Chapter 24
The Development, Implementation, and Oversight of Artificial Intelligence in Health Care: Legal and Ethical Issues

Jenna Becker, Sara Gerke, and I. Glenn Cohen

Abstract Artificial Intelligence (AI), especially of the machine learning (ML) variety, is used by health care organizations to assist with a number of tasks, including diagnosing patients and optimizing operational workflows. AI products already proliferate the health care market, with usage increasing as the technology matures. Although AI may potentially revolutionize health care, the use of AI in health settings also leads to risks ranging from violating patient privacy to implementing a biased algorithm. This chapter begins with a broad overview of health care AI and how it is currently used. We then adopt a "lifecycle" approach to discussing issues with health care AI. We start by discussing the legal and ethical issues pertaining to how data to build AI are gathered in health care settings, focusing on privacy. Next, we turn to issues in algorithm development, especially algorithmic bias. We then discuss AI deployment to treat patients, focusing on informed consent. Finally, we will discuss existing oversight mechanisms for health AI in the United States: liability and regulation.

Keywords Artificial Intelligence · Health care · Machine learning · Data & health AI · Oversight

An Overview of Health Care AI

Although AI lacks a clear definition (Scherer, 2016), our discussion of AI centers around software that can reason on its own, process and identify images, or process and analyze text. A subset of AI, machine learning software, can learn and improve

J. Becker (✉) · I. G. Cohen
Harvard Law School, Harvard University, Cambridge, MA, USA
e-mail: jebecker@jd22.law.harvard.edu; igcohen@law.harvard.edu

S. Gerke
Penn State Dickinson Law, Carlisle, PA, USA
e-mail: sgerke@psu.edu

E. Valdés, J. A. Lecaros (eds.), *Handbook of Bioethical Decisions. Volume I*, Collaborative Bioethics 2, https://doi.org/10.1007/978-3-031-29451-8_24

as it is used, recognizing patterns in data (Hao, 2018). AI/ML is increasingly used in health care, from clinical support to administrative optimization. The potential for health AI is certainly great. AI-based software can be used to improve diagnostic accuracy, identify complex clinical trends, and decrease costs for health systems. However, as a novel technology, questions abound surrounding AI development and its use in health care.

Clinical AI software may be used for a wide range of purposes. These products are used today to aid in the diagnosis or treatment of patients (FDA, 2018a), detect diseases like strokes from medical images (FDA, 2018b), or predict a patient's risk of deterioration from an illness like COVID-19 (Brodwin, 2020). The Food and Drug Administration (FDA) has already cleared or approved at least over 500 AI products as medical devices (FDA, 2022a). As we will discuss in Section V, FDA only regulates a subset of clinical AI products. Therefore, the number of FDA-authorized AI products does not demonstrate the full scope of AI usage in clinical settings.

Although much of this chapter focuses on the clinical applications of AI, AI is also used in health care administration, for example: to schedule patient appointments (Murray et al., 2020), assign hospital beds (Forneas, 2018), or allocate care management resources (Obermeyer et al., 2019). Although non-clinical in nature, these algorithms can certainly impact a patient's access to care.

The development and implementation of health care AI follow a few standard steps. The AI developer must acquire data to train, validate, and test the algorithm. The AI developer must develop the algorithm and train it on the data set, as well as validate and test the model. Then a health care organization implements the AI-based software in the real world. But these seemingly straightforward steps raise a number of questions. How do developers obtain health data? When is patient privacy violated by developer use of health data? Is the data set representative of the broader patient population? In what ways can development practices create bias in health AI? Must providers obtain informed consent from patients before each AI use? How do legal and regulatory systems oversee the effectiveness of these products and their safe use? We discuss these questions in the following sections.

Obtaining Data for Health AI

An initial step when developing an AI product is obtaining relevant data for algorithm training, validation, and testing. In the health care context, this can be particularly fraught due to patient privacy considerations.

(a) Training Data: Where It Comes from and Where It's Going

AI/ML is generally trained on large data sets to ensure model accuracy. Developers may obtain these data from a number of sources. Primary health data, like patient diagnoses, clinicians' notes, and laboratory results, are often found in electronic health records (EHRs), controlled by health care organizations. The rise of ambient data collection in hospitals via audio and video collection has led to another rich

source of hospital-controlled data (Gerke et al., 2020a). Primary health data can also be found in health insurance claims, as well as laboratory and pharmacy records.

Health care AI may also be trained on data from non-traditional sources. Patient health apps, like glucose monitoring or menstrual tracking apps, store troves of user-generated data. Life insurance companies also have access to large amounts of patient health data. Finally, organizations with access to large amounts of health data may have data that can be used to make inferences about a patient's health (Price & Cohen, 2019). An individual's search history or consumer data may reveal intimate health information, such as whether the person is pregnant or lives with chronic illness.

