] The argument is that biases in AI algorithms could theoretically be contained if the right technical corrections are applied such as refining multimodal data inputs, expanding limit thresholds, or alpha testing using more realistic real-world examples.[79] We share Wong's critique of techno-dominant "fixes" for algorithmic fairness: "Since decisions on fairness measure and the related techniques for algorithms essentially involve choices between competing values, 'fairness' in algorithmic fairness should be conceptualized first and foremost as a political issue and to be [resolved politically]."[80]

The context within which validation occurs also matters for evidentiary quality and rigor. Should the FDA have the authority to approve a new device if applicants successfully demonstrate safety and efficacy but only tested the device at well-resourced academic medical centers? An analogy could be drawn here about approval for therapies for rare genetic disease. Often, such therapies require expensive equipment or diagnostic sequencing capabilities that are unlikely to be available everywhere.[81] While the FDA does not have the authority to withhold approval for an otherwise safe and effective drug because of cost or site availability of necessary equipment, these practical considerations can still shape coverage decisions and wider uptake by healthcare systems.[82]

Insofar as the device could have similar applicability in low or under-resourced clinical settings, we argue that a single validation setting is insufficient to meet evidentiary standards for efficacy. On account of both efficacy and fairness, every AI-enabled device should include performance and other validation testing across differently resourced settings.

[78] Brianna Richardson & Juan E. Gilbert, *A Framework for Fairness: A Systematic Review of Existing Fair AI Solutions*, 1 J. ARTIFICIAL INTELLIGENCE RSCH. 1, 13 (2021).

[79] *See id.*

[80] Pak-Hang Wong, *Democratizing Algorithmic Fairness*, 33 PHIL. TECH. 225, 225–26 (2019).

[81] Kate Antrobus, *How We Can Make Gene Therapies Available to All*, WORLD ECON. F. (Oct. 21, 2021), https://www.Weforum.Org/agenda/2021/10/how-we-can-make-gene-therapies-available-to-all/.

[82] *Frequently Asked Questions about CDER*, U.S. FOOD & DRUG ADMIN. (Oct. 28, 2019), https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/frequently-asked-questions-about-cder#:~:text=We%20understand%20that%20high%20drug,by%20manufacturers%2C%20distributors%20and%20retailers.

C. Implementation and monitoring

Fair AI deployment necessitates attention to fair allocation of the benefits and burdens that result from AI decision supports or devices. Fairness is thus a driving force also for continuous monitoring of AI-enabled devices which have already been deployed.[83] Periodic monitoring and auditing are essential to ensure individuals and groups are fairly treated when AI tools are used to inform decisions about how to distribute clinical resources, goods, or services.[84] Continuous evaluation and auditing of algorithmic systems enables developers to identify and correct for biases that can occur after initial validation and which can measurably affect device performance over time.[85] Healthcare demographics as well as social determinants of health also evolve. Ongoing monitoring helps developers take better account for shifts in population characteristics and adapt AI models in line with these changes to ensure they remain effective.[86]

III. ESTABLISHING SAFETY AND EFFICACY OF MEDICAL AI

The existing and emerging regulatory frameworks for approving AI/ML-enabled medical devices emphasize safety and efficacy of the devices.[87] This goal can be achieved through different requirements set by such regulatory frameworks, including data representativeness to address data bias.[88] In response, approval of new devices might be dependent on collecting new data through clinical trials to show the safety and efficacy for the intended target populations.

Notably, this requirement may not apply to all new devices. In fact, for many devices, manufacturers can rely on existing clinical data

[83] MacCarthy, *supra* note 74.

[84] *See* MacCarthy, *supra* note 74.

[85] *See* MacCarthy, *supra* note 74.

[86] *See generally* Jasmine Chiat Ling Ong et al., *Artifical Intelligence, ChatGPT, and Other Large Language Models for Social Determinants of Health: Current State and Future Directions*, 5 CELL REP. MED. 1, 6 (2024)

[87] *How FDA Regulates Artificial Intelligence in Medical Products*, PEW TRUSTS (Aug. 5, 2021), https://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2021/08/how-fda-regulates-artificial-intelligence-in-medical-products.

