21 Hous. J. Health L. & Policy

Houston Journal of Health Law & Policy

AI'S LEGITIMATE INTEREST: TOWARDS A PUBLIC BENEFIT PRIVACY MODEL

Charlotte A. Tschider*

* Loyola University Chicago School of Law, Beazley Institute for Health Law & Policy. The Author would like to thank the Houston Journal for Health Law and Policy Symposium participants, Glenn Cohen and the Harvard Law School Health Law, Bioethics, and Biotechnology Workshop for their excellent feedback on an early draft of this paper, participants of Washington University's Cordell Institute Symposium on "The Future of Consent for Privacy and Health," especially Woodrow Hartzog, and participants of the University of North Carolina Symposium on "Privacy Norms Across Borders and Boundaries," especially Anne Klinefelter.

ABSTRACT

Health data uses are on the rise. Increasingly more often, data are used for a variety of operational, diagnostic, and technical uses, as in the Internet of Health Things. Never has quality data been more necessary: large data stores now power the most advanced artificial intelligence applications, applications that may enable early diagnosis of chronic diseases and enable personalized medical treatment. These data, both personally identifiable and de-identified, have the potential to dramatically improve the quality, effectiveness, and safety of artificial intelligence.

Existing privacy laws do not 1) effectively protect the privacy interests of individuals and 2) provide the flexibility needed to support artificial intelligence applications. This paper identifies some of the key challenges with existing privacy laws, including the ineffectiveness of de-identification and data minimization protocols in practice and issues with notice and consent as they apply to artificial intelligence applications, then proposes an alternative privacy model. This model specifically rejects a notice and consent model in favor of legitimate interest analysis. This approach introduces a more restrictive application of health privacy law while adopting a flexible, interest-balancing approach to permit additional data uses that primarily benefit individuals and communities.

INTRODUCTION

Health data uses are increasing, and data are becoming even more essential in health applications. Health data elements are unlike other types of consumer data because they can be used for new uses: quality and efficiency in care, improvements in diagnostic processes, or cross-product treatment and system efficacy. Artificial intelligence (AI) has the potential to revolutionize healthcare through data use. But how do organizations collect the vast data volume needed to power AI when privacy law could impede data flow?

Much has been written on the degree to which algorithmic decision-making should be more transparent to facilitate fairness and non-discrimination goals.[1] Others, including this Author, have focused more generally on the inadequacies of the U.S. privacy system for healthcare technologies.[2] Scholars have highlighted the likelihood of AI-caused injuries, including who or what should be held legally accountable for such injuries, without discussing the natural tension between safety and privacy rights: increased access to personal

[1] *See generally* Mason Marks, *Algorithmic Disability Discrimination*, in DISABILITY, HEALTH, LAW, AND BIOETHICS (2020) (I. Glenn Cohen et al. eds., Cambridge Univ. Press 2020) (identifying key risks to individuals with disabilities by using AI to treat people with disabilities differently, such as exploiting them); Andrew Selbst & Salon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORD. L. REV. 1085, 1118 (2018) (highlighting the 'inherent good ' associated with explainability to understand decisions that impact individual options); Anya Prince & Daniel Schwarcz , *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257(2020) (describing the potential for discrimination even when data that could lead to discrimination are not directly collected by an organization); Frank Pasquale, *Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society*, 78 OHIO ST. L. J. 1243, 1247 (2017) (explaining how algorithms can lead to unfairness and discrimination issues, such as 'algorithmic nuisance'); Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent*, 93 TEX. L. REV. 85, 140-46 (2014) (introducing IoT's key areas of concern); Charlotte A. Tschider, *Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age*, 96 DEN. L. REV. 87, 97-98 (2018) (analyzing the potential for discrimination associated with big data feeding AI algorithmic decision-making).

[2] *See, e.g.*, Nicolas P. Terry, *Will the Internet of Things Transform Healthcare?*, 19 VAND. J. ENT. & TECH. L. 327, 338-39 (2016) (positing that most mobile and software providers will not be subject to HIPAA); Nicolas P. Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX 65, 80 (2014) (noting HIPAA's broad exceptions to specific privacy obligations, for example 'laundered, 'or inferential, health data); Charlotte A. Tschider, *Enhancing Cybersecurity for the Digital Health Marketplace*, 26 ANNALS HEALTH L. 1, 1, 10, 16, 29 (2017) (concluding that many digital health devices will only be subject to general FTC oversight).

information often results in safer and more efficacious AI products.[3] Each of these investigations, while critically important, does not directly examine this tension or recommend a specific course of action for balancing privacy and safety interests in an effective way.[4] Data use for purposes of improving products or offering new products that benefit public health may indeed be justification for data use beyond originally disclosed purposes, such as what has been described in a privacy notice or notice of privacy practices.

Healthcare privacy is an essential part of developing trust in healthcare treatment and facilitating effective insurance transactions. Without effective privacy commitments in healthcare, patients may not be willing to offer accurate information to enable effective diagnosis and treatment or may not seek health treatment at all.[5] Indeed, individually identifiable health data elements are especially sensitive because they are a digital extension of a person's physical and mental body and bodily function, or *datafication*.[6] Individually

[3] *See generally* W. Nicholson Price, *Medical Malpractice and Black-box Medicine*, BIG DATA, HEALTH LAW, AND BIOETHICS (2018) (I. Glenn Cohen et al. eds., Cambridge Univ. Press 2018) (describing potential malpractice suits related to the use of AI in medicine); *see* Andrew D. Selbst, *Negligence and AI's Human Uses*, 100 B.U. L. REV. 1315, 1329 (2020); *see* Rebecca Crootof, *Internet of Torts*, 69 DUKE L. J. 583, 607 (2019) (describing IoT system liability in health applications, many of which will increasingly be connected to AI systems); W. Nicholson Price et al., *Potential Liability for Physicians Using Artificial Intelligence*, 322 JAMA 1765 (Oct. 2019); Pasquale, *supra* note 1, at 1247.

[4] Ryan Calo does offer some perspective related to safety and privacy, specifically the "data parity problem," wherein data will be consumed in large volume by AI, and these data must be supplied from somewhere, likely data sources that have substantially more power and reach than the individuals about whom data are collected. *See* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS 399, 424 (2017). Mason Marks similarly addresses both safety and privacy as separate considerations in the context of suicide prevention. Mason Marks, *Artificial Intelligence Based Suicide Prevention*, 21 YALE J. L. & TECH. 98, 111(2019). Although tremendously important in terms of the issues that apply to AI-enabled robotics and AI suicide prevention, respectively, Calo and Marks do not squarely address collision of safety and privacy where competing interests may cut in favor of privacy on one hand or safety on the other.

[5] W. Nicholson Price, II & I. Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 NAT. MED. 37 (2019) (noting the importance of trust in any big data transactions, both from the perspectives of the patient and the physician).

[6] Indeed, datafication as it pertains to patients is intended to ultimately benefit the patient. The degree of such a benefit, however, is often up for debate. *See* Kristen Ostherr, *Privacy, Data Mining, and Data Profiling in Online Patient Narratives*, 4 CATALYST: FEMINISM, THEORY, TECHNOSCI. 1, 2-5 (2018), https://catalystjournal.org/index.php/catalyst/article/view/29628/html.

identifiable health data are also at increased risk of misuse, unauthorized disclosure, and use for discriminatory purposes.[7]

Data, however, are also crucial to modern healthcare. Data are used not only for directly providing healthcare but also for measuring quality in such transactions, advancing efficiency goals, and reducing costs.[8] Data are used for improving product functionality and for creating new AI-enabled products. Although for certain types of uses, data need not be highly sensitive or even individually identifiable, AI systems will likely use at least *some* individually identifiable data.[9] The necessity of such data complicates privacy compliance: usually the designers of AI systems do not know which data will be useful at the time of collection.[10] Further, the black-box nature of AI, either through dynamic inscrutability or trade secrecy, makes it nearly impossible to disclose the extent to which data are actually used.[11]

---

[7] *See* Marks, *supra* note 1, at 18; Tschider, *supra* note 1, at 122-23; Charlotte Tschider & Krista Kennedy, *Data Discrimination: The International Regulatory Impasse of AI-Enabled Medical Wearables*, *in* LEGAL, SOC. & ETHICAL PERSP. ON HEALTH & TECH. (Motahareh Fathisalout Bollon & Anna Berti Suman eds., USMB 2020) (describing key issues related to reliance of audiology patients on doctors and audiologists, including the fact that data use and associated data ecosystems are generally opaque to principal health workers, let alone patients); *see* Price & Cohen, *supra* note 5, at 37 (separating potential risk into deontological and consequentialist concerns, wherein consequentialist concerns include tangible negative consequences).

[8] *See generally* Alessandro Capone et al., *Health Data Entanglement and Artificial Intelligence-Based Analysis: A Brand New Methodology to Improve the Effectiveness of Healthcare Services*, 167 CLIN. TER. 102 (2016) (describing the value of health data for quality and operations purposes).

[9] Guy Pearce, *Beware the Privacy Violations in Artificial Intelligence Applications*, ISACA NOW BLOG (May 28, 2021), https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2021/beware-the-privacy-violations-in-artificial-intelligence-applications.

[10] *See* U.S. DEP'T OF HEALTH & HUMAN SERVS., *SUMMARY OF THE HIPAA PRIVACY RULE* 1526-27 (2013), https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[11] Black-box medicine is medicine conducted through opaque, automated methods. For example, AI diagnostic tools could diagnose cancer with a 90% probability from mammogram images without explaining how that diagnosis was made. W. Nicholson Price, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 421 (2015) (introducing the concept of black-box medicine); W. Nicholson Price, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421,423 (2017) (identifying approached to regulate opaque medical AI); *see* Roger Allan Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOMM. & TECH. L. REV. 1 (2016) (identifying models for reviewing medical AI, including issues related to privacy in disclosure of data associated with medical AI); Charlotte Tschider, *Beyond the Black Box*, 98 DENV. L. REV. 683 (2020) (describing AI's dynamic inscrutability and advocating for alternative models to determine AI quality).

In this paper, I directly examine this intersection of data use interests and privacy interests and recommend two example models that the U.S. Department of Health and Humans Services (HHS) and the Federal Trade Commission (FTC) could use to better analyze whether data use is "necessary." I also recommend how both agencies might better govern secondary data use when such use has a "substantial public benefit," specifically through an interest-balancing approach. These two models illustrate how interest-balancing could work within the current HIPAA and FTC Fair Information Practices models, or what features a new policy model could have to balance these interests.

Part I introduces AI technology in healthcare, offering case studies of AI applications. Part II describes the history and function of health privacy regulation, introducing how the U.S. model was not designed for AI use with big data. Part III explores the inconsistent and frequently diverging aims of healthcare privacy and AI safety and reliability, specifically how responsible AI development requires an evolution of thought regarding healthcare privacy. In Part IV, I recommend an approach for balancing the interests of data use while simultaneously toughening privacy requirements in HHS, FTC, and state law models. Such data interest approaches should balance interests in favor of patient and public benefit rather than promoting opaque organizational data benefits.

I. AI AND BIG DATA APPLICATIONS

Popular television shows and movies on AI, such as the *Westworld* HBO TV series and the 2014 film *Ex Machina*, illustrate the popular association between AI and sentience.[12] Yet, the most common AI technology involves algorithms designed based on data stored in a database, and recommending decisions, commonly called *machine*

[12] Denise Chow et al., *'Westworld 'Science Advisor Shares His Vision of Robots and the Future of AI,* NBC NEWS *(2018), https://www.nbcnews.com/mach/science/westworld-science-adviser-shares-his-vision-robots-future-ai-ncna883321;* Alex Garland, *Alex Garland of 'Ex Machina 'Talks About Artificial Intelligence*, THE N.Y. TIMES (Apr. 22, 2015); Bobby Azarian, *The Myth of Sentient Machines*, PSYCHOLOGY TODAY (June 1, 2016), https://www.psychologytoday.com/us/blog/mind-in-the-machine/201606/the-myth-sentient-machines.

learning.[13] Machine learning is the most commonly used AI approach in healthcare today, and it can be used for any number of tasks from task automation to analytics.[14]

The foundation of Artificial Intelligence (AI) is data.[15] From relatively simple, nonlinear algorithms, to supervised or unsupervised machine learning and deep learning through advanced neural networks, data are used to create algorithms that render a decisional result.[16] Machine learning applications use exceptionally large volumes of data, which are analyzed by a machine learning utility to determine interrelationships between these data.[17] As data change, frequently so does the algorithm.

The remarkable ability of AI to "self-learn" through reassessing big data relationships then updating its algorithms is what separates AI from traditional data science and human-designed algorithms. The most complex problems typically require the most complex AI solutions. And, the more complex the AI system and algorithm, the more important data volume and data quality become.[18]

A. Artificial Intelligence's Big Data Dependency

Although the idea of a digital computer began with Charles Babbage's idea for a "digital engine," an entirely mechanical machine, the concept was not realized until Alan Turing proposed the concept of a "universal computer."[19] The universal computer could be adapted for multiple purposes, with storage, an operating executive unit to direct core behavior, and controls or rules that govern behavior.[20]

---

[13] Bernard Marr, *What Is the Difference Between Artificial Intelligence and Machine Learning?*, FORBES (Dec. 6, 2016), https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#101fde3b2742.

[14] Tirthajyoti Sarkar, *AI and Machine Learning for Healthcare*, TOWARDS DATA SCI. (Apr. 24, 2020), https://towardsdatascience.com/ai-and-machine-learning-for-healthcare-7a70fb3acb67.

[15] Jack M. Balkin, *2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data*, 78 OHIO L. J. 1217, 1221 (2017).

[16] Jason Brownlee, *How Much Training Data is Required for Machine Learning?* MACHINE LEARNING MASTERY (July 24, 2017), https://machinelearningmastery.com/much-training-data-required-machine-learning/; *see* Tschider, *supra* note 11, at 690.

[17] *See* Tschider, *supra* note 11, at 690-92.

[18] For a more comprehensive explanation of AI type and function, see *id.*

[19] Alan M. Turing, *Computing Machinery and Intelligence*, 49 MIND 433, 439 (1950).

[20] *Id.* at 437.