With the rise of AI usage in health care settings, the market for health data has flourished. While health care organizations can and do develop their own AI, software companies have increasingly entered the field. Over the last few years, Google has partnered with several large health systems (Japsen, 2019; Dave, 2019; Evans, 2021) to create health AI products. Several EHR vendors have released integrated AI products. Startups have been developing health AI in a range of areas, from precision medicine to patient engagement (Toews, 2020). Thus, in many cases, AI development requires health care organizations to share patient data with third parties.

(b) Data Sharing: Protecting Patient Privacy and Autonomy

The rapid growth of health care AI development has led to questions surrounding patient privacy. First, how does sharing health data outside a health care system impact patient privacy, and how do current privacy laws guard against potential privacy harms? Second, should sharing data to develop health care AI require patient approval?

(i) Health Data Privacy in the United States

Defining privacy is a surprisingly complex task, and scholars have debated the definition of privacy for decades. But one prominent theory of privacy, useful for our purposes, defines it as "contextual integrity," where norms of information sharing are governed by the context surrounding information flow (Nissenbaum, 2004). A privacy violation occurs when these contextual norms are violated, such as when an unintended party gains access to the information.

In the case of health data, the consequences of a privacy violation can be severe. Individuals may experience social stigma and embarrassment, employment discrimination, or even be denied life insurance due to contextual privacy violations involving health information. If health privacy is under-protected, individuals are more likely to find themselves subject to such privacy harms. But if health privacy is overprotected, technological innovation may be dampened, and the benefits of applying AI/ML to large health data sets may be lost.

Health data privacy in the United States is primarily governed by the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. HIPAA's privacy protections, developed in 1996, fall short in today's technological context in two key ways.

First, HIPAA only applies to specific "covered entities" and their business associates (Gerke et al., 2020b). Under this framework, not all health data is protected by HIPAA – only individually identifiable health information controlled by specific types of organizations, like most health care providers, health plans, and healthcare clearinghouses (Price & Cohen, 2019; Gerke et al., 2020b; 45 C.F.R. §§ 160.102, 160.103). However, as described above, many more types of organizations have access to health data today, including life insurance companies and technology companies, such as Google, that are generally not considered to be "covered entities."

Second, HIPAA does not adequately protect against reidentification risk for data shared with third parties (Price & Cohen, 2019; Gerke et al., 2020b). Under the Privacy Rule, covered entities may share health data with third parties if that data is deidentified. Deidentification under HIPAA's Safe Harbor merely requires the removal of 18 discrete data elements like names, social security numbers, and telephone numbers (45 C.F.R. §§ 164.502, 164.514(a)–(b)). When health data sets deidentified under the Safe Harbor standard are combined with sufficiently large external data sets that also contain information about the patient, it may be possible in some cases to reidentify the patient – thus it is probably more accurate to say that we have made data harder to reidentify than to treat it as truly completely deidentified. This may be particularly true for some kinds of information that is relatively unique, like genetic information.

These shortcomings are meaningful in the AI context. HIPAA may not apply to a large number of AI developers with access to large health data sets, leaving individual's health data unprotected. Further, health data shared with AI vendors like Google, who have access to large amounts of consumer and in some instances location data, may be at a higher risk of being reidentifiable in some circumstances (Dinerstein v. Google, 2020). The expansion of AI development and data sharing has the potential to lead to real patient privacy harms, and federal law does not fully protect against these harms.

(ii) Patient Consent for Data Sharing

In general, under HIPAA, health systems are, for example, *not* required to obtain patient consent to share deidentified patient data with third parties. Requiring patient consent for all data use and sharing would, perhaps, increase patients' autonomy over their health information. But this requirement, if meaningful, would come at a significant cost.

Obtaining meaningful patient consent each time a patient's data are used and shared to create AI products keyed to specific uses would be monumental and could lead to statistically significant gaps in data sets.[1] If a large health system sees a million patients a year, reaching out to each patient in their EHR data set would be

[1] By "meaningful" we intend to distinguish at the extreme what we might think of as pro forma consent. For example, where the first time a patient enters a health care facility, they sign a form they likely never read that they consent to future use of their data with the identifiers stripped; if they have read the form, chances are they really do not understand the risks and benefits, because how could they without being given specifics about intended uses, what other data sets are present that may be triangulated with this data set, the cybersecurity practices of various data holders, etc.?

a time-consuming and expensive task. The health system would be less likely to want to use the health data for developing AI, which may impede innovation. There is also some concern that a regime that allowed use only with consent might generate important gaps between the data set that is generated and the true full population of patients (Cohen, 2018). Certain patient populations, like patients with stigmatized diagnoses, may be less likely to approve their health data being shared or used outside the context of their own care (Watts, 2019). This could lead to AI products being less accurate for these under-represented populations.