[88] Shea Brown et al., *Bias Mitigation in Data Sets* (July 8, 2021), https://osf.io/preprints/socarxiv/z8qrb.

to demonstrate that devices meet safety and efficacy standards.[89] For example, in the EU, the EU Medical Devices Regulation requires only devices that fall under the higher risk category level (Class IIa or higher) to provide such evidence.[90] In the US, the most common pathway for device approval is the 510(k) premarket submission process, where device manufacturers can claim a device is "substantially equivalent" to a previously approved device, thus indirectly relying on previously submitted data to satisfy regulatory standards.[91] Some AI/ML-enabled software may fall outside this regulatory scope.[92] Clinical Decision Support (CDS) tools, for example, provide recommendations to healthcare providers who may then independently review the evidentiary basis for such recommendations.

Concerns about data non-representativeness also emerge when using data from existing databases to show the safety and efficacy of new devices. For example, one research team published in Nature Genetics that algorithms designed for polygenic risk calculations show sub-optimal results for people descending from non-European ancestry. This was perhaps unsurprising given nearly 79% of all the participants in the training database had European genetic ancestry, while they account for only 16% of global population diversity.[93]

When it comes to conducting new clinical trials and generating evidence for safety and efficacy of new devices, the specific requirements for data representativeness likely needs further clarification. Under section 513(a) of the Federal Food, Drug, and Cosmetic Act (FDCA), premarket approval applications for medical devices must contain "reasonable assurance of safety and effectiveness" that is

[89] Sathesh Kumar Annamalai, *Navigating Equivalence and Ensuring Biological Equivalence in the EU MDR: A Comprehensive Guide for Medical Device Manufacturers*, LinkedIn (Aug. 12, 2023), https://www.linkedin.com/pulse/navigating-equivalence-ensuring-biological-eu-mdr-guide-annamalai/.

[90] *Id.*; Regulation (EU) 2017/745 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC [2017] OJ L 117/1.

[91] AMANDA K. SARATA, CONG. RSCH. SERV., R47374, FDA REGULATION OF MEDICAL DEVICES (2023).

[92] Kyle J. McKibbin et al., *Reconciling Diversity in Health and Genomic Data Collection with the Regulation of AI in Clinical Genomics*, 26 GENETIC MED. (2024).

[93] Alice R. Martin et al., *Clinical use of current polygenic risk scores may exacerbate health disparities*, 51 NATURE GENETICS 584-91 (2019) doi:10.1038/s41588-019-0379-x; FLOREZ ET AL., *supra* note 8, at 303–305.

determined after "weighing any probable benefit to health from the use of the device against any probable risk of injury or illness from such use," among other relevant factors.[94] Sponsors principally demonstrate safety and efficacy by submitting valid scientific evidence.[95] FDA staff review the application and determine if the data submitted support the sponsors' claims of clinical significance, intended use, and indications for use, and substantiate that the device yields probable benefits that outweigh probable risks.[96] FDCA rules do not require data diversity in clinical evaluations.[97] However, the FDA did issue guidance in 2017, recommending trial sponsors "enroll diverse populations including representative proportions of relevant age, racial, and ethnic subgroups, which are consistent with the intended use population of the device" to meet the FDA's expectations of a well-designed clinical study.[98]

In the EU, the In Vitro Diagnostic Medical Regulation (IVDR) requires manufacturers to include information about the "representativity of a target population" in their clinical performance studies, but there has been no recent elaboration on this provision.[99] Looking at a legally binding EU Commission decision implementing common technical specifications for the In Vitro Diagnostic Devices Directive (the predecessor to the IVDR), it likely means studies should be performed

[94] Federal Food, Drug, and Cosmetic Act, 21 § U.S.C. § 360(c).

[95] 21 C.F.R. § 860.7(c)(2) (2024) (valid scientific evidence is defined as "evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use.").

[96] *Id.*

[97] *Id.*; Indeed, the FDA may be even less likely to specify such a requirement in the wake of the recent US Supreme Court ruling in *Loper Bright Enterprises v. Raimondo* that now allows courts to clarify interpretations in federal statutes, a role that was previously deferred to federal agencies. Loper Bright Enterprises v. Raimondo, 144 S. Ct. 2244, 2263 (2024).

[98] U.S. FOOD & DRUG ADMIN., EVALUATION AND REPORTING OF AGE-, RACE-, AND ETHNICITY-SPECIFIC DATA IN MEDICAL DEVICE CLINICAL STUDIES: GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF (2017).