Turing described a universal computer equipped with an infinite data store, which could be called an "infinitive capacity computer."[21]

The universal computer, or the Universal Turing Machine, anticipated continuous learning by permitting controls or rules to be updated like data, using a meta-logical interpreter.[22] This continuous learning capacity is key for AI in differentiating between human-designed algorithms and self-executing AI.[23] These systems leverage substantial computing power, as Turing suggested,[24] coupled with big data stores, or databases designed for broad data collection and use.[25] In short, modern AI is exactly as it was originally envisioned by Turing: AI's success or failure is directly connected to the quality and volume of data used to train AI algorithms or improve its accuracy over time.[26]

Powerful computers excel at evaluating large volumes of data to identify referential relationships. For example, an AI data set including radiological images may include thousands, if not more, images.[27] Each image carries with it data points that must be analyzed in relationship to others, and an AI algorithm may analyze imaging data points as well as other inputs, such as specific symptoms or lab results.[28] For example, a typical diagnostic process for colon cancer would involve analyzing lab tests for blood in a patient's stool, blood tests indicating anemia, colonoscopy results, and images. However, an AI diagnostic test could potentially diagnose colon cancer

---

[21] *Id.* at 438-39.

[22] *Id.* at 440.

[23] *Id.* at 439; Ben Lorica, *Why Continuous Learning is Key to AI*, O'REILLY (Aug. 7, 2017), https://www.oreilly.com/radar/why-continuous-learning-is-key-to-ai/.

[24] *Id.* at 445.

[25] Big data are described in terms of volume, velocity, and variety, designed specifically for high availability use along with accommodating a diversity and great number of data elements. *Big Data, Artificial Intelligence, Machine Learning and Data Protection*, INFO. COMM'R'S OFF. 1, 6-7.

[26] Willem Sundblad, *Data Is the Foundation for Artificial Intelligence and Machine Learning*, FORBES (Oct. 18, 2018), https://www.forbes.com/sites/willemsundbladeurope/2018/10/18/data-is-the-foundation-for-artificial-intelligence-and-machine-learning/.

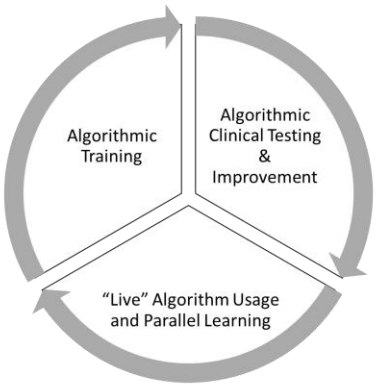
[27] Ronald Summers & Andrew Murphy et al., *Imaging Data Sets (Artificial Intelligence)*, RADIOPAEDIA (last visited Apr. 30, 2021), https://radiopaedia.org/articles/imaging-data-sets-artificial-intelligence.

[28] D. Douglas Miller & Eric W. Brown, *How Cognitive Machines Can Augment Medical Imaging*, 212 AJR 9, 10-11 (Jan. 2019), https://www.ajronline.org/doi/pdf/10.2214/AJR.18.19914.

probability before symptoms manifest by using alternative patient data. It may also be able to analyze images more effectively, especially for "borderline" cases.

Organizations do not only use data to create algorithms; data are needed both in continuous supply and long-term for purposes of algorithmic learning or tuning.[29] Further, the system-generated algorithm is continuously adapting as new data are generated and added, modifying the algorithm.[30] See Figure 1, which describes when in an AI's lifecycle data are used.

Figure 1: AI Data Lifecycle[31]



[29] Unsupervised machine learning and optimally functioning neural networks become more effective over time as they learn. To avoid potential issues related to safety and efficacy, some medical devices are "locking" algorithms after they have been trained on clinical data, to avoid resubmission for U.S. Food & Drug Administration review processes. However, AI's potential is tied to self-learning, whether self-learning on a separate system for purposes of later FDA resubmission as a "material change," or an AI that continuously changes in its live version.

[30] John Schreifer, *Six Questions About Machine Learning and AI for Warehouse Management*, LUCAS (Apr. 14, 2021), https://www.lucasware.com/six-questions-about-machine-learning-for-warehouse-management/.

[31] *See, e.g.*, Pratik Shah et al., *Artificial Intelligence and Machine Learning in Clinical Development: a Translational Perspective*, NPJ DIGITAL MED. 1, 2-3 (describing use of machine learning in clinical and real-world applications after initial algorithmic development, especially for start-ups and to meet FDA requirements). Artificial intelligence in healthcare is still reviewed using existing U.S. Food & Drug Administration processes, which require the submission of research data and outcomes with an Investigational Device Exception to perform clinical research. After clinical research has concluded, the device must be submitted for approval in a Pre-Market Approval, 510(k), or De Novo classification process. After this time, whether or

Algorithmic training uses data for purposes of developing the AI algorithm. In this stage, data are stored in a big data set, upon which an algorithm runs for a predetermined period of time.[32] After an algorithm has undergone initial testing, if it is intended to be used for patients, it may be tested in clinical testing protocol, wherein it will be tested for safety and efficacy.[33] After the algorithm is approved, it either may be configured to adapt on the fly, *dynamically*, where data are continuously added to the big data set to improve the algorithm or may collect data and create a shadow algorithm to be released a later time.[34] Regardless of the model used, data are essential to every part of the AI algorithmic lifecycle, from creation to improvement.

B. AI Safety and Discrimination Concerns

Neural-networked AI systems, which are responsible for the most complex of diagnostic medical tasks, have the ability to intake these data across thousands of patients, apply advanced inferences and weightings across data points, and predict certain outcomes, such as whether an individual likely has breast cancer.[35] The data set needs to be very robust because comprehensive analytics depend on big data, including for use in AI.[36]

---

not an AI-enabled medical device is updated (and dynamically "learning") is based on whether the AI algorithms are locked upon submission to the FDA. *See Artificial Intelligence and Machine Learning in Software as a Medical Device.* U.S. FOOD & DRUG ADMIN. (Jan. 12, 2021), https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device; *see also* Charlotte A. Tschider, *Medical Device Artificial Intelligence: The New Tort Frontier*, 46 BYU L. REV. 1551, 1572 (2021).

[32] Adrian Yijie Xu, *How to Train Your Model: A Novice's Guide to Selecting the Correct Machine Learning Algorithm for Your Problem*, GRADIENT CRESCENT (Mar. 6, 2019), https://medium.com/gradientcrescent/the-right-tool-for-the-job-a-novice-guide-to-selecting-the-correct-machine-learning-algorithm-for-60613c7f7b0b (describing model selection and running code on a data set).

[33] *See* Tschider, *supra* note 31, at 1581.

[34] *Id.* at 1572-73.

[35] *See* Tschider, *supra* note 11, at 692-93; Farhad Malik, *Neural Network Bias and Weights*, FINTECHEXPLAINED (May 18, 2019), https://medium.com/fintechexplained/neural-networks-bias-and-weights-10b53e6285da; Christoph I Lee & Joann G. Elmore, *Artificial Intelligence for Breast Cancer Imaging: The New Frontier?*, 111 J. NAT'L. CANCER INST. 875-76 (2019), doi: 10.1093/jnci/djy223.

[36] Wendy Netter Epstein & Charlotte Tschider, *We Need to Do More with Hospitals 'Data, But There Are Better Ways*, HARV. L. PETRIE-FLOM CTR. BILL OF HEALTH BLOG (July 7, 2021), https://blog.petrieflom.law.harvard.edu/2021/07/07/hospital-data-big-tech/. The crucial use of big data for AI application is described as *data essentialism.*

Data volume and quality have a significant effect on the reliability, safety, and fairness of an AI system because the system encodes data relationships and weightings in its algorithms.[37] Of course, the volume and variety of data needed to make a system reliable and avoid these issues is highly specific to a given system's chief goals.[38] Similar to a human mind that makes assumptions without enough data inputs results in cognitive biases,[39] algorithms without sufficient data volume or without quality data are more likely to make algorithmic assumptions,[40] which may produce dangerous or discriminatory results.[41]

Training data, which data scientists use to create and refine initial AI algorithms, may begin from initial inferences based on data that include discriminatory practices.[42] By training AI on such data, the AI itself may codify discrimination and perpetuate it through later decisions, except with the guise of technical objectivity.[43] Even when such data sets do not explicitly include sensitive individually identifiable data elements, individuals may still experience

---

[37] *See* Tschider, *supra* note 11, at 693.

[38] Bernard Marr, *Why AI Would Be Nothing Without Big Data*, FORBES (June 9, 2017, 12:29 AM), https://www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/#1d7677c54f6d; Joshua New, *AI Needs Better Data, Not Just More Data*, CTR. FOR DATA INNOVATION (Mar. 20, 2019), https://www.datainnovation.org/2019/03/ai-needs-better-data-not-just-more-data/.

[39] *See* Jessica Stillman, *6 Cognitive Biases That Are Messing Up Your Decision Making*, INC. (Nov. 22, 2016), https://www.inc.com/jessica-stillman/6-cognitive-biases-that-are-messing-up-your-decision-making.html.

[40] Matthew Stewart, *The Limitations of Machine Learning*, TOWARDS DATA SCI. (July 29, 2019), https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6.

[41] *See* Charlotte A. Tschider, *Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age*, 96 DEN. L. REV. 87, 98-100 (2018).

[42] Sandra Wachter & Brent Middelstadt, *A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 543-48 (2019) (proposing a right to know and rectify inferences).

[43] Douglas McNair & W. Nicholson Price, II, *Health Care AI: Law, Regulation, and Policy*, *in* ARTIFICIAL INTELLIGENCE IN HEALTHCARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL 181,184 (Michael Matheny et al. eds., NAT'L. ACAD. MED 2019) [hereinafter, ARTIFICIAL INTELLIGENCE IN HEALTHCARE]. Paul Teich, *Artificial Intelligence Can Reinforce Bias, Cloud Giants Announce Tools for AI Fairness*, FORBES (Sept. 24, 2018, 6:00 AM), https://www.forbes.com/sites/paulteich/2018/09/24/artificial-intelligence-can-reinforce-bias-cloud-giants-announce-tools-for-ai-fairness/#bd6835e9d21f.

discrimination by proxy,[44] simply because data fed to the algorithmic is inaccurate or unfair.[45]

Although the need for data to avoid issues related to safety, efficacy, and unfairness may seem obvious, there are additional nuances that strengthen the case for more, better, and varied data sets. One significant issue in the diagnostic realm involves transferring diagnostic and treatment AI between differentiated contextual environments.

For AI systems, two of the most crucial choices an AI designer must make are 1) the training data set and 2) the mechanisms for immediate feedback and correction.[46] If an AI system trains on data from hospitals with a high degree of resources, such as the newest technologies and the most highly trained practitioners, the model the AI system creates will be oriented towards high-resource use and may not be as effective as one trained on low-resource environments.[47] This means that training data should be representative for the population where the AI might be used,[48] similar to how clinical trials for certain populations often are the populations to whom certain drugs may be marketed.

In different contextual environments and with highly differentiated equipment, facilities, and patient populations, the efficacy of an algorithm and its attendant recommended treatments might be less.[49] For example, an algorithm that has been developed using data from patients with access to top hospitals and the best specialists may not include data from socially or financially

---

[44] *See generally* Anya Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257 (2020) (describing the likelihood of discrimination when big data and highly powerful AI systems can result in discriminatory application of decisions to protected groups).

[45] Derek A. Haas et al., *3 Myths About Machine Learning in Health Care*, HARV. BUS. REV. (Nov. 13, 2019), https://hbr.org/2019/11/3-myths-about-machine-learning-in-health-care.

[46] Challen et al., *Artificial Intelligence, Bias and Clinical Safety*, 28 BMJ QUALITY & SAFETY 238 (2018), available at: http://dx.doi.org/10.1136/bmjqs-2018-008370.

[47] *See* W. Nicholson Price II, *Medical AI and Contextual Bias*, 33 HARV. J. L. & TECH. 65, 96-100 (2019).

[48] *Id.* at 115.

[49] *Id.* at 96-100 (describing in great detail the potential risks associated with applying a trained algorithm within new and different contexts, and suggesting issues related to contextual change could be avoided through collecting data representative of these contexts and responsibly training algorithms based on these data).

disadvantaged patients, many of which are highly vulnerable due to certain risk factors or co-morbidities.[50]

Treatment recommendation following diagnosis may also be tuned to high-resourced facilities.[51] This seems to suggest that differentiated data collection requires more comprehensive and contextual data collection to develop contextually extensible AI algorithms. Moreover, contextual data collection is necessary to address safety and efficacy concerns, as well as the potential for unfairness through application of substandard AI to certain low-resourced communities or protected classes.[52]

AI does not simply require data, but it requires that data are available from a variety of data populations and collection contexts.[53] Data must be quality, well-organized, appropriately labeled, and reliably sourced to ensure AI systems perform safely, efficaciously, and fairly.[54] If AI data scientists cannot collect or use *good* data, patients will not reap the potential benefits AI has the potential to provide and society will not see the economic benefits of AI investments.

C. Health AI Case Studies

The healthcare marketplace has transformed in recent years, positioning itself for safer, more reliable, non-invasive, and cost-effective solutions through artificial intelligence. AI healthcare, therefore, has dominated AI investment—$4 billion in 2019,[55]

[50] *Id.* at 96-97.

[51] *Id.*

[52] Michael Matheny et. al., *Artificial Intelligence in Healthcare: Hope not Hype, Promise not Peril*, *in* ARTIFICIAL INTELLIGENCE IN HEALTHCARE 217-18.

[53] Indeed, broad, representative data collection is at the center of data ethics commitments. *See* Luciano Floridi & Mariarosaria Taddeo, *What Is Data Ethics?*, 374 PHILO. TRANS. ROYAL SOC'Y A: MATHEMATICAL, PHYS. & ENG. SCI. 1, 3 (2016), available at: https://doi.org/10.1098/rsta.2016.0360.

[54] Paramita (Guha) Ghosh, *Challenges of Quality Data in the AI Ecosystem*, DATAVERSITY (Nov. 12, 2019), https://www.dataversity.net/challenges-of-data-quality-in-the-ai-ecosystem/

[55] Heather Landi, *Investors Poured $4B into Healthcare AI Startups in 2019*, FIERCE HEALTHCARE (Jan. 22, 2020, 5:55 PM), https://www.fiercehealthcare.com/tech/investors-poured-4b-into-healthcare-ai-startups-2019.

positioned to reach $6.6 billion by 2021.[56] In fact, after new revelations of the value of low-touch and remote medical care during the COVID-19 pandemic, the healthcare AI marketplace is anticipated to be valued at $51.3 billion by 2027.[57] If investment follows these market predictions, some analysts have predicted AI will save $150 billion annually, by 2026.[58] Ultimately, AI may provide more efficient and effective healthcare solutions, but it also could dramatically reduce healthcare costs. That is, if these solutions are safe, effective, and fair.

Healthcare AI technologies include operational support, diagnostics, and kinetics.[59] The more advanced the category and associated AI functions, the more dependent on data the algorithm becomes. Except for self-driving cars and automated industrial applications, such as electrical grid operation, some of the most complex AI applications are in the healthcare sector, which means that health data are essential ingredients for AI health applications.