How to resolve the debate between patient autonomy and the benefit of access to these data is contested, but we are skeptical that paper pro forma informed consent does much to right the balance. One of us has argued that patients have a duty to share their health care data for AI and analytics purposes in some instances – where the "user" will be government or a hospital system that is directly aimed at the public good and can provide strong protection against hacking or malicious reidentification (Cohen, 2018). Assuming the risk of reidentification can be reduced, either by removing additional identifiers or through agreements with third parties, the risks to individual privacy may be low if non-zero. The potential gains from AI innovation in health care are significant, perhaps outweighing the risk of reidentification. Patient privacy and autonomy may be protected in other ways. For example, hospital-level data governance boards, made up of both patients and experts, could be utilized to protect patient interests while also not requiring individual patient input (Price & Cohen, 2019). In that model, a trained and informed group of stakeholders would weigh privacy risks against the potential technological benefits rather than relying on individual patient consent (Cohen & Mello, 2019).

But the debate is far from resolved, and indeed across the world we are seeing very different approaches.

AI Development: Data Representativeness and Algorithmic Bias

Although AI has the potential to improve health outcomes across patient populations, the risk of AI bias is also very real. This bias can develop in several ways. First, without data sets that are representative of the patient populations served by the AI, its predictions may be less accurate for those groups. Second, errors in algorithm development, such as using proxy variables, can lead to biased outcomes. Finally, AI can exacerbate existing inequities in health care, reflecting an already biased system.

(a) Data Representation

AI bias can be caused by a lack of representation in AI training data. If an algorithm is trained on data that is not reflective of the environments in which it is used, recommendations and output will be less accurate.

Patient populations vary by health system. Differences in race, ethnicity, socioeconomic status, or health conditions can lead an algorithm that performs well in one health system to perform poorly in another health care organization. For example, if an algorithm designed to detect skin cancer from photographs is trained on data from a health system with primarily White patients, the algorithm will likely degrade in performance when deployed at institutions with greater racial diversity. This would lead to bias, as the algorithm would detect cancer more accurately for White patients than for Black and Brown patients.

Patient populations, as well as treatment patterns and practices, can also vary by location. A recent study demonstrated that the majority of peer-reviewed deep learning image-based diagnostic software was trained on data from patients in California, New York, and Massachusetts (Kaushal et al., 2020). Algorithms trained only on data from certain locations may not be easily generalizable to other locations.

Professor Nicholson Price has argued that AI trained in "high resource" environments, like academic medical centers, are less effective when deployed to lower-resource settings (Price, 2019). First, patient populations differ between the institution supplying the training data and the organization deploying the algorithm. This is similar to the diversity issue discussed above, where demographic differences between patient populations may lead to bias. Second, the recommendations supplied by AI products from high resource contexts may be inappropriate in lower resource settings. An algorithm may recommend treatment that is not available in the health care organization, or it may recommend more expensive procedures over less costly but effective procedures.

The issue of data representation could be alleviated by training AI on data from a diverse group of health care organizations. But this is certainly easier said than done. Health systems developing their own AI products may struggle to find partner organizations willing to share their patient data. Smaller AI vendors may lack relationships with a large number of health systems. Or, developers may find that partnering with more famous health care organizations helps when advertising new AI products. Federal programs, like NIH's All of Us initiative, aim to help create and distribute inclusive, deidentified data sets that AI developers can use for algorithm training. But until such a program comes fully to fruition, training AI on broadly representative data may be out of reach for some developers.

(b) Algorithm Development: Labeling Bias

Issues of AI bias may also arise due to decisions made when developing an algorithm. A prime example of bias caused by algorithmic decision-making is "labeling bias." Labeling bias can occur when AI developers use proxy variables, factors used in place of the actual quantities attempting to be measured. The disconnect between what the algorithm is in fact measuring and what the algorithm is intended to measure can lead to bias (Obermeyer et al., 2021).

In a particularly notorious example, researchers found labeling bias in a widely-used algorithm used to refer patients for care management services that was developed to measure a patient's risk for requiring significant health care resources (Obermeyer et al., 2019). But rather than predicting patient health

outcomes, the algorithm instead used a patient's predicted cost as a proxy for health (Obermeyer et al., 2019). Under that framework, developers appeared to assume that lower predicted health care costs indicated better predicted health. However, health care costs for an individual patient do not only vary based on the patient's health. Cost of care also varies based on the patient's access to care. Because Black patients face unequal access to care, this use of a proxy variable led the algorithm to under-identify Black patients for increased care management resources.