[99] *Guidance on general principles of clinical evidence for In Vitro Diagnostic medical devices (IVDs)*, EUR. COMM'N (Jan. 2022) https://health.ec.europa.eu/system/files/2022-01/mdcg_2022-2_en.pdf.

on a population "equivalent to the European population as a whole."[100]

Besides general medical devices regulations, the emerging AI specific regulations may also impact the discussions about fairness in medical devices. Such regulations have adopted alternative approaches across jurisdictions. In the EU, the recently adopted AI Act is a sector agnostic regulation and applies to any type of AI devices, including those considered as high-risk in the medical field.[101] To define what is high-risk in the medical context, the regulation refers to the classifications provided by the MDR, meaning that this regulation will apply in tandem with EU MDR to the high-risk medical devices.[102] The AI Act for its part introduces data quality and transparency related requirements including bias monitoring, which can address some of the data bias related concerns.[103] The Act does not, however, define exactly how data representativeness or population descriptors should be interpreted.[104] For example, challenges in developing accurate population descriptors have been echoed in a recent report by a Committee of the US National Academy of Sciences, Engineering and Medicine (NASEM), which provided a conceptual framework for improving the way population descriptors are used in genetics and genomics.[105]

Regulations other than those governing medical devices, such as data protection and privacy regulations, may impact data collection for the purpose of new device approvals. For example, the EU General Data Protection Regulation (GDPR) sets stricter rules for collecting sensitive data, including health data and variables such as race and

[100] 2002 O.J. (L 131) 17.

[101] *EU AI Act: First Regulation on Artificial Intelligence, supra* note 28.

[102] Sherin Sayed & Dr. Stefanie Greifeneder, *Medical devices in the context of the European Commission's AI Regulation draft*, TAYLORWESSING (Sept. 18, 2023), https://www.taylorwessing.com/en/insights-and-events/insights/2023/09/medical-devices.

[103] 2024 O.J. (L 2024/1689) (Artificial Intelligence Act).

[104] Hannah van Kolfschoten, *The AI cycle of Health Inequity and Digital Agism, Mitigating the Biases Through the EU Regulatory Framework on Medical Devices*, 10 J. L. BIOSCIENCES 1, 1–13 (2023).

[105] *Use of Race, Ethnicity and Ancestry as Population Descriptors in Genomic Research*, NAT'L ACADEMIES SCI., ENG'G MED., https://www.nationalacademies.org/our-work/use-of-race-ethnicity-and-ancestry-as-population-descriptors-in-genomics-research (last visited Feb. 19, 2024).

ethnicity.[106] This can in principle restrict data collection and sharing, which is needed to maintain monitoring on AI devices. In principle, processing of any type of health data has been considered sensitive data under the GDPR, Art. 9, requiring a legal basis for such data processing.[107] The EU GDPR recognizes these requirements and stresses a need for adopting "adequate safeguards" for such data processing.[108]

The relevant regulations are again silent on what would constitute adequate safeguards and provide narrowed examples of pseudonymization or encryption as the technical fixes.[109] In contrast, the US regulatory framework for privacy and health data protection (i.e. HIPAA) primarily apply rules for sharing de-identified health data.[110]

IV. OTHER OVERSIGHT BODIES WITH COMPLEMENTARY FAIRNESS RESPONSIBILITIES

While official regulatory bodies such as the EU Medical Device Regulation and the FDA/FDCA in the US oversee issuing regulatory approval for AI-enabled devices, there is room for other oversight bodies such as IRBs (also known as research ethics committees (REC)) and DACs to address some of the relevant concerns related to fairness upstream of regulatory review and approval.[111]

[106] Mahsa Shabani & Sami Yilmaz, *Lawfulness in Secondary Use of Health Data Interplay Between Three Regulatory Frameworks of GDPR, DGA & EHDS, Technology and Regulation*, 2022 TECHREG 128, 128-34 (2022); Corrette Ploem & Jeanine Suurmond, *Registering Ethnicity for COVID-19 Research: Is the Law an Obstacle?,* 370 Brit. Med. J. (2020).

[107] *Id.*; Mahsa Shabani & Pascal Borry, *Rules for Processing Genetic Data for Research Purposes in View of the New EU General Data Protection Regulation*, 26 EUR. J. HUM. GENETICS 149, 149-56 (2018).