Individually identifiable health data are increasingly used for a variety of important purposes, and it is not always possible to render

---

[56] *AI and Healthcare: A Giant Opportunity*, FORBES (Feb. 11, 2019, 12:47 PM), https://www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-a-giant-opportunity/#3d308f284c68.

[57] METICULOUS RESEARCH, HEALTHCARE ARTIFICIAL INTELLIGENCE (AI) MARKET WORTH $51.3 BILLION BY 2027 – EXCLUSIVE REPORT COVERING PRE AND POST COVID-19 MARKET ANALYSIS BY METICULOUS RESEARCH, CISION PR NEWSWIRE (Aug. 27, 2020, 9:07 AM), https://www.prnewswire.com/news-releases/healthcare-artificial-intelligence-ai-market-worth-51-3-billion-by-2027—exclusive-report-covering-pre-and-post-covid-19-market-analysis-by-meticulous-research-301119739.html.

[58] *Future AI Opportunities for Improving Care Delivery, Cost, and Efficacy*, HEALTH IT ANALYTICS (July 29, 2019), https://healthitanalytics.com/news/future-ai-opportunities-for-improving-care-delivery-cost-and-efficacy.

[59] PHILIPS, *USING AI TO MEET OPERATIONAL, CLINICAL GOALS* 5-6 (Feb. 2018), https://www.philips.com/c-dam/b2bhc/master/seamless-care/Q1-HIM/AI_Updated_02032018.pdf. AI-enabled medical devices that introduce some physical functionality are "kinetic," in that they have a physical, rather than mental manifestation of AI automation. For example, decisional systems that identify congestive heart failure function differently than an artificial pancreas, which physically releases insulin to the body. *See* Darrell M. West & John R. Allen, *How Artificial Intelligence Is Transforming the World*, BROOKINGS (Apr. 24, 2018); Nimri et al., *Insulin Dose Optimization Using an Automated Artificial Intelligence-Based Decision Support System in Youths with Type 1 Diabetes*, 26 NAT MED. 1380, 1380-81 (Sept. 2020), available at: https://www.nature.com/articles/s41591-020-1045-7. Healthcare robots similarly occupy "kinetic" functionality based on AI, which may include robotic-assisted surgery. INTEL, *Robotics in Healthcare to Improve Patient Outcomes*, https://www.intel.com/content/www/us/en/healthcare-it/robotics-in-healthcare.html (last accessed: July 28, 2021).

such data non-identifiable yet retain its usefulness. The following cases illustrate a variety of data uses that specifically demonstrate both a) the range of uses for individually identifiable health data and b) the necessity of such data to fulfill legitimate health goals.

1. Value-Based Healthcare

Value-based healthcare has been a central focus of healthcare development for private insurers and government health plans. Under the Health Information Technology for Clinical Health (HITECH) Act of 2009, healthcare organizations had been directed to make health data electronic, in the form of electronic medical records (eMR) or electronic health records (eHR).[60] In 2010, with the passage of the Patient Protection and Affordable Care Act (ACA) that created Accountable Care Organizations (ACOs), the United States government has invested heavily in data development and exchange.[61]

Most recently, in 2015, the Medicare Access and CHIP Authorization Act consolidated a mix of quality and efficiency reporting indicators to simplify evaluation of organization and physician performance: Merit-Based Incentive Programs (MIPs) and Alternative Payment Models (AMPs), both of which rely heavily on submission of detailed patient encounter information.[62]

Although typically AI is not discussed in relation to value-based healthcare, automation AI is a natural extension of value-based healthcare goals and is positioned to save as much as $71 billion annually through virtual nursing assistants, administrative workflow assistance, fraud detection, and dosage error reduction.[63] Connected machines enabling remote monitoring could save an additional $14 billion per year.[64]

---

[60] 42 U.S.C. § 300jj (2020).

[61] Taylor Burke, *Accountable Care Organizations*, 126 PUB. HEALTH REPS. 875, 875-76 (2011).

[62] Niam Yaraghi, *MACRA Proposed Rule Creates More Problems Than It Solves*, HEALTH AFFS. BLOG (Oct. 12, 2016), https://www.healthaffairs.org/do/10.1377/hblog20161012.057043/full/ (providing that organizations under MACRA may submit PHI without patient authorization and that AI has the potential to better identify opportunities for improvement to MACRA scores through enhanced quality or efficiency).

[63] *See Future AI Opportunities*, *supra* note 58.

[64] *Id*.

AI will likely be used in conjunction with connected machines: sensors and other Internet of Health Things technologies used for remote monitoring.[65] Increasingly, AI and other advanced technologies have the potential to be used for data analysis and to augment value-based solutions, or *Value = (Quality + Outcomes)/Cost*.[66] Nicolas P. Terry has described this model as the "New Iron Triangle," wherein Terry advocates for a more advanced and nuanced policy model that expands the original focus on access, quality, and cost containment to account for technological development.[67] With the addition of AI natural language processing (NLP) approaches, some of the most notorious sources of cost-containment issues, such as medical coding inaccuracies, healthcare providers have the opportunity to dramatically increase efficiency while simultaneously reducing expensive mistakes.[68]

Terry's model is exemplified with current Accountable Care Organization initiatives. For example, CareAngel has positioned itself as the first AI and Voice-Powered Virtual Nurse Assistant, which is intended to reduce clinic visits and enable "aging in place" for older adults.[69] *See* Figure 2, below, which illustrates the technology flow from AI engagement with a patient to data aggregation and reporting to healthcare providers.

---

[65] Nicolas P. Terry, *Appification, AI, and Healthcare's New Iron Triangle*, 20 J. HEALTHCARE L. & POL'Y 117, 129-31 (2018).

[66] *Id.* at 124.

[67] *Id.* at 120-21.

[68] Elliot B. Sloane & Ricardo J. Silva, *Artificial Intelligence in Medical Devices and Clinical Decision Support Systems, in* CLINICAL ENGINEERING HANDBOOK 556, 560 (Ernesto Iadanza ed., 2d ed. 2020).

[69] *ACOs Use Virtual Nurse Assistants to Improve Patient Engagement and Outcomes*, CAREANGEL (Nov. 16, 2018, 3:11 PM), https://www.careangel.com/blog/what-it-means-to-be-an-accountable-care-organization (describing CareAngel's role in ACOs); *The Power of AI and Voice*, CAREANGEL, https://www.careangel.com/ai-and-voice-powered-virtual-nurse-assistant (last visited Mar. 14, 2021).

Figure 2: CareAngel AI-Enabled Virtual Assistant[70]

A necessary function of the virtual assistant is interacting with the patient; in fact, one of the major value-add functions of the CareAngel is to report individually identifiable information to a system, which determines the risk to the individual based on advanced AI-enabled algorithms and notifies a patient's care team when indicators demonstrate high risk to the patient.[71] CareAngel claims that this functionality will result in 24 times existing clinical capacity and saves three to four hours of clinician time per day.[72]

Without AI, CareAngel cannot fulfill its value proposition. If AI using NLP did not direct the virtual assistant role, time savings would not be realized. Without advanced machine-learning analytics, presumably risk identification and reporting would be less accurate. Similarly, data collected must be identifiable in nature: without knowing who the patient is, it is impossible to fulfill CareAngel's chief aims.

2. *Healthcare Diagnostic Applications*

AI-enabled healthcare diagnostic applications are positioned to dramatically improve the accuracy, repeatability, and transferability of expert knowledge, or *democratizing* medicine.[73] These applications are also positioned to save as much as $8 billion per year, though

---

70 *The Power of AI and Voice*, *supra* note 69.

71 *Id.*

72 *Id.*

73 W. Nicholson Price II, *Artificial Intelligence in the Medical System: Four Roles for Potential Transformation*, 18 YALE J. HEALTH POL'Y, L. & ETHICS 122, 126-27; 21 YALE J. L. & TECH. 122, 126-27 (2019).

greater accuracy is the most persuasive benefit.[74] Indeed, precision medicine offers the most potential for both diagnosis and treatment protocol for serious diseases that often have a high mortality rate if diagnosed incorrectly or too late.[75] AI-enabled healthcare diagnostic applications include data from patient intake to lab results, diagnosis, and expert recommendations for treatment.[76] Depending on the type of diagnostic applications, additional data could include imaging data from radiological or other imaging systems, blood test results, and a variety of other data depending on the type of disease or diagnosis.

Diagnostic algorithms now apply to a wide variety of diagnostic applications. Cancer diagnostics have been a primary focus for AI, leveraging partnerships with big technology companies, and populating big data infrastructure with diagnostic data, imaging, and treatment data from some of the most successful oncologists.[77] One example of these algorithms is QuantX, the first breast cancer imaging diagnostic tool to be cleared by the United States Food and Drug Administration (FDA),[78] and the tool is used by radiologists to improve the accuracy of diagnoses, rather than replacing these specialists.[79] In a clinical study, QuantX resulted in a 39 percent reduction of overlooked breast cancers and a 20 percent diagnostic improvement.[80]

Imaging algorithmic training usually requires actual diagnostic data, such as radiological images, as well as individually identifiable information from the individual.[81] QuantX, for example, analyzes MR

---

74 *Future AI Opportunities, supra* note 58.

75 Thomas Davenport & Ravi Kalakota, *The Potential for Artificial Intelligence in Healthcare*, 6 FUTURE HEALTHCARE J. 94, 94-95 (2019).

76 *See* Price, *supra* note 73, at 127 (describing the functionality of the IDx-DR system).

77 Neil Savage, *Another Set of Eyes for Cancer Diagnostics*, 579 NATURE 14, 15 (2020).

78 Melissa Locker, *This AI Breast Cancer Diagnostic Tool is the First to Get FDA Clearance*, FAST CO. (July 17, 2019), https://www.fastcompany.com/90377791/quantx-is-first-ai-breast-cancer-diagnostic-tool-cleared-by-fda.

79 Jack Carfagno, *5 FDA Approved Uses of AI in Healthcare*, DOCWIRENEWS (July 18, 2019), https://www.docwirenews.com/docwire-pick/future-of-medicine-picks/fda-approved-uses-of-ai-in-healthcare/.

80 *Id.*

81 Jenifer Sunrise Winter & Elizabeth Davidson, *Governance of Artificial Intelligence and Personal Health Information*, 21 DIGITAL POL'Y, REG. & GOVERNANCE (SPEC. ISSUE) 280 (2019), https://doi.org/10.1108/DPRG-08-2018-0048.

image data, including segments and user-selected regions of interest.[82] However, it also requires information about positive and negative tests and information about the individuals.[83] It is likely that at least some of this information is individually identifiable. Part of the data set used for training the QuantX algorithm originated from a previous but similar study.[84]

3. AI-Enabled Robotic Surgery

AI has also become the new foundation for a variety of Internet of Things devices as well as devices connected to internal hospital networks. AI has optimized how devices function, enhanced human-computer interactions, and transformed previously human-only activities. Medical devices stand to benefit most from AI as such devices supplant nursing care, clinic visits, and even surgery. AI is increasingly being used for disease management as well as treatment.[85]

Robotic surgery is one physical treatment technology that has the potential to reduce recovery time while simultaneously reducing surgical errors.[86] Robotic surgery also has the potential to reduce surgical costs by up to 29 percent, as much as a $40 billion in savings annually.[87] In robotic surgery, a patient usually receives care from a primary physician, who transfers health records about the individual to a specialized surgeon who is trained in robotic surgery.

As part of any surgical process involving the use of a surgical robot, information about the individual must be shared with a medical

---

[82] FDA, DEN170022, EVALUATION OF AUTOMATIC CLASS III DESIGNATION FOR QUANTX DECISION SUMMARY 1-2, https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170022.pdf (last visited Oct. 2, 2020).

[83] *Id.*

[84] *Id.* at 15.

[85] *See* Alvin Powell, *AI Revolution in Medicine*, HARV. GAZETTE (Nov. 11, 2020), https://news.harvard.edu/gazette/story/2020/11/risks-and-benefits-of-an-ai-revolution-in-medicine/.

[86] Rafael E. Perez & Steven D. Schwaitzberg, *Robotic Surgery: Finding Value in 2019 and Beyond*, 4 ANNALS LAPAROSCOPIC & ENDOSCOPIC SURGERY 1, 1 (2019).

[87] *Id.* at 5 tbl.1 (citing Zhamak Khorgami et al., *Extra Costs of Robotic Surgery in Minor and Major Surgeries: An Analysis of National Inpatient Sample*, 225 J. AM. C. SURGEONS e86 (2017)).

device manufacturer prior to surgery.[88] Trainers or device supervisors may be present in the surgery to ensure no issues occur during the surgical process.[89] The surgical robot, to improve its movements and detection sensors both during the surgery and after surgery, must also collect data during surgery which will likely be valuable outside the surgical environment.[90]

For example, a kinetic surgical movement that worked most effectively on a 33-year-old with a genetic predisposition for cardiomyopathy and Type-1 diabetic comorbidity may not be as effective on a 92-year-old with kidney disease, even if both patients are receiving the same surgery.[91] In aggregate, understanding effectiveness over time and across facilities may also be useful, which could require retention of facility and treatment date data.

Ultimately, high-volume data stores are an essential ingredient for all AI solutions and models, to varying extents.[92] Without easy access to data, algorithms will likely be less efficacious, and for some applications, may be downright dangerous. Quality measures may be inexact, leading to strategies that reduce available resources to individuals who need these. Diagnostic tools may be ineffective, resulting in misdiagnosis or unnecessary surgical interventions. Robotic surgeries may be less exact, leading to devastating surgical outcomes.

Although individually identifiable health data has the potential to be misused, especially in aggregate, the countervailing value of data for safety and efficacy purposes may encourage regulators to balance these interests more effectively.

---

[88] Ford et al., *Rightsizing the Role of Medical Device Reps in the OR*, THE SOURCE (2021), https://healthtrustpg.com/thesource/cqo/rightsizing-role-medical-device-reps/.

[89] *Id.*

[90] Claire Jarvis, *Robots are Surging in Popularity. So Will Their Data*, UNDARK (Aug. 15, 2019), https://undark.org/2019/08/15/surgical-robots-are-suring-in-popularity/.

[91] Robotic surgery, like most AI applications, are by their nature personalized. *See* James Warner, *Thanks to AI, Medical Treatments Are Becoming More Personalized*, TNW (Dec. 11, 2019), https://thenextweb.com/syndication/2019/12/11/thanks-to-ai-medical-treatments-are-becoming-more-personalized/.