Labeling bias can arise in a variety of health care settings. Although eliminating the use of proxy variables to address the potential for labeling bias is ideal, it can be challenging, if not impossible for algorithm developers to measure the "ideal target" in certain scenarios (Obermeyer et al., 2021). For example, if an emergency department triage algorithm is designed to predict the resources an incoming emergency patient will use, rather than whether the patient actually requires immediate care, the resulting algorithm may be biased based on a number of factors that impact resource consumption, including race and insurance status (Obermeyer et al., 2021). However, whether a patient needs immediate care may be difficult to measure, and these algorithms may require the use of proxy variables to approximate the "ideal target." Therefore, it is important for developers that use proxy variables to analyze their algorithms for potential bias (Obermeyer et al., 2021).

(c) Existing Bias and Disparities

Finally, health AI may be biased based on existing bias and disparities in the health care system. An algorithm's training data may be perfectly representative, but if some patient populations systemically receive poorer care than other patients, that bias will be learned and reflected in algorithmic output. Health care in the United States is racist, from medical school curricula to the historic segregation of hospitals and clinics (Benjamin, 2019). Professor Deborah Hellman has argued that using AI in such settings "compounds injustice" (Hellman, 2021). First, the data itself may reflect bias. For example, if physicians are less likely to accurately diagnose Black patients with skin cancer (McFarling, 2020), a skin cancer detection algorithm trained to learn based on prior physician diagnoses will be similarly biased. Second, the data may reflect the impact of systemic injustice on individual health. This could lead an algorithm to recommend certain treatments or resources at a higher rate for some subgroups over others, which may similarly lead an AI to be biased.

While all these sources of bias are important, an all-things-considered judgment about algorithms must also consider the extent of bias in the status quo non-AI-assisted forms of medicine that the AI seeks to improve. It may *both* be true that an AI is biased (in the sense that it performs less well for X group than Y group) and that it is *less* biased than the standard practice of medicine in a field, such that its use all-things-considered reduces bias. The Perfect should not be the enemy of the Good. But what if it both reduces bias for some groups (even the majority of groups) but exacerbates bias for some groups? How should we consider the trade-offs here? More general political theories about distribution can be helpful – one could imagine, for example, a Prioritarian theory of bias distribution where reductions in bias

to the least well-off group count "more." While most of the existing literature has focused on bias connected to what the law treats as suspect classes – race and gender – there is no reason to believe that these are the only biases rampant in AI adoption. Should, for example, a bias unrelated to a suspect classification (or only weakly associated with it), such as bias against rural patients or patients with pets, count as the worrying kind of bias in this analysis? Part of the question is how much we think the obligation to correct for bias is primarily about accuracy versus being about a way to compensate for prior forms of injustice. While the current interest in bias in health care AI is laudable, there is still plenty of first-order questions such as these for bioethicists to consider as they examine which biases to tolerate versus target.

Using AI: Is Informed Consent Required?

Once AI is developed and deployed within health care systems, we must ask whether patients should be informed on the use of AI in their care.

In the United States, the doctrine of informed consent determines what information must be disclosed to patients in the provision of their care. In standard contexts, such as surgery, this often entails a discussion of the risks and benefits of a procedure. If a patient is not sufficiently informed, a physician may be held liable for a breach of their duty to obtain informed consent.

What physicians must disclose to patients to meet informed consent requirements is primarily based on case law and varies by jurisdiction. In some jurisdictions, physicians must disclose information that a "reasonable physician" would disclose (Cohen, 2020). Other jurisdictions require physicians to disclose risks that would be "material" to the patient (Cohen, 2020). Finally, a few states limit informed consent requirements to surgical and other invasive procedures (Cohen, 2020).

Applying the doctrine of informed consent to health AI is not particularly straightforward. Let's say a physician uses an AI product as a guide in decision-making, such as in considering an AI-based recommendation as to whether to recommend a specific surgical procedure as opposed to watchful waiting. This AI-based recommendation may be one of many data points a physician reviews when making their decision for which surgical procedure to recommend to a patient. A "reasonable physician" would not generally disclose all of the factors they considered and their entire reasoning process to a patient. Is there something special about AI's contribution as opposed to, say, experience with prior patients or medical journal articles? Similarly, many of the things that go into the "old school black box" – the physician brain deciding what to recommend – are not things we typically think of as "material" for informed consent purposes. Should AI be treated differently because of particular patient sensitivity to AI involvement in care?

Legally speaking, the failure to disclose the part that AI played in a recommendation is unlikely to give rise to tort liability for failure to provide informed consent (Cohen, 2020). But ethical obligations often appropriately go beyond the legal

floor. Would it be more ethical to be very explicit about the role of AI in their decision-making? The answer is far from clear. Over-disclosure of AI usage, even when AI use is not material to a patient, may make it challenging for patients to meaningfully evaluate risks (Cohen, 2020). As AI becomes more prevalent in health care, patients may be so inundated by disclosures that they are unable to analyze the risks of each product.

But some scenarios may arise where patients may reasonably expect to be informed of AI usage along with its associated risks. For example, a patient may find a physician's AI use material if a health system maintains a policy requiring physicians to follow the recommendation of AI-based software. Rather than weighing the recommendation of the software along with the physicians' own knowledge and training, the AI product would become the sole determinant of a patient's care plan.