[108] Shabani & Yilmaz, *supra* note 106; Shabani & Borry, *supra* note 107.

[109] *See generally* Shabani & Yilmaz, *supra* note 106; *see generally* Shabani & Borry, *supra* note 107.

[110] U.S. DEPT. OF HEALTH AND HUM.SERV., GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA) PRIVACY RULE (Nov. 26, 2012).

[111] Marie-Charlotte Bouesseau et al, *Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants*, WORLD HEALTH ORG. (2011), https://www.ncbi.nlm.nih.gov/books/NBK310668/#ch1.s1chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf.

IRB/RECs oversee approval for research involving humans.[112] Some AI/ML-enabled tools may require testing under new clinical trials and generate new data for the purpose of regulatory approvals. In that sense, such research protocols would need to be reviewed by competent IRBs to ensure that the ethical aspects of such trials have been duly addressed - including consent forms, privacy related aspects, and the balance of risks and benefits for the research participants.[113] IRBs and their equivalent ethics review oversight bodies worldwide, we argue, are best placed to evaluate study-specific aspects related fairness when reviewing the research protocols for new AI/ML-enabled tools.[114] As IRB/REC review is an integral part of ensuring human research protections, it provides an opportunity to embed fairness related oversight on a global level, rather than to jurisdiction specific medical devices regulations.[115] In so doing, fairness related elements could be integrated in the international guidelines, such as SPIRIT-IT and CONSORT-AI, for the design and reporting of AI clinical trials[116] where such a review is formally missing from these guidelines.[117]

Data Access Committees (DACs) are provide another important oversight layer by managing requests for controlled-access data.[118] Depending on the organizational structure, DACs may work at data repositories or locally at research institutions, and have myriad duties and administrative responsibilities.[119] One aspect of fair AI relates to

[112] Gretchen E Parker, *A Framework for Navigating Institutional Review Board (IRB) Oversight in the Complicated Zone of Research*, CUREUS (2016), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5120963/.

[113] Francis McKay, *Artificial Intelligence and Medical Research Databases: Ethical Review by Data Access Committees*, 24 BMC MED ETHICS 49 (2023); *see EU AI Act: First Regulation on Artificial Intelligence*, *supra* note 28.

[114] *See generally*, McKay *supra* note 113.

[115] *See generally*, McKay *supra* note 113.

[116] Xiaoxuan Liu, *Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: the CONSORT-AI Extension*, 26 NATURE MED. 1364, 1364–74 (2020).

[117] Samantha Cruz Rivera et al., *Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: the SPIRIT-AI Extension,* 26 NATURE MED. 1351, 1351–63 (2020); Isabel Chien et al., *Multi-disciplinary fairness considerations in machine learning for clinical trials*, FACCT (2022).

[118] Mahsa Shabani et al., *Who Should Have Access to Genomic Data and How Should They be Held Accountable? Perspectives of Data Access Committee Members and Experts*, 24 EUR. J. HUM. GENETICS 1671, 1671 (2016).

[119] Jonathan Lawson et al., *Achieving Procedural Parity in Managing Access to Genomic and*

how sensitive health data was managed and processed. DACs may be assigned to review data access requests for developing new AI/ML-enabled devices.[120] In that sense, DAC members could ensure whether adequate safeguards are in place for sensitive data processing pursuant to the development and substantiation of AI/ML-enabled medical devices.[121]

V. INTRODUCING A FAIRNESS 'CHECKPOINT' IN REGULATORY REVIEW AND APPROVAL OF AI/ML-ENABLED MEDICAL

As we have shown in the previous sections, regulatory approvals for AI/ML-enabled medical devices rest on whether the applicants can satisfy evidentiary thresholds for safety and efficacy.[122] Regulatory bodies maintain oversight responsibilities for ensuring, among other things, that new medical devices comply with the highest safety standards and work as indicated.[123] Clinical evidence used to substantiate both safety and efficacy of AI/ML-enabled medical devices must therefore have all the hallmarks of quality and rigor. Who is represented in training datasets and where real-world validation takes place are issues of fairness at the level of data and which regulatory approvals should, in our view, be conditioned for AI/ML-enabled medical devices.