[92] Sabyasachi Dash et al., *Big Data in Healthcare: Management, Analysis and Future Prospects*, J. BIG DATA 1, 3 (2019).

II. HEALTH DATA REGULATION AND ITS LIMITATIONS

While data privacy laws aim to promote important goals, including reducing the potential for abuse through data overcollection and use, current health data regulations in the United States were not designed for big data and AI data use. The current regulatory model, as currently interpreted and employed, does not adequately promote privacy interests or enable effective data use for important public benefits. Data privacy laws should protect individuals while also promoting responsible data use.

A. Privacy "Risk" in Health Data

Health data *are* exceptional.[93] Health data are unique in that they exist because our bodies produce and perform to create them. However, technology is required to collect, use, store, transfer, retain, and ultimately delete data produced by our bodies. Some technologies continuously siphon data from our bodies through pervasive connectivity.[94] These data, then, become *dis*embodied, separating the person from their data when such data are processed and stored.

Health data are similarly exceptional because their unauthorized use poses high risk: 1) misuse by medical professionals or staff; 2) data stolen due to security vulnerabilities or poor security practices; and 3) other concrete impacts resulting from improper data use or sharing, such as employment discrimination, insurance coverage issues, legal issues, or personal impacts. Because these risks are the consequence of unauthorized data use and poor data security practices, they are considered consequentialist.[95]

Health privacy laws codify responses to consequentialist risks in two ways: 1) sectoral privacy laws establish key regulatory responsibilities for defined organizations to ensure they are aware of

---

[93] *See generally* Nicolas P. Terry, *Big Data Proxies and Health Privacy Exceptionalism*, 24 HEALTH MATRIX 65, 66 (2014) (describing the nature of big data in health, from direct identifiers to proxies for identifiability and the attendant regulations at the federal and state levels that and strategies for resolving existing issues).

[94] Andrea M. Matwyshyn, *The Internet of Bodies*, 61 WM. & MARY L. REV. 77, 81, 82, 116 (2019) (exploring the realities of security issues related to pervasive connectivity, what Matwyshyn calls "the gratuitous Internet problem").

[95] *See* Price & Cohen, *supra* note 5, at 39.

their responsibilities,[96] and 2) these laws specify binding requirements defined organizations must implement, usually preventative procedural activities, or management-level controls.[97] For example, organizations must verify the identity of a health data requestor prior to disclosing identifiable information and organizations must engage in appropriate management of legitimate access to health records and access termination.[98] Laws like the Health Insurance Portability and Accountability Act (HIPAA) require compliance with the Security Rule to reduce the probability of data misuse and data breach.

In addition to these consequentialist risks, which primarily focus on tangible negative consequences, health data privacy risks also consist of deontological risks.[99] Deontological risks are risks that exist in and of themselves and do not depend on negative consequences.[100] For example, if an organization collects an individual's data without their knowledge, this could harm the individual's autonomy, or ability to make a choice, even though no specific negative consequences, such as data theft, result. Or perhaps a data breach occurs, exposing an individual's data to a large group, but no direct negative harm results, such as unfair treatment in the workplace.[101] That individual's privacy was compromised, even if the individual was not directly harmed.

Unfairness, as in exclusion or differential treatment, may fall under this risk category, as these risks are injurious in and of themselves to autonomy even if no further injury occurs in a consequential manner.[102] Health privacy laws address deontological concerns through data collection and use restrictions and limitations on data sharing. For example, under many laws and broad Federal

---

[96] 45 C.F.R. § 160.103. Covered entity and business associate are the two relevant roles specified under the Health Insurance Portability and Accountability Act of 1996.

[97] David Thaw, *The Efficacy of Cybersecurity Regulation*, 30 GA. ST. L. REV. 287, 324-25 (2014).

[98] *See* 45 C.F.R. §§ 164.514(h),164.308(a)(4)(i). The HIPAA Security Rule, for example, includes Administrative safeguards that are primarily procedural. *See* Charlotte A. Tschider, 26 ANNALS HEALTH L. 1, 14 (2017) (describing the substantial focus of the Security Rule on administrative specifications rather than technical requirements).

[99] *See* Price & Cohen, *supra* note 5, at 39.

[100] *Id*. at 3-4.

[101] *Id.*

[102] *See* Mason Marks, *Artificial Intelligence-Based Suicide Prevention,* 18 YALE J. HEALTH POL'Y LAW & ETHICS 98, 117-18, 21 YALE J. L & TECH. 98, 117-18 (2019) (describing issues related to differential treatment as 'autonomy risks').

Trade Commission (FTC) authority, organizations must notify individuals about their data handling practices, including potential uses, and categories of third parties receiving data.[103] Notably, privacy laws and regulatory health device clearance processes do not address deontological concerns involving differential treatment, unfairness, and discrimination due to automated decisioning.[104]

At the federal level, the Health Insurance Portability and Accountability Act (HIPAA) codifies requirements that address both consequentialist and deontological concerns. For example, HIPAA requires data minimization for data collection and use consistent with treatment, payment, and healthcare operations (required healthcare activities, where data are essential for care and claims processing).[105] The U.S. Department of Health and Human Services (HHS) has also codified a de-identification "safe harbor," which incentivizes reducing identifiability of data sets by permitting broad use when typically sensitive health data elements are not used.[106] These collectively reduce the potential for negative consequences and independently set limits on data collection and use.

On the surface, this combination of approaches blends a practical approach to reducing risk while reinforcing privacy civil rights through individual choice.[107] Unfortunately, the current model

---

[103] U.S. DEP'T OF HEALTH & HUMAN SERVS., SUMMARY OF THE HIPAA PRIVACY RULE 9 (2013), https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[104] *See* Mason Marks, *Algorithmic Disability Discrimination*, DISABILITY, HEALTH, LAW & BIOETHICS 242-44, 246 (Cambridge Univ. Press 2020), available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3338209; Charlotte A. Tschider, *Medical Device Artificial Intelligence: The New Tort Frontier*, 46 BYU L. REV. 1551, 1571 (2021) (describing the reliance of the FDA on process-based, ineffectual solutions that do not prevent discrimination).

[105] SUMMARY OF THE HIPAA PRIVACY RULE, *supra* note 103, at 4.

[106] U.S. DEPT. OF HEALTH & HUMAN SERVS., GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA) PRIVACY RULE (Nov. 6, 2015), https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

[107] However, choice is not necessarily achieved by these models. *See* Helen Nissenbaum, *Privacy as Contextual Inquiry*, 79 WASH. L. REV. 119, 130 (2004); Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1894 (2013); Neil Richards & Woodrow Hartzog, *The Pathologies of Digital Consent,* 96 WASH. L. REV. 1461, 1492; Charlotte A. Tschider, *The Consent Myth: Improving Choice for Patients of the Future*, 96 WASH. UNIV. L. REV. 1505, 1528 (2019).

complicates legitimate data use necessary for complex technology and, by extension, AI, while being highly ineffective for its autonomy risk-mitigating and civil rights goals.[108] Four key areas where privacy laws are essentially ineffective are notice, consent, consent revocation (or notice and consent), and de-identification.[109]

B. Regulatory Applicability

Medical devices, including medical software, medical diagnostic or treatment applications, and AI-enabled wearable, implanted, or standalone medical devices, are generally regulated for safety and efficacy by the FDA.[110] Although the FDA partially regulates cybersecurity issues, it does not directly regulate for privacy.[111]

---

[108] *See* Terry, *supra* note 93, at 97 (quoting Viktor Mayer-Schönberger and Kenneth Cukier's concerns over big data).

[109] *Id*. Mayer-Schönberger and Cukier describe this as *anonymization*, but within the United States regulatory system the terminology is de-identification. To be sure, these standards are distinctly different, especially when one reviews European Union laws and guidance on the topic. In the main, de-identification and anonymization goals are the same: to strip identifying characteristics, so that the rendered data set does not identify an individual to a high probability. *See generally* Tschider, *supra* note 41 (describing these topics and their limitations for purposes of privacy, security, and anti-discrimination goals). It should be mentioned that although this paper does not directly address discrimination and other autonomy risks, I position solving these issues via broad disclosure of processes by which AI are created while hosting live AI to promote AI system testing by competitors. *See* Tschider, *supra* note 11, at 715. *C.f.* W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775 (2021) (proposing broad data disclosure and non-intuitive explanation to promote innovation, which could improve exploration of algorithmic efficacy).

[110] It should be noted that security requirements applicable in this area also do not follow historical paths for *reasonable* security, largely due to both the big data infrastructure and unusual characteristics of AI technologies. For a more comprehensive discussion of security issues and potential solutions, *see* Charlotte Tschider, *Deus ex Machina: Regulating Cybersecurity and Artificial Intelligence for Patients of the Future*, 5 SAVANNAH L. REV. 177, 192 (2018) (describing issues in AI regulation by the FDA specifically related to cybersecurity cyber-kinetic attacks of medical devices).

[111] *Id*. It is encouraging that the FDA has begun reviewing cybersecurity for medical devices, as this model will support privacy goals. However, the main concern for cybersecurity from the FDA's perspective is device safety and efficacy, not privacy. General privacy requirements, such as "reasonable privacy," may be included for specific device types, but it is often not reviewed as part of device clearance or registration under 510(k) and PMA processes, even for big data implementations and AI use. This is unsurprising given that even AI has not been fully regulated, except for a discussion paper specific to imaging AI. *See* U.S. FOOD & DRUG. ADMIN., PROPOSED REGULATORY FRAMEWORK FOR MODIFICATIONS TO ARTIFICIAL

Healthcare privacy is enforced under the Department of Health and Human Services 'Office for Civil Rights (OCR), for organizations specifically defined under HIPAA.[112]

Notably, HHS does not regulate most health data uses under HIPAA, which includes many consumer-facing health products,[113] because these organizations do not meet the definitions of "covered entity" or "business associate."[114] The effect of these specific definitions is that many organizations producing AI health technologies are alternatively regulated under the Federal Trade Commission Act's Section 5, which commissions the FTC to investigate and prosecute unfair or deceptive trade practices.[115] The FTC has created guidelines for health data apps and more broadly for privacy through the Fair Information Practices.[116] Although neither are directly legally binding, they do provide some indication to how the FTC interprets privacy practices, which are usually represented in consent decrees.[117]

Although these parallel regulation tracks do address the privacy of health data to some extent, neither appropriately regulates health AI.[118] Ultimately, the calculus of risk and benefit in health AI data is unique, necessitating more nuanced privacy models than the models

---

INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) 7-8 (last visited Oct. 20, 2020). Although comparatively speaking, the FDA does partially regulate cybersecurity for these systems, cybersecurity requirements have been similarly short-sighted.

112 The Office for Civil Rights (OCR), a division of HHS enforces HIPAA, although HHS generally establishes rules associated with the law, as described in Part II.

113 *See* W. Nicholson Price II & Margot E. Kaminski et al., *Shadow Health Records Meet New Data Privacy Laws*, 363 SCI. 448, 448 (Feb. 1, 2019).

114 *See* Tschider, *supra* note 110, at 201-02; *see infra* Part II(D) and accompanying notes.

115 17 C.F.R. § 248.3(a)(1)–(2) (2018).

116 *Mobile Health Apps Interactive Tool*, U.S. FED. TRADE COMM'N (Apr. 2016), https://www.ftc.gov/tips-advice/business-center/guidance/mobile-health-apps-interactive-tool; U.S. FED. TRADE COMM'N [hereinafter, FTC PRIVACY ONLINE], PRIVACY ONLINE: A REPORT TO CONGRESS 7-8 (1998), https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf (describing the FIPs in detail).

117 Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COL. L. REV., 583, 600–04 (2014) (describing the expansion of jurisdiction and FTC enforcement creating a de facto role of the FTC as primary privacy authority).

118 *See generally* Tschider, *supra* note 108 (describing the overreliance of HIPAA and the FTC on the concept of consent).

that have populated privacy laws and practices for nearly two decades.[119]

C. The Health Insurance Portability and Accountability Act

Although the Health Insurance Portability and Accountability Act (HIPAA) and its most recent update, the Health Information Technology for Economic and Clinical Health Act of 2009 (HITECH),[120] regulate organizations that qualify as a covered entity (healthcare providers, healthcare clearinghouses, and health plans) and their third parties (business associates), the collective legislation has been hailed one of the most comprehensive in the United States.[121]

1. History and Purpose of Health Data Collection under HIPAA

When HIPAA was passed in 1996, Congress was only beginning to understand the potential for a computing future. For example, the House Ways and Means Committee, in reviewing HIPAA's text, specifically noted the need to protect patient privacy, but it was contextualized within healthcare delivery: "Confidentiality—In determining what information is required, the Secretary shall include procedures to assure that the privacy of individuals receiving healthcare services is appropriately protected."[122] Ways and Means also noted the importance of data in transferability to different providers or insurers, or the "portability" in the Health Insurance *Portability* and Accountability Act.[123]

HIPAA, both in name and as described in the Committee's review of the proposed law, was created initially to promote the portability of insurance coverage from one provider to another in an effort to avoid

---

[119] It should be noted that some sectors may have somewhat similar risk profiles, such as the automotive industry for self-driving cars. However, with healthcare technologies, the data collected have the potential to impact individuals more severely due to the highly sensitive and externally valuable nature of individually identifiable health data.

[120] Health Information Technology and Quality, 42 U.S.C.A. § 300jj (West 2020).

[121] Daniel J. Solove, *HIPAA Mighty and Flawed: Regulation Has Wide-Reaching Impact on the Healthcare Industry*, 84 J. AHIMA 30, 30 (2013), http://library.ahima.org/doc?oid=106326#.YEj7SWhKg2w. As Solove describes, HIPAA does not manage data gatekeeping particularly well—while on one hand HIPAA may restrict access, in others it may not permit legitimate and useful access to PHI.

[122] H.R. REP. NO. 104-496(I), at 24 (1996).

[123] *Id.* at 34-35.

job lock, or the risk of losing health insurance continuity when changing jobs.[124] Necessarily, transferred insurance coverage meant transferring identifiable health data, and there would be no need for insurance at all without providing healthcare.

In addition to the transfer of data for purposes of healthcare provisioning and insurance reimbursement, consistency of health data for reimbursement and billing purposes relied on third-party review of medical coding prior to data transfers, as in independent audits.[125] Healthcare clearinghouses review non-standard data and assigned codes to reduce errors and streamline reimbursement and billing processes.[126] It is no surprise that these organizations, in 1996, were defined exclusively as covered entities, as they are essential to healthcare provisioning and reimbursement and within the original Act's primary contemplation.[127]

From its initial beginnings, HIPAA was intended to facilitate data sharing for practices core to provisioning healthcare, and the HITECH Act's update to HIPAA focused on strengthening its protections while simultaneously moving towards digitization of health data.[128] HHS describes HITECH as "promot[ing] the adoption and meaningful use of health information technology."[129]

---

[124] *Id.* at 1, 280.