Disclosure might be more important when an AI product plays an outsized role in a patient's care. For example, assume a physician relies on an AI recommendation, as if the AI-based software is a specialist with relevant expertise. The patient should perhaps be informed that the physician lacks sufficient expertise and is relying on an AI product as a quasi-member of the patient's care team (Cohen, 2020). Some scholars have argued that physicians should be required to elucidate the role played by AI in a patient's care (Schiff & Borenstein, 2019).

There has been a particular concern in the law and ethics of AI with "black-box" systems, where AI is not interpretable nor explainable, such as many neural net systems (Babic et al., 2021). Should a physician disclose to the patient that an AI was involved in the care and the reason why the AI made the recommendation it did was *not* one the physician could explain even if she wanted to? Patients may not trust such an opaque recommendation. On the other hand, physicians regularly rely on products they do not understand, including aspirin. Explanation is just one epistemic warrant that something will be good for a patient. If a provider does not understand *how* a particular drug or device works, they may still be confident that the product *does* work, based on clinical trials or other evidence that underly regulatory approval (London, 2019). However, in the current regulatory world, much of the AI used in health care has *not* gone through rigorous clinical trials or a searching regulatory review. Should we "default" into disclosure for such AI systems? Is there a way to make that consent meaningful, especially given the opaque nature of these systems?

Finally, does the analysis of informed consent change when a system is used to help make decisions to allocate rivalrous goods such as an organ, an ICU bed, etc.? If a particular patient refuses to allow AI involvement in that decision-making, this affects not only what they will receive but also the distribution to other claimants. Is this an instance where "informed consent" should be bifurcated – patients should be informed about AI involvement in their care, but if they want to be considered for the allocation *not* be given an opportunity to opt-out of AI analysis?

These are heady questions bioethics has only begun to wrestle with.

Oversight of Health AI

At least two existing mechanisms can be used to oversee health AI development and use in the United States. Physicians, health care organizations, and AI developers may be held liable in tort when patients are harmed by health AI usage. Further, some health AI products are currently regulated by FDA.

(a) Liability for Health AI Use, Implementation, and Development

New health care technologies like AI often lead to complex questions surrounding liability. Physicians, health systems, and software developers (among other actors) may be held liable for patient injury caused by health AI (Maliha et al., 2021). Can the United States' liability system adapt to balance patient protection from dangerous products while also encouraging innovation and the adoption of innovative technologies?

(i) Physician Liability

Physicians may be held liable for medical malpractice if their use of an AI product leads to patient harm. For example, if a physician follows the recommendation of a patient deterioration algorithm that suggests a specific intervention and that intervention harms the patient, the physician may be held liable for the injury. However, under the current liability framework, a doctor would not *always* be liable in this scenario. Instead, liability often depends on whether a physician followed the standard of care expected from such a clinician.

Some scholars (including two of us) have suggested that the current rules of tort liability will prompt physicians afraid of malpractice to use AI merely for confirmatory purposes, to follow the current standard of care (Price et al., 2019). Of course, this narrow use would significantly limit the potential benefits of AI usage whose main goal is to improve overall outcomes in medical care and/or to tailor care to the needs of specific patient populations. For example, if an algorithm used to predict patient deterioration suggests an intervention that deviates from standard practice but leads to a higher survival rate for critically ill patients, we *want* the physician to depart from the standard of care in that case. More generally, it is important that the liability framework for physicians should not deter physicians from using AI when it improves patient care.

Of course, determinations about departures of the standard of care are often in the hands of juries. A recent study found, using individuals playing the role of jurors, that physician liability for AI usage is influenced by whether the AI output deviates from the standard of care, but that the standard of care is not the only factor considered by juries (Tobia et al., 2021). Physicians are less likely to be held liable for harm caused by following the recommendation of an AI product that aligns with the standard of care. But physicians are not necessarily shielded from liability by rejecting all AI recommendations that deviate from the standard of care. Instead, the study indicated that jurors would also give significant weight to whether the

physician followed the AI recommendation, displaying a greater trust in AI among lay individuals than anticipated (Tobia et al., 2021; Price et al., 2021).

Because of the centrality of the standard of care for physician liability for AI usage, it is important to acknowledge how the standard of care can change over time. Physicians should encourage their professional organizations to take active steps to evaluate practice-specific algorithms, and in so doing, may shape the law's understanding of when following and disregarding an AI are appropriate or not.

(ii) Health System Liability

Hospitals and health systems may also be held liable for AI usage and implementation. Health systems may be the better actor to accrue liability for harms caused by the use and implementation of products like medical AI. First, hospitals and health care organizations are likely more financially equipped than individual physicians to pay for damages and increasing insurance rates. Second, removing the burden from individual physicians may help encourage physicians to use new AI products. Finally, health systems are already responsible for the safety of their medical equipment and clinician training.