We argue that the existing requirements for safety and efficacy need to accentuate fairness related aspects, including data diversity and transparency for more broader populations than a narrowly defined population for intended use. We propose this can be done by introducing fairness checkpoints in the process of regulatory approvals and monitoring. In the next sections we further describe where this hypothetical checkpoint could embed in existing review pipelines

Related Health Data: A Global Survey of Data Access Committee Members, 00 BIOPRESERVATION BIOBANKING 1 (2023).

[120] Groves, *supra* note 31.

[121] Mahsa Shabani, *Will the European Health Data Space Change Data Sharing Rules?*, 375 SCI. 1357, 1357–59 (2022).

[122] *How FDA Regulates Artificial Intelligence in Medical Products*, *supra* note 87.

[123] *See How FDA Regulates Artificial Intelligence in Medical Products*, *supra* note 87; *see EU AI Act: First Regulation on Artificial Intelligence, supra* note 28.

using FDA and EU Medical devices approvals as regulatory benchmarks, and comment on the anticipated effects on application decisions.

By introducing a new checkpoint in the review and approval process for AI software and devices, we argue that federal regulators could reorient incentives for device developers to design for fairer outcomes. The proposed checkpoint is one downstream solution to identified problems of accountability and continuous monitoring of AI systems. As such, its success depends largely on complementary initiatives targeted at ensuring safe and trustworthy AI development far upstream of regulatory approval and during early model training and validation.

Building on proposals furthered in the AI ethics literature, we envision three possible checkpoint models that could achieve this desired effect: the airport security model, the fast-track model, and the nutrition label model.[124]

A. Airport model

One approach to implementing the fairness checkpoint could be modeled after modern security and surveillance entry points that authorize access to restricted spaces or services. Consider an analogy of traveling on a commercial airline through any domestic or international airport. Merely purchasing or possessing an airline ticket is insufficient to gain access to passenger only areas of the airport. Rather, passengers gain access only after passing a security checkpoint and presenting standardized credentials (e.g. valid government-issued identification). A central regulator imposes specific rules for entry (e.g. passengers must not carry dangerous weapons, liquids more than a certain volume etc.) while extension agents enforce those rules at the point of entry. Just as access at an airport is contingent on meeting security requirements, the results of a fairness checkpoint could determine whether applicants proceed to subsequent next stages of the approval process for AI/ML-based medical devices. The airport model is advantageous because all approved devices would be vetted using standardized criteria and procedures as a condition of market authorization, providing assurance that the device meets at least minimum

[124] *See* Sara Gerke, *Nutrition Facts Labels for Artificial Intelligence/Machine Learning-Based Medical Devices—The Urgent Need for Labeling Standards*, 91 GEO. WASH. L. REV. 79 (2023).

fairness requirements. While attractive, the airport model poses a unique challenge for implementation in that regulators would need to agree on common criteria and metrics for assessing fairness and communicate these to prospective applicants. The airport model would also require a significant lead time to ensure sponsors could appropriately design trials around fairness benchmarks.

B. Fast track model.

The airport model makes device approval contingent on earning a passing grade for fairness. Another potential checkpoint model could instead privilege regulatory approval for those devices proven safe and effective and which score highest on a fairness assessment. Sponsors could therefore benefit from having their approvals fast tracked in recognition of the device's specific attention to fairness, including diverse data representation, performance validation, and implementation testing in varied clinical contexts. The fast-track model would operate in parallel with ordinary review processes without impeding the potential for approval for those applicants that are not fast tracked. Therefore, this model would allow for review of device applications that demonstrate safety and efficacy but perhaps score lower on fairness measures. In this way, it overcomes at least one limitation of the airport model by not imposing any new barriers to approval. However, some, but not all, devices may be evaluated for fairness under the fast-track model and the same potential for algorithmic biases and under-representativeness in datasets remain for those non fast-tracked reviews. When devices are evaluated, sponsors may also use unstandardized measures to assess fairness, preventing direct comparisons of fairness outcomes.