[125] PHYSICIANS ADVOCACY INST., INC., MEDICAL AUDITS: WHAT PHYSICIANS NEED TO KNOW (June 2013), https://www.ncmedsoc.org/wp-content/uploads/2013/06/PAIMedicalAudits.pdf.

[126] Standards for Privacy of Individually Identifiable Health Information, 45 C. F. R. pts. 160 & 164 (2002).

[127] Commenters on the proposed 2000 Privacy Rule advocated for expanding the definition to apply to any organization that receives or maintains PHI. However, the reference to such entities under 1173(a)(1) in the 1996 Act seemed to have tied HHS 'hands with respect to the Privacy Rule. *See* U.S. DEP'T OF HEALTH & HUM. SERVS., *Standards for Privacy of Individually Identifiable Health Information. Final Rule Preamble* [hereinafter, HHS Preamble] (Dec. 28, 2000), https://aspe.hhs.gov/report/standards-privacy-individually-identifiable-health-information-final-privacy-rule-preamble. Furthermore, HIPAA was passed as an update to ERISA, which regulates private insurance, and includes, for example, the Consolidated Omnibus Budget Reconciliation Act (COBRA), which provided for continuation of insurance coverage following employment separation. The contextual passage of HIPAA, therefore, reinforced its focus on regulating healthcare and reimbursement related to insurance coverage.

[128] *Main Goals of HITECH: Everything You Need to Know*, RSI SEC. BLOG (Dec. 6, 2019), https://blog.rsisecurity.com/main-goals-of-hitech-everything-you-need-to-know/.

[129] U.S. DEP'T OF HEALTH & HUMAN SERVS., HITECH ACT ENFORCEMENT INTERIM FINAL RULE

Despite this focus on insurance portability, Congress had the foresight to anticipate potential risks attendant to health data digitization. With the advent of computerized technologies, storage and transfer of these data introduced new security risks and complicated existing patient privacy issues. After being unable to draft rules specific to privacy and security themselves, Congress appointed HHS to create the Privacy and Security Rules.[130]

2. *Covered Entities and Their Business Associates*

HIPAA extends from a covered entity to its third-party business associates indirectly through the covered entity's contract with its business associates[131] under a Business Associate Agreement, in the event a business associate is located outside of the United States, and directly when business associates are located within the United States.[132]

An organization may be a covered entity or business associate under HIPAA but not be subject to its requirements if the information collected, used, transferred, or stored is not Protected Health Information (PHI).[133] PHI is individually identifiable health data pertaining to previous or current mental or physical health conditions.[134] Although the definition is broad, information is not PHI when it has been de-identified—stripped of its identifying characteristics—and HIPAA-regulated organizations are not required to comply with HIPAA for properly de-identified data sets.[135] For

(June 16, 2017), https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html.

130 SUMMARY OF THE HIPAA PRIVACY RULE, *supra* note 103.

131 *Id.* A business associate is a named organization that processes, transmits, or stores individually identifiable health data, also known as Protected Health Information, or PHI, under HIPAA.

132 *Id.* Prior to HITECH's passage, HIPAA directly regulated covered entities but not their business associates. The HITECH Act extended HIPAA's obligations directly to business associates within its regulatory purview, specifically U.S.-based business associates.

133 45 C.F.R. § 160.103 (2014).

134 *Id.*

135 45 C.F.R. §§ 164.502(d)(2), 164.514(a) and (b). It should be noted that de-identification does not change the covered entity or business associate status of an entity; rather, it renders the application of the three rules—the Privacy Rule, Security Rule, and Data Breach Notification Rule—moot. Under HIPAA, there is nothing to secure and no rights to protect when identification characteristics have been stripped.

example, using de-identified data for additional purposes will not trigger a requirement to execute individual patient authorizations.

Similarly, organizations that do not fall under the definition of a covered entity (or a non-covered entity organization's business associates) will not be regulated under HIPAA, even if they collect, use, transfer, or store individually identifiable health data that would otherwise meet the definition of PHI. Ultimately, HIPAA-regulated status requires two reinforcing requirements: 1) status as a covered entity or as a covered entity's business associate and 2) data that are considered PHI.

Therefore, organizations that collect, use, process, or store health data but are not covered entities or their business associates and organizations that may be covered entities or business associates but manage or possess de-identified data according to HIPAA's De-identification Safe Harbor standard, are not required to follow HIPAA. Rather, these organizations are subject to the FTC's Section 5 enforcement.[136]

3. Consent and the HIPAA Privacy Rule

The Privacy Rule's discussion crossed two administrations and was subject to separate notice/comment periods, wherein the subject of consent prompted much debate. The Privacy Rule was eventually passed in 2000, after 52,000 public comments, and updated in 2002 at the direction of a new administration.[137]

One of the most significant 2002 updates was making consent to a Notice of Privacy Practices, which covered standard services like treatment, payment, and healthcare operations, a good-faith acknowledgement of receipt rather than an overt requirement for healthcare providers.[138] Despite patients, patient advocates, and physicians generally supporting a consent requirement, health plans,

---

[136] In fact, the enforcement arm of HHS, the Office for Civil Rights (OCR) seems to view the FTC's enforcement abilities as co-extensive with its own, even for practices squarely within HIPAA's ambit. U.S. DEP'T OF HEALTH & HUMAN SERVS. OFF. CIV. RTS, SHARING CONSUMER HEALTH INFORMATION? LOOK TO HIPAA AND THE FTC ACT (Oct. 2016), https://www.hhs.gov/sites/default/files/pdf-0219_sharing-health-info-hippa-ftcact%20508.pdf.

[137] U.S. DEPT. OF HEALTH & HUMAN SERVS, 66 Fed. Reg. 40 (Feb. 28, 2001) ; *see* SUMMARY OF THE HIPAA PRIVACY RULE, *supra* note 103.

[138] SUMMARY OF THE HIPAA PRIVACY RULE, *supra* note 103.

employers, and institutional providers resisted a consent requirement.[139] HHS ultimately sided with institutional commenters but adopted consent for authorization of PHI record release while enabling good-faith acknowledgement for basic healthcare functions:

> We stated our concern that the blanket consents that individuals sign today provide these individuals with neither notice nor control over how their information is to be used. While we retain those concerns, we also understand that for many who participate in the health care system, the acts of providing and obtaining consent represent important values that these parties wish to retain. Many individuals argued that providing consent enhances their control; many advocates argued that the act of consent focuses patient attention on the transaction; and many health care providers argued that obtaining consent is part of ethical behavior.[140]

Ultimately, the 2000 Final Rule included consent for healthcare providers. At the time, it was the understanding of HHS that individually identifiable health data were already the subject of consent-based models at most providers and that patients were familiar with the model of notice and consent due to ethical obligations healthcare providers typically must fulfill.[141] Despite these parallels, the 2002 update to the privacy rule stripped explicit consent for treatment, payment, and healthcare operations purposes. Furthermore, because HIPAA does not preempt more restrictive privacy requirements, states subsequently passed laws requiring explicit consent.

Of course, not all consent in the privacy context is the same. Primary use is data use that is tightly connected to core services provided and uses are tightly related to those services, while secondary use concerns data use outside these core services.[142] For

---

[139] HHS Preamble, *supra* note 127.

[140] *Id.*

[141] This article does not address any potential changes to informed consent processes in AI, although effectiveness in informed consent is essential for purposes of non-privacy risks. For a comprehensive treatment of informed consent, *see* Valerie Gutmann Koch, *Eliminating Liability for Lack of Informed Consent to Medical Treatment*, 53 U. RICH. L. REV. 1211 (2019). For a specific discussion on AI and informed consent, *see* I. Glenn Cohen, *Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?* 108 GEO. L. J. 1425 (2020).

[142] Charlotte A. Tschider, *The Consent Myth: Improving Choice for Patients of the Future*, 96 WASH. L. REV. 1505, 1514 (2018).

some uses, consent is required, for others it is not.[143] For example, if data are used to perform a blood test, this is likely reasonably connected to the reason why a patient is visiting a healthcare provider, and only a good-faith acknowledgement is needed because data use is reasonably expected. However, if data are used to engage in marketing activities, these activities are likely not expected by a patient. But what happens if secondary uses are not expected by a patient but are nevertheless highly beneficial to them and consent through formal authorization is difficult or impossible to facilitate?

A HIPAA Notice of Privacy Practices is provided for treatment, payment, and healthcare operations, which are reasonably predictable activities that are also necessary and crucial to provisioning healthcare, or primary uses.[144] However, HIPAA does contemplate exigent circumstances, such as when providing a Notice of Privacy Practices is not practical or useful.[145] In these situations, a Notice of Privacy Practices is provided when it is reasonably possible to do so, such as when an emergency situation has stabilized or within a reasonable time after a beneficiary has selected their insurance benefits.

A Notice of Privacy Practices requires organizations to disclose information related to their own practices, categories of third parties providing service on behalf of the organization, a covered entity's obligations, the covered entity's contact details, and articulation of a patient's rights with respect to their data.[146] HIPAA requires some specificity with regard to disclosed uses:

> (A) A description, including at least one example, of the types of uses and disclosures that the covered entity is permitted by this subpart to make for each of the following purposes: treatment, payment, and health care operations.
>
> (B) A description of each of the other purposes for which the covered entity is permitted or required by this subpart to use or disclose protected health information without the individual's written authorization. [. . .]
>
> (D) For each purpose described in paragraph (b)(1)(ii)(A) or (B) of this section, the description must include sufficient detail to place

143 *Id.* at 1515.

144 45 C.F.R. § 164.520(a)-(b) (2013).

145 45 C.F.R. § 164.520 (2013).

146 45 C.F.R. § 164.520(b) (2013).

> the individual on notice of the uses and disclosures that are permitted or required by this subpart and other applicable law.
>
> (E) A description of the types of uses and disclosures that require an authorization [. . .] a statement that other uses and disclosures not described in the notice will be made only with the individual's written authorization, and a statement that the individual may revoke an authorization [. . .][147]

Primary use should not surprise a patient: it is core to providing the service so long as the minimum necessary rule applies and data collection does not dramatically exceed what is needed to provide the service.[148] Secondary use, however, is auxiliary, tangential, and unanticipated by the patient.[149] Although it may be connected in some way to primary services provided, data use outside primary treatment, payment, and healthcare operations generally would not be reasonably expected by a patient and usually require greater disclosure and explicit consent.[150] For example, seeing a doctor about a potential respiratory infection and sharing your information for that purpose is not the same as your doctor then submitting these data to a pharmaceutical company developing new nasal decongestant products.

Under HIPAA, no consent is needed for primary use, but secondary use requires an authorization document combined with explicit, written consent and an expiration date or event.[151] In contrast with Notice requirements, an authorization document requires far more detail, including the specific third party to which PHI will be disclosed (if applicable), more specificity in use disclosure, and termination date or activities.[152] Any PHI data use outside treatment, payment, or healthcare operations requires an organization to execute an authorization.[153]

---

147 45 C.F.R. § 164.520(b)(1)(ii) (2013).

148 *See* Tschider, *supra* note 142.

149 *Secondary Use of Your Personal Information*, OFF. INFO. & PRIVACY COMM'R FOR B.C. (Apr. 17, 2018), https://www.oipc.bc.ca/news/secondary-use-of-your-personal-information.

150 HIPAA requires authorization for these uses, and functionally most secondary uses require additional consent be collected. 45 C.F.R §§ 164.508(a)(2)-(4), (b)(5) (2013).

151 *Id.*

152 45 C.F.R § 164.508 (2013).

153 *Id.*

Privacy notices like the Notice of Privacy Practices and the authorization document are designed to assist patients in assessing the potential for health data privacy risks. However, notice and consent only works effectively when patients have a meaningful choice.[154] This means, for example, that patients may select an alternative covered entity and avoid risk posed by the covered entity's data handling practices. It may also mean that a patient may refuse to sign an authorization. However, there are numerous reasons why meaningful choice is not as effective in a healthcare setting, and notice and consent is generally ineffective as primary privacy mechanisms.

4. Identifiability and De-Identification

An often ignored but foundational pre-supposition of HIPAA is that organizations collecting data adhere to the "minimum necessary" rule.[155] Although frequently this rule seems to apply to specified uses as disclosed in the Notice of Privacy Practices and authorization documents, it is derived from confidentiality codes in medicine that apply broadly.[156] The minimum necessary rule properly applies to all actions and activities related to the PHI data ecosystem including collection, use, and disclosure.[157] Despite these clear requirements, organizations may not consider how data sharing, data retention, or identifiability status affects a continuing obligation to follow the minimum necessary rule. After-all, if data are not duplicated through sharing, do not exist at all, or are rendered non-identifiable, there are far fewer risks to patients.[158]

---

154 *See* Tschider, *supra* note 142, at 1519-28.

155 45 C.F.R. §§ 164.502(b), 164.514(d).

156 U.S. DEP'T OF HEALTH & HUMAN SERVS., MINIMUM NECESSARY REQUIREMENT (Apr. 4, 2003), https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/minimum-necessary-requirement/index.html.

157 U.S. DEP'T OF HEALTH & HUMAN SERVS., OFF. FOR CIV. RTS., COLLECTION, USE, AND DISCLOSURE LIMITATION: THE HIPAA PRIVACY RULE AND ELECTRONIC HEALTH INFORMATION EXCHANGE IN A NETWORKED ENVIRONMENT 1, https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/special/healthit/collectionusedisclosure.pdf (last visited Oct. 19, 2020).

158 Mary Branscombe, *Data Deletion: Your Data Strategy's Greatest Defense*, CIO MAG. (July 3, 2019), https://www.cio.com/article/3405129/data-deletion-your-data-strategys-greatest-defense.html.

Without appropriately implementing the minimum necessary rule, consequentialist and deontological risks dramatically increase for patients. After all, simply having more PHI stored in a database than is actually necessary increases the probability that such PHI could be misused unintentionally or deliberately.[159] It also means such data could be subject to a data breach.[160] If third-party relationships result in several transfers of PHI to a wide variety of organizations, some of which could have differing privacy or security practices, the possibility of misuse also increases, especially when such data are identifiable or retained longer than needed.