Hospitals can use their resources to ensure that AI products are implemented safely and that clinicians are properly trained on their use. For example, a hospital system may be held liable if they fail to train physicians on the use of AI, leading to patient harm. Or, a health care organization that does not ensure that an AI system is safe for its intended use, such as a children's hospital implementing an AI tool designed for adult patients without retraining or testing, would likely be liable for any resulting injury (Maliha et al., 2021). The potential for liability may discourage some hospital systems from implementing AI-based products. But hospitals are no more likely to accrue liability for AI usage than for implementing any other novel medical device.

(iii) AI Developer Liability

AI developers may be held liable for product defects, although the barriers to establishing liability over AI vendors are relatively high. A key barrier to liability may be that the bulk of medical AI in use today is used to *aid* physician practice and decisions rather than being used on its own to directly treat patients (Price, 2017). Further, unlike tangible products, software products do not easily fit into the existing products liability framework (Brown & Miller, 2014). Finally, as the regulation of these products ramps up, AI developers are less likely to be held liable for product failures that harm patients (Maliha et al., 2021). Instead, that liability is more likely to fall back on physicians using the software and health systems implementing the product. Perhaps, especially in the case of black-box algorithms where physicians and hospital systems may be unable to sufficiently audit the effectiveness of AI-based recommendations, liability will shift to hold developers accountable for errors (Maliha et al., 2021). However, as our system of liability currently stands, liability concerns should not deter AI developers from continuing to create innovative health care algorithms. But concerns about liability may be so low that

developers are not sufficiently incentivized to develop safe products without an effective regulatory regime.

(b) FDA Authorization of AI-Based Medical Devices

Regulation is another oversight mechanism that can help ensure health AI products are safe and effective. FDA regulates medical devices, including health care AI that qualifies as Software as a Medical Device (SaMD) under the Federal Food, Drug, and Cosmetic Act (FDCA). Although FDA has the authority to regulate AI products, the agency's authority over health AI is somewhat limited, and the agency's regulatory plans remain unclear.

(i) Health AI Constituting a Medical Device

FDA's authority over health care AI is relatively narrow under the FDCA. FDA regulates devices designed for "use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease" (FDCA § 201(h)(1)). Although this authority certainly covers a broad array of AI-based software, software that impacts patient care through use in administrative or operational contexts fall well outside FDA's authority. The Twenty-First Century Cures Act narrowed the medical device definition in 2016. For example, in general, software that supports or provides recommendations to clinicians is not considered to be a medical device and thus not regulable by FDA if the product also provides an explanation of its recommendation that is understandable by the intended user (FDCA § 520(o)(1)(E)).

Of the AI-based software that may be regulated by FDA, the agency maintains discretion over which algorithms it will actually regulate. FDA released guidance in September 2022, which expanded the scope of clinical decision support software the agency intends to regulate (FDA, 2022b).

(ii) FDA Regulatory Structure and Challenges

The regulation of AI-based software leads to a number of unique challenges not faced by FDA in its regulation of tangible medical devices. FDA's regulatory plans for software devices, and AI-based software more specifically, remain in flux.

FDA's traditional device review mechanisms, as FDA has noted (Gottlieb, 2017), do not translate well to the oversight of AI-based medical devices, especially "adaptive" AI algorithms that learn and update with use. FDA piloted a certification program specifically for software devices called the Software Pre-Certification Program (Pre-Cert) (FDA, 2019a). The Pre-Cert program would have allowed algorithm developers that demonstrate excellence in key areas like product quality and patient safety to be eligible for a more streamlined premarket review of their software devices or no premarket review at all. However, after the completion of the pilot, FDA sunseted the Pre-Cert program (FDA, 2022c), leaving open questions surrounding how FDA will regulate software devices.

A couple of key issues remain unaddressed by FDA's current plans. First, how FDA will ensure the safety of algorithm updates, especially for adaptive AI, is unclear. Although FDA has released a discussion paper surrounding updates for AI/

ML-based SaMD (FDA, 2019b) and a recent Action Plan (FDA, 2021), much still needs to be figured out, such as how to continuously ensure the safety and effectiveness of these devices (Babic et al., 2019; Gerke, 2021).

Second, truly understanding the impact of health AI in practice not only requires an understanding of whether the medical device itself is accurate, but also on a wide range of external factors, like the accuracy of medical record input data, how clinicians will react to device recommendations, and the longer-term impact on patient outcomes. Addressing these contextual variables requires a "system view" approach (Gerke et al., 2020c) whereby regulators would, for example, require more frequent human factors testing.