C. The labeling model.

The third proposed model, the labeling model, was first proposed by Sara Gerke[125] for AI/ML-enabled medical devices that have obtained approval for market use to improve transparency and, among other things, "ensure that users know how to properly use the device and assess its benefits, potential risks, and limitations."[126] The same

[125] *Id.*

[126] *Id.*

labeling could likewise provide a useful service at the pre-market approval stage, where regulators benefit from knowing the specific characteristics of datasets used to train algorithms or enable unsupervised learning, including dataset composition, representativeness, size, etc. Once determined mechanically safe and effective, AI/ML-enabled medical devices could be given a label that reports out eleven development "facts" that Gerke proposes.[127] These facts—akin to facts found on a nutrition label—are relevant for assessing fairness at the level of the data used for device modeling. In addition to the eleven facts, the label could also detail the scope and size of training datasets, what validation tests were performed and where, and an indications checklist. The purpose of the label is to help users make better informed decisions about the appropriateness of the device at the point of use. While this model improves transparency, it assumes that prospective users (e.g. clinicians, patients, laboratory technicians) have the analytical skills to interpret the facts in context. With greater transparency around the device's developmental and investigational history, the labeling model supports the evolution of best practices in line with new technologies and new data inputs.

VI. CONCLUSION

We contend that device regulators should consider algorithmic fairness at the level of data used to support clinical evidence of safety and efficacy. At this moment, the adequacy of safeguards set by national medical devices regulations is a matter of discussion. Critics have shown that the traditional risk-based regulatory oversight for medical devices have been disrupted by the intricacies of designing, testing, and implementing AI systems *in situ*.[128] Many regulatory

[127] *Id.* at 163 (Gerke proposes eleven developmental 'facts' that should be included in AI labeling: (1) Model Identifiers; (2) Model Type; (3) Model Characteristics; (4) Indications for Use; (5) Validation and Model Performance; (6) Details on the Data Sets; (7) Preparation Before Use and Application; (8) Model Limitations, Warnings, and Precautions; (9) Alternative Choices; (10) Privacy and Security; and (11) Additional Information).

[128] Alan G. Fraser et al., *Artificial Intelligence in Medical Device Software and High-Risk Medical Devices - A Review of Definitions, Expert Recommendations and Regulatory Initiatives*, 20 EXPERT REV. MED. DEVICES 467, 467–68 (2023); Michael Bretthauer et al., *The New European Medical Device Regulation: Balancing Innovation and Patient Safety*, 176 ANNALS INTERNAL MED. 844 (2023); Anastasiya Kiseleva & Paul Quin, *Are You AI's Favorite? EU Legal Implications of Biased AI Systems in Clinical Genetics and Genomics*, 5 EPLR 155 (2021); Sven Van

frameworks for addressing AI systems in health, such as the FDA medical device requirements in the US, the European Union's Medical Devices Regulations (MDR/IVDR), and European AI Act, do not account for a broader concept of fairness or are limited to addressing bias, without further elaboration.[129] Therefore, there is an urgent need to solidify the current regulatory oversight on fairness based on an inclusive concept of fairness, and address all of its underlying aspects including, vulnerability, and discrimination.

In response, we proposed three possible models for a fairness "checkpoint" that regulatory bodies could implement into existing review and approval processes. We acknowledge that implementing any one of the fairness checkpoints may come with operational challenges. Measures of fairness can be dynamic as new data of potentially higher quality or representativeness are entered as algorithmic inputs to train or refine AI models. Establishing standards for what constitutes "fair" AI/ML device development and use presents another operational barrier as well as a conceptual challenge. The numerous fair AI frameworks are testament to the difficulty of achieving such consensus, to say nothing of how standards would be uniformly applied and updated over time. Sponsors could also adopt different metrics than those used by the regulatory bodies to assess the same fairness outcomes. This could result in discrepant outcomes of a fairness assessment that could be costly for sponsors and delay the review and approval process for otherwise safe and effective medical devices that could improve health.

Issues of fairness permeate the AI device development and deployment pipeline and have important consequences for safety and efficacy as we have argued. For these reasons, fairness deserves serious consideration as a condition of device approval to ensure that all prospective patients can equitably benefit from all that AI innovations have to offer.

Laere et al., *Clinical Decision Support and New Regulatory Frameworks for Medical Devices: Are We Ready for It? - A Viewpoint Paper*, 11 INT. J. HEALTH POL'Y MGMT. 3159, 3159–62 (2021).

[129] *See How FDA Regulates Artificial Intelligence in Medical Products*, *supra* note 87 at 5; Annamalai, *supra* note 89; *EU AI Act: First Regulation on Artificial Intelligence, supra* note 28.

ACKNOWLEDGEMENTS

We sincerely thank the editors and co-contributors to this volume, as well as colleagues across our institutions who participated in a peer review workshop that helped strengthen the writing and crystallize the ideas presented in this paper.