Although frequently organizations focus on the notice and (where applicable) consent model (disclosure) to procedurally approximate patient choice for data collection, these same organizations often do not exert much diligence after data have been collected. Historically, HHS addressed this issue by creating the De-identification Safe Harbor ('Safe Harbor'), which permitted complete data use (including sharing and selling data) if data organizations collected or managed were de-identified.[161]

The question of identifiability was a point of discussion when HHS solicited public feedback on the Privacy Rule. Specifically, commenters advocated to define PHI as directly individually identifiable personal information, rather than include indirectly identifiable data.[162] Commenters saw value in indirectly identifiable personal information for research purposes, data that do not directly identify an individual person.[163] Usually, directly identifiable data might include a name or other pervasive identifier like a medical

---

159 This is precisely why many risk qualification and quantification approaches for cybersecurity actually include "number of records" as an input into the calculation. *See* Miryam Meir, *The 2 Types of Risk Assessment Methodology*, SEC. SCORECARD BLOG (June 15, 2020), https://securityscorecard.com/blog/types-of-risk-assessment-methodology; Joey Beachum, *Top Under-the-radar Cybersecurity Threats You May Not See Coming*, HUBBARD DECISION RES. (June 17, 2019), https://hubbardresearch.com/category/htma/how-to-measure-anything-in-cybersecurity-risk/.

160 *How to Prevent A Data Breach at Your Business*, INSUREON SMALL BUS. BLOG (2021), https://www.insureon.com/blog/how-to-prevent-a-data-breach-at-your-business.

161 U.S. DEP'T OF HEALTH & HUMAN SERVS., OFF. FOR CIV. RTS., GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (Nov. 6, 2015).

162 *See* HHS Preamble, *supra* note 127.

163 *Id.*

record number. However, indirectly identifiable data are data that may provide information about an individual that is personal in nature but generally is not used to identify an individual.[164]

De-identification processes render individually identifiable data no longer identifiable, within an acceptable level of risk to the individual, and the Safe Harbor offers two options for demonstrating de-identification: 1) removal of 18 identifiers from a data set or 2) expert determination.[165] The predefined 18 identifiers are commonly collected data elements that are pervasive identifiers of an individual.[166] For example, a person's birth date is considered a pervasive identifier, but a numerical age is not unless it is over age 89.[167] Unless a covered entity or business associate independently knows that a data set is identifiable, stripping these 18 identifiers renders a data set de-identified under the Safe Harbor.[168]

Usually reserved for scenarios when a covered entity or business associate cannot effectively strip all 18 identifiers, an alternative path to de-identification is expert determination.[169] Expert determination relies on statistical analysis of a data set to determine whether the risk of reidentification is negligible, rendered by an independent party.[170] The expert determination path may be used when a data element that should be stripped by the Safe Harbor method must be retained for some business reason. For example, the implantation date for an AI-enabled artificial pancreas or the hospital where it was implanted might be tremendously valuable for determining if an AI's algorithmic update caused potential safety issues in that time period. It may also help to determine which hospitals are most effectively working with

---

164 *See* U.S. DEP'T OF HEALTH & HUMAN SERVS., *supra* note 161.

165 45 C.F.R. § 164.514(b).

166 *Id.*

167 *Id.*

168 The independent knowledge caveat to the de-identification Safe Harbor is a curious addition, in that it seems to almost anticipate the use of big data and AI algorithms. Independently knowing that a data set has the ability to identify an individual usually means the data scientist is using that data *to* identify or target an individual or an algorithm can select an individual from their representative data, or *reidentification*. A BA's rights to de-identify or otherwise use data supplied to it, even after de-identification, are usually specified within a data use agreement, which allocates contractual rights with respect to data.

169 *See* U.S. DEP'T OF HEALTH & HUMAN SERVS., *supra* note 161.

170 *Id.*

product, enabling replication of strong processes. Expert determination could demonstrate that although these direct identifiers (date of implantation and implantation location) are retained, the data set overall poses very low risk of reidentification, permitting data to be shared or used without restriction.

The minimum necessary rule combined with the availability of de-identification is perhaps the most valuable privacy contribution under HIPAA for AI. It is the relationship between constraint (minimum necessary) and affordance (de-identification) that most accurately represents the twin challenges of AI: data maximization to create reliable, safe, and accurate AI while minimizing data use to avoid patient privacy harms. Any AI-friendly privacy model must permit data use when it benefits the patient and restrict data use when data retention poses risk without accompanying benefit.

D. The Federal Trade Commission Act

Despite its limited application to these named organizations, HIPAA is still the most comprehensive health data law currently in existence at the federal or state level, codifying reasonably specific yet flexible privacy and security requirements.[171] However, no federal healthcare privacy law applies to data collected broadly outside covered entities and their business associates, such as health app technology providers or medical device manufacturers that are not covered entities and their business associates.[172]

As a result, the FTC and state regulators have stepped in, though their privacy models replicate the limitations of HIPAA such as notice and consent.[173] Notably, while the FTC does enforce Section 5 for privacy and security practices, the inexactitude of *when* the FTC enforces and *what* is enforced is left up to administrative discretion.[174]

---

171 *See* Tschider, *supra* note 2, at 12; Solove, *supra* note 121.

172 *See* Tschider, *supra* note 142, at 1515-16.

173 *Id.* at 1515-17.

174 Through its consent orders, the FTC has established a type of "common law," at least as it pertains to FTC actions for privacy. This body of administrative enforcement has created some degree of predictability as to the FTC's interpretation of unfair or deceptive trade practices under Section 5 for privacy activities, which generally follow the Fair Information Practices (FIPs). *See* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 627 (2011) (illustrating how the FTC has established a body

Much of the specific direction given to organizations collecting, using, processing, or storing individually identifiable, non-HIPAA regulated personal health data, is rendered non-binding, such as the FTC's Fair Information Practices, health app guidance, and consent decrees enforcing these non-binding sources of truth.[175]

A closer look at the Fair Information Practices illustrates a fairly simplistic view of privacy, absent effective context, although to be fair, the FTC's understanding of privacy has come a long way since 1998. Notice and consent, similar to HIPAA, procedurally automates choice without really providing individuals with meaningful choice.[176] Data minimization is not explicitly mentioned, only purpose limitations, wherein data collection and use should be limited to the purposes specified in the privacy notice.[177] Overall, this can be interpreted to mean that, so long as an organization identifies itself, communicates its purpose for collecting data, provides the ability for someone to "opt out," and provides an individual with a list of data it has collected about them when requested, it can collect highly identifiable and sensitive data without restriction and retain it indefinitely.

E. The California Consumer Protection Act

States have begun legislating to address the gap in broad personal information protection. For example, the highly publicized and often criticized California Consumer Protection Act (CCPA) requires more rigorous privacy protections when organizations that receive personal information from California residents or are located in California.[178] Organizations complying with CCPA may opt to extend its protections across the U.S. for ease of overall data and privacy operations

of law that functions in part like the common law). The concept of common law has not necessarily extended to privacy protections through improved security. *See* Justin (Gus) Hurwitz, *Data Security and the FTC's UnCommon Law*, 101 IOWA L. REV. 955, 968-971 (2016) (illustrating why the FTC has not successfully extended its approach to privacy to security, as well). Hurwitz describes how administrative enforcement does not have the same function as the common law in establishing predictive outcomes, as in the common law. *Id*. at 984.

175 *See* FTC PRIVACY ONLINE, *supra* note 116, at 7-11.

176 *See* Charlotte A. Tschider, *The Consent Myth: Improving Choice for Patients of the Future*, 96 WASH. L. REV. 1505, 1516 (2018); FTC PRIVACY ONLINE, *supra* note 116, at 7.

177 *See* FTC PRIVACY ONLINE, *supra* note 116, at 15.

178 CAL. CIV. CODE § 1798.185(a)(1)-(2), (4), (7).

management, and indeed other states have begun to replicate it.[179] Under the CCPA, organizations otherwise subject to HIPAA are excused from its requirements, presumably only for their HIPAA-regulated activities.[180]

CCPA does not perform much better than general FTC guidance, although its requirements are codified and apply across sectors.[181] This approach offers a state-based "catch-all" for organizations that are not regulated under HIPAA, though unfortunately much of CCPA replicates the same issues with notice and consent and anonymization, effectively restricting data usability which may not be beneficial for AI goals.

For example, the CCPA mandates a detailed privacy notice and does not require consent unless personal information will be sold or used for a secondary purpose not disclosed at the time of notice, or if the individual is a minor.[182] Notably, it does not include any explicit data minimization requirement, though it encourages de-identification activities through a "reasonableness" standard for de-identification.[183] This standard prohibits reidentification of the consumer, unlike the HHS de-identification safe harbor, which permits reidentification.

F. Collective Privacy Approaches

Overall, HIPAA, the FTC Fair Information Practices, and the CCPA replicate long-standing privacy strategies, which do not effectively balance AI interests with privacy interests. *See* Table 1 for a comparison between the regulations. In Part III, we will discuss how

---

[179] Kayvan Alikhani, *California's CCPA Triggers A Tsunami of State-Level Data Privacy Laws*, FORBES (Feb. 20, 2020), https://www.forbes.com/sites/forbestechcouncil/2020/02/20/californias-ccpa-triggers-a-tsunami-of-state-level-data-privacy-laws/#209913dd6cad.

[180] CAL. CIV. CODE § 1798.145(c)(1). There is some debate as to whether the CCPA offers exemptions for any data deemed PHI or whether an organization must be regulated by HHS for both PHI and covered entity/business associate status. It seems likely that the CCPA's drafters desired not to duplicate or overregulate in the HIPAA space.

[181] The CCPA does include an express carve-out for HIPAA and state healthcare law-regulated organizations, seemingly acknowledging that HIPAA includes far more specificity than the CCPA. CAL. CIV. CODE § 1798.145(c)(1)(A).

[182] CAL. CIV. CODE §§ 999.305, 999.307 (notice of financial incentives to providing personal information), 999.330, 1798.120(c).

[183] CAL. CIV. CODE § 1798.140(h).

these combined approaches are poorly suited for AI technologies and, in particular, patient interests.

Table 1: Legal Comparison

Law	HIPAA	FTC	CCPA
Applicability	Covered entities and their business associates when access/process/ store/retain PHI	All organizations doing business in the U.S.	Organizations doing business with California residents or operating in California (with $$ requirements)
Notice	Notice of Privacy Practices; Authorization	Privacy Notice	Privacy Notice
Consent	Acknowledgement of receipt; Consent; consent revocation	Implied consent; opt-out consent revocation	Implied consent except in specific circumstances; opt-out consent revocation
Minimum Necessary	Yes	No	No
De-Identification available	Yes, Safe Harbor (18 identifiers; expert determination)	No specific direction, presumed available	Yes, reasonable technology + no reidentification

III. INCONSISTENT PRIVACY AND AI AIMS

HIPAA, the FTC Fair Information Practices, and the CCPA all codify a model that focuses on privacy without appropriately balancing individual rights and potential risks with data needs for safe, effective, and fair healthcare AI, or what this author calls *data essentialism*.[184] Notice and consent, data minimization, and de-

[184] *See* Epstein & Tschider, *supra* note 36.

identification demonstrate key incompatibilities between AI technologies and privacy law.

A. Notice and Consent

As described in Part II, notice and consent in the United States is the preferred procedural mechanism for individual choice. The role of consent, at least in the United States, is to manifest some agreement to a proposed scheme of individually identifiable data processing.

Notice and consent, unfortunately, is not without limitations. It is ineffective precisely due to five key problems, or "the consent myth": 1) voluntariness, 2) structural limitations, 3) cognition issues, 4) exogeneity, and 5) temporal issues.[185]

1. Lack of Patient Choice in the Healthcare Industry

Healthcare is one area where the concept of "choice" becomes murky, in large part because of the inherent disparity in knowledge, access to information, and complexity of relationships between healthcare providers, insurers, and medical device manufacturers that create medical AI. This complexity creates fewer medical choices (and sometimes only one choice) for patients. Fewer insurance and medical choices result in fewer medical device options, which collectively limit the ability of patients to influence their healthcare privacy options overall.

Patients depend on their doctors and healthcare providers, which necessitates trust, a key element of any privacy-based relationship, especially for fiduciaries like physicians.[186] However, although a healthcare provider may think a particular treatment is best or consistent with a new standard of care, treatment options may be limited by what the patient's insurance will cover (whether government-provided or private insurer), which may influence

---

[185] Tschider, *supra* note 142, at 1519-28. Richards and Hartzog similarly note the concept of "unwitting consent," which also explains issues of voluntariness, structural limitations, cognitive issues, and exogeneity problems. Richards & Hartzog, *supra* note 107, at 1478-84. They also note the concerns of coercion in data collection and use, which reduce the function of voluntariness, especially when alternative options are limited. *Id.* at 1486-87. Finally, some versions of consent are not consent at all: such as when an individual is not capable or may be incapacitated. *Id.* at 1490-91.

[186] *See generally* ARI EZRA WALDMAN, PRIVACY AS TRUST (Cambridge Univ. Press: 2018) (describing the essential nature of trust in relational constructs).

whether a doctor recommends a course of treatment or not.[187] Indeed, insurers may offer more favorable reimbursement to healthcare providers for certain medical devices and may decline reimbursement altogether for others.[188]

Reimbursement for the use of health AI outside clinical trials, as in diagnostics or medical device use, is dependent on the type of insurance provided (e.g., HMO, PPO, Medicaid) and preexisting reimbursement models between insurers and providers, which are often opaque to patients.[189] Often patients do not have many insurance options available, whether from an employer, on the "open" Patient Protection and Affordable Care Act (ACA) marketplace, or from the government, all of which tie covered persons to an insurer's reimbursement policies, including preferred diagnostics or devices.[190] For example, a Health Maintenance Organization (HMO)-based insurance plan usually requires an individual to select a primary

[187] Susan B. Yeon, *The Scope of Medicare Reimbursement for New Medical Devices: Impact on Device Availability and the Standard of Care*, LEDA AT HARV. L. SCH., https://dash.harvard.edu/bitstream/handle/1/8852165/syeon.html?sequence=2 (last visited July 31, 2021).

[188] *When Insurers and Doctors Haggle Over Medicaid Costs, Patients Pay the Price*, MEDICALXPRESS (July 26, 2021), https://medicalxpress.com/news/2021-07-doctors-haggle-medicaid-patients-price.html.

[189] *Reimbursement: A Medical Device Company's Worst Nightmare?*, MASTERCONTROL (Aug. 12, 2015), https://www.mastercontrol.com/gxp-lifeline/reimbursement-a-medical-device-company's-worst-nightmare-/; *Understanding Reimbursement for Medical Devices: Coding, Coverage, Payment, and Payors,* THE ATTICUS GROUP BLOG (Jan 1, 2017), https://theatticusgroup.net/understanding-reimbursement-medical-devices-coding-coverage-payment-payors/; David P. Lind, *Secret Contracts Between Insurers and Providers – Who Benefits?*, HEARTLAND HEALTH RES. INST. BLOG (Nov. 13, 2018), https://hhri.net/secret-contracts-between-insurers-and-providers-who-benefits/.