Conclusion

AI has the potential to transform health care, improving patient outcomes and reducing administrative inefficiencies. But a number of issues remain unsettled, such as protecting patient privacy, preventing algorithmic bias, whether to obtain informed consent, and establishing effective oversight structures. These issues must be addressed to ensure the safe, effective, and ethical deployment of health care AI.

Acknowledgments I.G.C. was supported by a grant from the Collaborative Research Program for Biomedical Innovation Law, a scientifically independent collaborative research program supported by a Novo Nordisk Foundation grant (NNF17SA0027784). I.G.C. was also supported by Diagnosing in the Home: The Ethical, Legal, and Regulatory Challenges and Opportunities of Digital Home Health, a grant from the Gordon and Betty Moore Foundation (grant agreement number 9974). S.G. reports grants from the European Union (Grant Agreement no. 101057321 and no. 101057099), the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (Grant Agreement no. 3R01EB027650-03S1), and the Rock Ethics Institute at Penn State University.

References

Babic, B., et al. (2019). Algorithms on regulatory lockdown in medicine. *Science, 366*, 1202–1204.

Babic, B., et al. (2021). Beware explanations from AI in health care. *Science, 373*, 284–286.

Benjamin, R. (2019). Assessing risk, automating racism. *Science, 366*, 421–422.

Brodwin, E. (2020). Health systems are using AI to predict severe Covid-19 cases. But limited data could produce unreliable results. *STAT News*. https://www.statnews.com/2020/11/18/covid-19-algorithms-reliability-epic-cerner/. Accessed 13 July 2021.

Brown, S. H., & Miller, R. A. (2014). Legal and regulatory issues related to the use of clinical software in health care delivery. In R. Greenes (Ed.), *Clinical decision support* (2nd ed., pp. 711–740). Elsevier.

Cohen, I. G. (2018). Is there a duty to share health care data? In I. G. Cohen et al. (Eds.), *Big data, health law, and bioethics* (pp. 209–222). Cambridge University Press.

Cohen, I. G. (2020). Informed consent and medical artificial intelligence: What to tell the patient? *The Georgetown Law Journal, 108*(6), 1425–1469.

Cohen, I. G., & Mello, M. M. (2019). Big data, big tech, and protecting patient privacy. *JAMA, 322*(12), 1141–1142.

Dave, P. (2019). Google signs healthcare data and cloud computing deal with Ascension. *Reuters*. https://www.reuters.com/article/us-alphabet-ascension-privacy/google-signs-healthcare-data-and-cloud-computing-deal-with-ascension-idUSKBN1XL2AT. Accessed 13 July 2021.

Dinerstein v. Google. (2020). 484 F.Supp.3d 561.

Evans, M. (2021). Google strikes Deal with hospital chain to develop healthcare algorithms. *Wall Street Journal*. https://www.wsj.com/articles/google-strikes-deal-with-hospital-chain-to-develop-healthcare-algorithms-11622030401. Accessed 13 July 2021.

FDA. (2018a). FDA. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye. Accessed 13 July 2021.

FDA. (2018b). *FDA permits marketing of clinical decision support software for alerting providers of a potential stroke in patients*. https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-clinical-decision-support-software-alerting-providers-potential-stroke. Accessed 13 July 2021.

FDA. (2019a). *Developing a Software Precertification Program: A Working Model v1.0*. https://www.fda.gov/media/119722/download. Accessed 14 July 2021.

FDA. (2019b). *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) – Discussion Paper and Request for Feedback*. https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf. Accessed 14 July 2021.

FDA. (2021). *Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device Action Plan*. https://www.fda.gov/media/145022/download. Accessed 14 July 2021.

FDA. (2022a). *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. Accessed 9 April 2023.

FDA. (2022b). *Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff*. https://www.fda.gov/media/109618/download. Accessed 9 April 2023.

FDA. (2022c). *Digital Health Software Precertification (Pre-Cert) Pilot Program*. https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-pilot-program. Last accessed 9 April 2023.

Forneas, N. (2018). *Improving hospital bed management with AI*. IBM. https://www.ibm.com/blogs/client-voices/improving-hospital-bed-management-ai/. Accessed 13 July 2021.

Gerke, S., et al. (2020a). Ethical and legal aspects of ambient intelligence in hospitals. *JAMA, 323*(7), 601–602.

Gerke, S., et al. (2020b). Ethical and legal challenges of artificial intelligence-driven healthcare. In A. Bohr & K. Memarzadeh (Eds.), *Artificial intelligence in healthcare* (1st ed., pp. 295–328). Elsevier.

Gerke, S., et al. (2020c). The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPJ Digital Medicine, 3*, 1–4.

Gerke, S. (2021). Health AI for Good Rather than Evil? The Need for a New Regulatory Framework for AI-Based Medical Devices. *20 Yale Journal of Health Policy, Law, and Ethics, 433*.

Gottlieb, S. (2017). *FDA announces new steps to empower consumers and advance digital healthcare*. FDA. https://www.fda.gov/news-events/fda-voices/fda-announces-new-steps-empower-consumers-and-advance-digital-healthcare. Accessed 14 July 2021.

Hao, K. (2018). What is machine learning? *MIT Technology Review*. https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/. Accessed 13 July 2021.

Hellman, D. (2021). *Big data and compounding injustice*. SSRN. Available from https://ssrn.com/abstract=3840175. Accessed 19 July 2021.

Japsen, B. (2019). Mayo Clinic, Google partner on digital health Analytics. *Forbes*. https://www.forbes.com/sites/brucejapsen/2019/09/10/mayo-clinic-google-partner-on-digital-health-analytics/. Accessed 13 July 2021.
Kaushal, A., et al. (2020). Geographic distribution of US cohorts used to train deep learning algorithms. *JAMA, 324*(12), 1212–1213.
London, A. J. (2019). Artificial intelligence and black-box medical decisions: Accuracy versus Explainability. *The Hastings Center Report, 49*(1), 15–21.
Maliha, G., et al. (2021). Artificial intelligence and liability in medicine: Balancing safety and innovation. *The Milbank Quarterly, 00*, 1–19.
McFarling, U. L. (2020). Dermatology faces a reckoning: Lack of darker skin in textbooks and journals harms care for patients of color. *STAT News*. https://www.statnews.com/2020/07/21/dermatology-faces-reckoning-lack-of-darker-skin-in-textbooks-journals-harms-patients-of-color/. Accessed 13 July 2021.
Murray, S. G., et al. (2020). Discrimination by artificial intelligence in a commercial electronic health record – A case study. Health *Affairs Blog*. https://www.healthaffairs.org/do/10.1377/hblog20200128.626576/full/. Accessed 13 July 2021.
Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*, 119–157.
Obermeyer, Z., et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*, 447–453.
Obermeyer, Z., et al. (2021). *Algorithmic bias playbook*. Center for Applied Artificial Intelligence at Chicago Booth. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook. Accessed 13 July 2021.
Price, W. N. (2017). Artificial intelligence in health care: Applications and legal implications. *The SciTech Lawyer, 14*(1), 10–13.
Price, W. N. (2019). Medical AI and contextual bias. *Harvard Journal of Law & Technology, 33*, 65–116.
Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine, 25*, 37–43.
Price, W. N., et al. (2019). Potential liability for physicians using artificial intelligence. *JAMA, 322*(18), 1765–1766.
Price, W. N., et al. (2021). How much can potential jurors tell us about liability for medical artificial intelligence? *Journal of Nuclear Medicine, 62*(1), 15–16.
Scherer, M. U. (2016). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law & Technology, 29*(2), 353–400.
Schiff, D., & Borenstein, J. (2019). How should clinicians communicate with patients about the roles of artificially intelligent team members? *AMA Journal of Ethics, 21*(2), E138–E145.
Tobia, K., et al. (2021). When does physician use of AI increase liability? *Journal of Nuclear Medicine, 62*(1), 17–21.
Toews, R. (2020). *These are the startups applying AI to transform healthcare*. https://www.forbes.com/sites/robtoews/2020/08/26/ai-will-revolutionize-healthcare-the-transformation-has-already-begun/. Accessed 13 July 2021.
Watts, G. (2019). Data sharing: Keeping patients on board. *Lancet Digit Health, 1*(7), E332–E333.

Jenna Becker, JD, was Student Fellow at the Petrie-Flom Center for Health Law Policy, Biotechnology & Bioethics, at Harvard Law School. Her research interests include the regulation of medical software and technologies.

Sara Gerke is an Assistant Professor of Law at Penn State Dickinson Law. Her current research focuses on the ethical and legal challenges of artificial intelligence and big data for health care and health law in the United States and Europe. Professor Gerke is a co-principal investigator in two EU-funded projects, CLASSICA (Validating AI in Classifying Cancer in Real-Time Surgery) and OperA (Optimizing Colorectal Cancer Prevention Through Personalized Treatment with Artificial

Intelligence). She is also a leadership team member of the Project on Precision Medicine, Artificial Intelligence, and the Law (PMAIL) at the Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School. Moreover, she is the co-investigator of the supplemental project "PREMIERE: A PREdictive Model Index and Exchange Repository" supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health and the Rock Ethics Institute at Penn State University.

I. Glenn Cohen is the James A. Attwood and Leslie Williams Professor of Law and a Deputy Dean at Harvard Law School. He is also the Faculty Director, Petrie-Flom Center for Health Law Policy, Biotechnology & Bioethics. He is the author of more than 200 articles and book chapters and is the author, co-author, editor, or co-editor of more than 18 books. He is one of the world's leading experts on the intersection of bioethics (sometimes also called "medical ethics") and the law, as well as health law.