[190] The Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2012, 78 Fed. Reg. 15410 (March 11, 2013) (codified at 45 C.F.R. pts. 153, 155, 156, 157, & 158). Under the ACA, individual markets have limited offerings of insurers available in that geographic location. For example, some counties only have one insurer available, though this rate has improved substantially since 2018. Daniel McDermott & Cynthia Cox, *Insurer Participation on the ACA Marketplaces, 2014-2021* (Nov. 23, 2020), https://www.kff.org/private-insurance/issue-brief/insurer-participation-on-the-aca-marketplaces-2014-2021/. The creation of Accountable Care Organizations (ACO) could also reduce competition as they incentivize healthcare consolidation. Issac D. Buck, *Furthering the Fiduciary Metaphor: The Duty of Providers to the Payers of Medicine,* 104 CAL. L. REV. 1043, 1078 (2016).

provider.[191] Preferred Provider Organizations (PPO) usually carry significant financial incentives to select a primary provider and receive care there.[192] Medicare and Medicaid certify specific providers to receive reimbursement funds.[193] These concepts reflect a higher degree of coercion than the typical marketplace because deviating from these limited options is financially undesirable.

Physicians who deviate from the reimbursement model of the patient's insurer risk protracted appeal processes or no reimbursement at all.[194] In some situations, the remaining cost is billed to the patient, in others the provider must absorb the cost.[195] Overall, although physicians often do and should recommend an appropriate care plan for an individual, reimbursement challenges likely affect what is ultimately recommended to patients.[196] The increased consolidation of providers similarly has reduced patient options.[197]

The lack of alternative options for AI technologies, such as diagnostic AI or connected medical devices like a surgical robot or an insulin pump, further limits a patient's choices. Diagnostic AI technologies specifically do not have many competitors precisely *because* they are transformative and cutting-edge. Connected medical devices, especially those incorporating AI, likely do not have many competitors because of the complexity of such technologies and market dynamics, as in many innovative products, alongside a heavy

---

191 *What Types of Health Plans Are Available?,* NH HEALTH COST (Apr. 9, 2018), https://nhhealthcost.nh.gov/guide/question/what-types-health-plans-are-available-0.

192 *Id.;* Adam Felman, *How are PPO and HMO Medicare Different?*, MED. NEWS TODAY (Apr. 15, 2020), https://www.medicalnewstoday.com/articles/how-are-ppo-and-hmo-medicare-different.

193 *Become a Medicare Provider or Supplier,* CTRS. FOR MEDICARE & MEDICAID (Dec. 1, 2020), https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/Become-a-Medicare-Provider-or-Supplier.

194 *See* Lind, *supra* note 189; *When Insurers and Doctors Haggle Over Medicaid Costs, supra* note 188.

195 Joshua W. Axene, *Paying Healthcare Providers: The Impact of Provider Reimbursement on Overall Cost of Care and Treatment Decisions*, AXENE HEALTH PARTNERS (2021), https://axenehp.com/paying-healthcare-providers-impact-provider-reimbursement-overall-cost-care-treatment-decisions/.

196 *Id.*

197 Jacqueline LaPointe, *Healthcare M&A Leads to 90% of Markets Being Highly Consolidated*, REVCYCLE INTELLIGENCE (Aug. 8, 2018), https://revcycleintelligence.com/news/healthcare-ma-leads-to-90-of-markets-being-highly-consolidated.

acquisition trend.[198] And indeed, it is highly difficult to enter this market as a new competitor due to substantial regulatory challenges.[199] For both diagnostic medical devices and implantable or affixed medical devices, the possibility of a legitimate "alternative choice" may not be available at all.

Patients are last in line regarding healthcare choices far *before* they must make a decision about whether they agree with any one of these organizations 'privacy policies. Insurers, which are available because an employer or the government has provided limited options, have relationships with a limited number of healthcare providers, which treat patients knowing that some medical products will not be reimbursed or may only be partially reimbursed. And few products may even be available for free or low cost to a patient of that type and with that diagnosis. In some cases, a patient will be left with only one option. In each and every step, patients have very little control over choices made about them, especially regarding privacy interests.

2. *Voluntariness and Coercion in Healthcare Personal Information (and Protected Health Information) Use*

The complexity of healthcare transactions influences how much choice a patient has, including how their personal information (or Protected Health Information (PHI)) will be used. Patients must provide personal information or PHI to all of these organizations—insurer, healthcare provider, medical device manufacturer (when a device is used) to receive and pay for healthcare. In the beginning of these relationships and ongoing upon material change, an insurance provider or health plan, healthcare provider, and, sometimes, medical device manufacturers, display privacy notices to patients.[200] These notices (unsurprisingly) are contracts of adhesion, too, as each organization's notice of privacy practices and authorization

---

[198] Andy Dixon & Tyler Bradshaw, *Fast Forward: Consolidation Continues in Medical Device Contract Manufacturing*, HARRISWILLIAMS (Mar. 2018), https://www.harriswilliams.com/article/fast-forward-consolidation-continues-medical-device-contract-manufacturing.

[199] Matthew Grennan & Robert Town, *Is the FDA Too Tough on Medical Device Makers?* KNOWLEDGE@WHARTON (June 25, 2015), https://knowledge.wharton.upenn.edu/article/the-just-right-zone-for-medical-device-regulation/.

[200] 45 C.F.R. § 164.520.

documents are identical (with respect to all similarly situated persons) and non-negotiable.[201]

Contracts of adhesion are an accepted form of consumer contracting used heavily in healthcare transactions, even though "take it or leave it" contractual terms are inherently coercive.[202] The main difference between healthcare and other consumer contracts is that the stakes are usually much higher. For example, under typical consumer circumstances, if a consumer does not like a privacy notice for a connected home thermostat system, the consumer could select a different product.[203] But in healthcare, the likelihood of a patient seeking care from another healthcare provider organization or selecting an alternative health technology due to unfavorable privacy terms in a contract of adhesion is extremely low.

In the case of prescribed medical devices, a patient is not usually in a position to refuse a device recommended by the patient's physician based on data practices: in part, because of the unavailability of this information to the physician, but mostly because the patient is in a position of trust and need. The alternative choice may be a choice that dramatically reduces the overall safety or efficacy of diagnosis or treatment or simply is not reimbursable.

Finally, there is the issue of ranked preferences: when a patient selects a medical device or an AI procedure, the patient likely ranks their safety or the technology's efficacy, such as reduction in recovery time, minimally invasive procedure, results accuracy, or even usability higher than comparatively more abstract data protection concerns.[204] The patient also likely selects the device that does not cost them significant out-of-pocket expenses. These ranked preferences should

---

201 Tschider, *supra* note 108, at 1519-20.

202 Nora K. Duncan, *Adhesion Contracts: A Twentieth Century Problem for a Nineteenth Century Code*, 34 LA. L. REV. 1081 (1974).

203 *Id.* at 1521. It should be noted that in most consumer transactions, coercive practices are used to obtain personal information. *See* Richards & Hartzog, *supra* note 107, at 1488-89.

204 *See generally* Patricia Flatley Brennan & Indiana Strombom, *Improving Health Care by Understanding Patient Preferences*, 5 J. AM. MED. INFO. ASS'N. 257, 259 (1998) ("While the value of understanding and using patient preferences in health care is well recognized, its implementation presents a daunting challenge to clinicians and patients alike. To imagine what a future state of health might be like and to determine the desirability of that future state are complex cognitive tasks. In addition, many patients lack experience in thinking about abstract concepts such as values and preferences. Attempting to do so under the stressful circumstance of the clinical encounter taxes the patient to an even greater degree.").

not necessarily lead us to believe that privacy is not important to patients, but rather signal that a patient is not in the kind of position to safeguard their own interests via private contracting. Simply, immediate salient concerns will present as more important, which may constitute a higher likelihood of coercion regarding more abstract or long-term concerns, such as privacy.[205]

3. AI Technology Provider Issues

Structural, cognition, and exogeneity issues are well-known for nearly all consumer devices. The sheer volume of available privacy notices and subsequent consent to these notices makes it nearly impossible for consumers to read them all: one study recorded the time to review every notice presented at 76 working days a year.[206] Cognition issues are related not only to whether or not notices are written in plain language but also due to the difficulty of individually assessing risk related to data collection, use, and retention.[207]

Risk is similarly complicated due to exogenous risk factors, which are only exacerbated in AI infrastructures.[208] A primary entity, like a hospital or clinic, will usually engage third parties to provide an AI solution, which means the primary entity that maintains a relationship with the patient will not usually completely understand how data are used within the solution.[209]

---

[205] *Id*.

[206] *See* Tschider, *supra* note 108, at 1522; Alexis C. Madrigal, *Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days*, THE ATLANTIC (Mar. 1, 2012), https://www.theatlantic.com/technology/archive/2012/03/re ading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/ [https://perma. cc/93FW-EEPA].

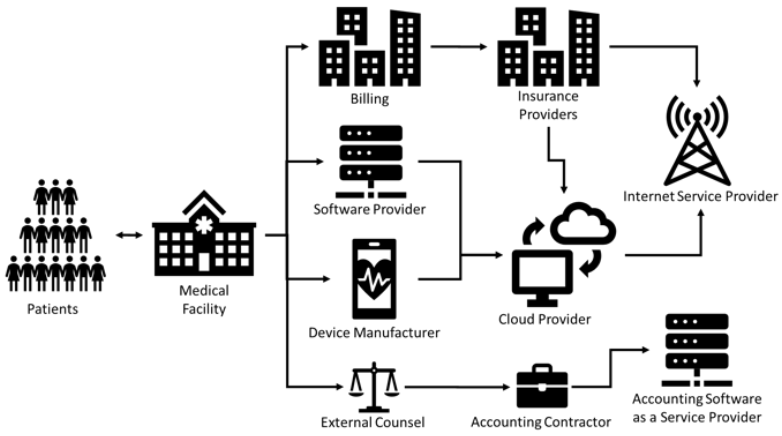
[207] Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1885 (2013) ("There is a more difficult problem with proposals for improved notice . . . [s]uch proposals neglect a fundamental dilemma of notice: making it simple and easy to understand conflicts with fully informing people about the consequences of giving up data, which are quite complex if explained in sufficient detail to be meaningful. People need a deeper understanding and background to make informed choices, [but] many privacy notices . . . are vague about future uses of data.").

[208] *See* Tschider, *supra* note 142, at 1524 (describing the inability of an individual person to approximate potential risks involving third party activities and contracts that govern third party relationships with a primary entity with whom a patient might do business).

[209] Although the healthcare provider will be considered a covered entity under HIPAA or a

Additionally, AI solutions depend on third parties (subcontractors), such as big data cloud providers or other infrastructure service providers to make decisions regarding these data.[210] Typically, hospital entities form contracts with AI solutions, which in turn contract with subcontractors, which contract with other subcontractors, which may or may not be regulated under any U.S. law.[211] The exogeneity of these practices dramatically affects patient privacy risk.[212] See Figure 3 for an example of third-party complexity.

Figure 3: Third-Party Relationships



As you can see in Figure 3, which is a highly simplified map of third-party relationships, from the perspective of the patient, describing third parties who may be involved in data handling practices may be tremendously difficult. In fact, the Notice of Privacy

primary entity from the perspective of the FTC and the CCPA, they also have comparatively less information with regards to how the system actually works, despite being responsible (typically) for providing a privacy notice. It should be noted that some AI providers may be coextensively considered covered entities under HIPAA, yet lack the direct relationship with the patient to adequately inform about potential risks.

210 *See* Tschider, *supra* note 2, at 2.

211 HIPAA addresses this concern through the Business Associate Agreement (BAA), which essentially exports HIPAA requirements via contractual terms. 45 C.F.R. § 164.308(b)(2).

212 *See* Tschider, *supra* note 142, at 1524-25.

Practices under HIPAA, the privacy notice under the Fair Information Practices, and a privacy notice under CCPA do not require disclosure of the actual names of third parties or their specific roles.[213] These laws only require a description of the *types* of third parties involved.

A HIPAA authorization may disclose the identity of the third party, but it also does not provide details of the third party's programs or provide sufficient information for even a sophisticated patient to adequately assess potential risk to their PHI.[214] Notably, the Fair Information Practices and the CCPA do not require any additional detail be provided for secondary use. The depth and relative opacity of subcontracting relationships, including details about their respective privacy and security programs, contractual obligations, and other data uses makes adequately transparent disclosure of primary and third-party data handling practices nearly impossible.

4. Temporality and Prior Notice

Finally, consent suffers from temporality concerns. The concept of notice and consent is premised on a legitimate and logical model that if a patient consents to data practices after being notified of them, this will mean that they have weighed their risks prior to making a decision.[215] For simple data uses with non-complex or no third-party relationships, the concept of *prior* notice coupled with consent works reasonably well to notify a patient of potential risks and for that patient to decide whether or not to proceed.

However, AI manufacturers, even for locked AI where clinical trials feed the creation of an initial algorithm, may not know which data are most useful.[216] And practically speaking, an AI manufacturer generally does not have any direct relationship or opportunity to

---

213 45 C.F.R. § 164.520(b); *see* FTC PRIVACY ONLINE, *supra* note 116.

214 45 C.F.R. § 164.512.

215 *See* Solove, *supra* note 208, at 1880 (defining "privacy self-management" as a "bundle of rights [that] . . . provide[s] people with control over their personal data [resulting in people deciding for themselves] how to weigh the costs and benefits of the collection, use, or disclosure of their information").

216 *See* Tschider, *supra* note 142, at 1527.

meaningfully interface with the patient or offer direct communication.[217]

For example, a system could collect medical record data other than data related to the specific AI that becomes incredibly important to the algorithm's function at a later time but does this through a combination of indirectly collected data from healthcare providers as well as from other sources. However, it may not be possible to accurately describe to patients why this individually identifiable information or PHI is being collected prior to collecting the data or even have an interface to do so.

Notice and consent exacerbates these key disclosure issues because it positions consent as "choice" when unencumbered choice is not possible. The reliance on notice and choice procedure as a stand-in for unencumbered choice misleads patients through a false sense of security. AI systems typify the very conditions for choice imperfections.

B. Data Minimization and Identifiability

Most algorithmic development efforts depend on access to data.[218] Data sets used to train algorithms for a diagnostic result may be reused to train algorithms for another diagnostic result.[219] For example, heart

---

[217] It may be possible for a manufacturer to post a privacy notice on their website for a particular device type, and indeed this activity is mandated by California, even prior to the passage of the CCPA.

[218] Indeed, data are essential to the development of AI. *See* W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775, 800-01 (2021).

[219] Diagnostic and kinetic AI products may use clinical data to tune these systems. However, most systems benefit from ongoing or transferred data use. For purposes of this article, we are focusing on data collected in a commercial setting, rather than clinical data. However, clinical data sharing and reuse is a common issue in the development phase for many technologies and, indeed, data use restrictions in contract and law can create significant issues for data availability for these purposes. *See, e.g.* Michael Mattioli, *The Data-Pooling Problem*, 32 BERK. TECH. L. J. 179, 200-01 (2017), https://btlj.org/data/articles2017/vol32/32_1/MATTIOLI_web.pdf (describing challenges related to restrictions in data sharing and availability for cancer research, including professional, competitive, and reputational concerns by Principal Investigators, rather than concerns over privacy). Governments like the U.S. have tried to address this with data and scholarly deposit contractual obligations, but it does not seem to have changed the research world dramatically. Charlotte Tschider, *Innovation in the Public Sphere: Reimagining Law and Economics to Solve the National Institutes of Health Publishing Controversy*, 1 J. L. & BIOSCI. 281

conditions may be diagnosed using similar data. An electrocardigram (ECG) machine is used to diagnose arrythmia, coronary artery disease, congenital heart defects, enlarged heart, pacemaker efficacy, or heart failure.[220] It is not hard to imagine, then, that data collected using an ECG or similar machine could be reused to develop new and different algorithms. Similarly, conditions like Alzheimer's or Parkinson's Disease, which manifest in various physical symptoms, could be diagnosed using data related to ambulation, neurological testing, brain imaging, and muscular performance. Without quality data, AI algorithms are less effective. Without data, AI algorithms cannot be created at all; for AI, all in-scope data needed to develop, train, or improve AI *are* necessary.

1. Data Minimization

It is well-known that AI's success is based on access to large, accurate, and well-labeled data stores, or data maximization. But part of AI's unique value proposition is to render personalized medicine to patients by increasing safe, effective, and fair services specific to the individual.[221] Precision medicine requires vast access to diverse and identifiable data elements, in particular electronic health records, so that AI systems can work most effectively. According to a recent study," [t]o implement effective personalized and population health with enhanced ability to positively impact patient outcomes, it is important to harness the power of electronic health records (EHR) by integrating disparate data sources and discovering patient-specific patterns of disease progression to provide real-time decision support."[222] The researchers went on to describe how the AI algorithm relied on data and infrastructure to create multiple applications:

(2014), https://academic.oup.com/jlb/issue/1/3 (describing the substantial financial impacts through lack of data and scholarly sharing for publicly funded research, despite contractual obligations to deposit these details in publicly available databases as a condition for funding).

220 *Electrocardiogram (ECG or EKG)*, MAYO CLINIC, https://www.mayoclinic.org/tests-procedures/ekg/about/pac-20384983 (last visited Apr. 24, 2021).

221 *See* Tschider, *supra* note 2, at 708.

222 Zeeshan Ahmed et al., *Artificial Intelligence with Multi-functional Machine Learning Platform Development for Better Healthcare and Precision Medicine*, DATABASE (OXFORD) 1, 2-4 (2020), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7078068/.

> [We] focused on deep learning algorithm implementation with increased data flow that allows machines to self-develop a complex function with improved predictability, as long as a large amount of data is fed as input. They developed a deep convoluted neural network for skin cancer detection, image analysis for diabetic retinopathy evaluation, smartphone-based AI platform to measure adherence in patients on direct oral anticoagulants and patient's visit length reduction.[223]

The concept of personalization is directly related to big data volumes to train, for example, AI machine learning classifiers. These classifiers, through data analysis, create the appropriate weightings and relationships necessary for AI functionality.[224]

Personalized medicine, by its very definition, requires *some* personalization, which requires access to identifiable data for a portion of the AI lifecycle: during algorithmic training and clinical trials, as well as when an individual patient "uses" the AI.[225] Although some AI used to aid in healthcare efficiency goals, accountable care strategies, or quality goals may be effectively de-identified,[226] personalization complicates the degree to which de-identification is possible or desirable.

2. Data De-Identification

As an affordance HIPAA and the CCPA offer, it is tempting to believe that organizations could render all data de-identified to simply avoid privacy obligations and reduce privacy risk to patients. However, the concept of de-identification, or removing data elements

---

[223] *Id.*

[224] *Id.* at 4.

[225] Clinical trials are typically subject to the Common Rule, which offers considerably more flexibility in establishing de-identification, as detailed in guidance from the Office of Human Research Protections (OHRP). Data are not "individually identifiable when they cannot be linked to specific individuals by the investigator(s) either directly or through coding systems." Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?* 10 AM. J. BIOETH. 1, 3 (2010), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3032399/. This model seems to open the door for pseudonymity, or individually identifiable information coded where re-identification is possible, just not to individuals leading a study. Despite this, data use after clinical trials by, for example, medical device or pharmaceutical companies, generally will be subject to HIPAA restrictions.

[226] Glenn Laffel, *Using De-Identified Health Information to Improve Care: What, How and Why*, PRACTICE FUSION (Apr. 30, 2010), https://www.practicefusion.com/blog/using-de-identified-patient-data-to/.

to render a data set either non-identifiable or at low risk of re-identification,[227] is complicated by AI's big data stores.[228] HHS 'Safe Harbor enables organizations to remove 18 identifiers from a data set to accomplish Safe Harbor status in use, transfer, or even sales.[229] Although the CCPA does not offer any specificity like HHS 'Safe Harbor, risk of reidentification may be difficult to demonstrate absent expert determination.

As a voracious consumer of data, AI has the potential to render a patient identifiable using a quantum of "de-identified" data elements.[230] For example, a patient's age, general location, location of treatment (but not treatment date), disease, complications, and social media connections could create a high probability of reidentification despite being de-identified according to HIPAA's Safe Harbor, especially when combined with public data or additional data sources.[231] For data like genetic information, which is increasingly used in AI diagnostics, de-identification may not even be possible.[232]

AI exacerbates existing de-identification limitations for three reasons: 1) AI is dependent on large and diverse (and often identifiable) data sets, and usually it is unknown at the time of collection which data are useful, 2) data sets are usually expanded using collected or purchased data sets from other organizations, such as insurers or other vertically integrated organizations, and 3) AI algorithms are often used specifically *to* reidentify individual patients from de-identified data.

---

227 De-identification is distinct from anonymization, which typically requires more removal of identifiable data than the 18 identifiers required under HIPAA's De-identification Safe Harbor.

228 *See* Tschider, *supra* note 1, at 104-109.

229 *Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, U.S. DEPT. OF HEALTH & HUMAN SERVS. (Nov. 6, 2015), https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.

230 Adam H. Greene, *More Data Please! The Challenges of Applying Health Information Privacy Laws to the Development of Artificial Intelligence*, PRIVACY & SEC. L. BLOG (Feb. 26, 2020), https://www.dwt.com/blogs/privacy—security-law-blog/2020/02/ai-healthcare-privacy-laws.

231 MélanieBourassa Forcier et al., *Integrating Artificial Intelligence into Health Care through Data Access: Can the GDPR Act as a Beacon for Policymakers?* 6 J. L. & BIOSCI. 317 (2019), https://academic.oup.com/jlb/article/6/1/317/5570026.

232 *Id.*

3. Contextualizing Data Minimization and De-Identification

The context of data use in AI healthcare scenarios means applying different conceptions of "reasonableness" with respect to data minimization and de-identification. Contextual privacy, as extensively explained by Helen Nissenbaum, involves information flow informed by normative constructs.[233] These norms are informed by information we know about the data subject, sender, recipient, information type, and method of transmission.[234] For example, contextual privacy might demand a different approach to privacy for an AI-enabled pacemaker than for someone using Facebook.

Within diverse healthcare environments, situational contexts are markedly different. For example, a diagnostic tool may not require retention of individually identifiable data or PHI after diagnostic tool use. Unsupervised machine learning algorithms and neural networks, however, may require continuous data feeding for long-term learning, and previously identifiable data may still be useful in de-identified form after diagnosis.

For Internet of Health Things leveraging AI infrastructure, such as a pacemaker or insulin pump, access and use of identifiable data may be strictly necessary to ensure effective personalization of an AI service. After service is cancelled, such identifiable data may be easily de-identified. In both contexts, the *appropriate* balance of use and de-identification may differ based on what conforms to HIPAA's minimum necessary rule.

Healthcare AI creates challenges for longstanding privacy constructs. To prioritize the patient's interests, including interests in safe, effective, and fair AI, privacy law must adapt to more flexible, context-based privacy that eschews insufficient procedural proxies for real choice and one-size-fits-all approaches to collection, use, retention, and de-identification.

---

[233] HELEN NISSENBAUM, PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE 4 (Stanford Univ. Press 2010).

[234] *Id.*

IV. RECOGNIZING LEGITIMATE INTERESTS IN AI SAFETY AND PATIENT PRIVACY

The competing interests of AI safety and patient privacy necessitate an interpretation of privacy law that accounts for differing contexts. Re-conceptualizing notice and consent, data minimization, and de-identification will simultaneously improve the effectiveness of existing privacy requirements while permitting more expansive use when such uses primarily benefit patients. The concept of legitimate interest balancing, long-contemplated in the European Union but not well-defined, offers a useful contextual lens to establish privacy models at the federal and state level.

Although the Article 29 Working Party addressed legitimate interests as a lawful basis for processing data as early as 2014,[235] the European Union's 2018 omnibus privacy regulation, the General Data Protection Regulation (GDPR), enhanced the concept of legitimate interest as a legal alternative to explicit consent.[236] Legitimate interest assessments typically involve evaluating specific benefits to individual persons (such as patients in the healthcare context) and to the organization, then weighing these benefits against one another:

1. Identify a legitimate interest
2. Show that the processing is necessary to achieve it; and
3. Balance it against the individual's interests, rights and freedoms[237]

Weighing these benefits in favor of individual people discourages disproportionate behavior leading to commoditization of patients and their data. Organizations should weigh these benefits at various times in the data lifecycle, ensuring weightings continue to primarily benefit human beings rather than organizations.

---

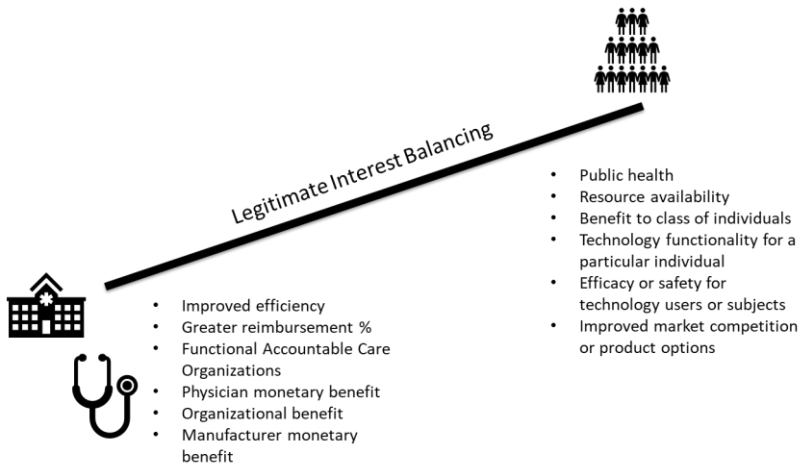
[235] Working Party 844/142014 O.J. (L 217) (EC).

[236] Regulations 2016/679, Art. 6 1(f) Recital 47 2016 O. J. (L 119) (EU). It should be noted that legitimate interest is only available as a lawful basis for processing personal information when benefits to the individual outweigh benefits to the organization. However, legitimate interest is not "read" into any other requirements, notably when consent is used. This potentially leaves the GDPR model open to the same consent issues U.S. privacy law faces. I position legitimate interest as a coextensive requirement intended to bolster ethical privacy practices and avoid consent abuses.

[237] *Legitimate Interests*, INFO. COMM'R'S OFF., https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/ (last visited Oct. 15, 2020).

Despite the language itself not being explicitly mentioned in HIPAA, the Fair Information Practices, or the CCPA, legitimate interest analysis offers a unique lens for evaluating the laws 'privacy features. As a core aspect of legitimate interest analysis, organizations must evaluate whether individually identifiable data or PHI are truly necessary for purposes of benefitting the patient.[238] For example, legitimate interest is often described as "interest balancing," or analysis of the relative data use benefits to an organization or a patient.[239] *See* Figure 4 for an example of the types of interests that may be balanced in this analysis.

Figure 4: Legitimate Interest Balancing



Interest balancing has the potential to offer important nuances in how, and to what extent data collection and use takes place in challenging areas of existing privacy models.

---

[238] *See* DATA PROTECTION NETWORK, GUIDANCE ON THE USE OF LEGITIMATE INTERESTS UNDER THE EU GENERAL DATA PROTECTION REGULATION VERSION 1.0 14 (2017), https://iapp.org/media/pdf/resource_center/DPN-Guidance-A4-Publication.pdf.

[239] *Id.* at 3.

A. Minimum Necessary Data

One initial benefit to legitimate interest balancing is how it can inform organizational conceptions of minimum necessary. Although minimum necessary does not explicitly apply in the Fair Information Practices and the CCPA, it is an important construct of some global privacy laws that addresses individual patient risks through limiting overcollection, use, and retention of such data.[240] For this reason, privacy laws should include a minimum necessary requirement, informed by legitimate interest balancing analysis.

Minimum necessary requirements, as informed by legitimate interest balancing, should extend throughout the organization-patient relationship, including to third parties and to the entire information management lifecycle. For example, organizations and their third parties should not be able to retain identifiable data longer than is necessary to satisfy the interests of the patient, the patient's group (e.g. individuals with congenital heart failure), or for substantial public benefit. Additional limitations on third-party data use will likely apply under traditional data use agreements included in these contracts.

Minimum necessary inherently has some contours of contextual privacy: some AI technologies necessitate greater data collection for safety, efficacy, and fairness purposes, but these data are essential to fulfill patient interests. Indeed, necessity does not necessarily reflect a limited quantum of data: some applications may require more than others. By exclusively including a minimum necessary requirement for these privacy models, organizations must truly consider both their data needs and the interests of the patients they serve. It also means that primary entities need to better understand their third-party relationships and ensure that third-party behavior is consistent with patient interests.

B. Data De-Identification

Data de-identification can also be positioned as an overt requirement related to data minimization. As a practical matter, data may still be useful but do not need to be as identifiable over time, as data exhibit a particular lifecycle. This means that first, we may need

---

[240] CHARLOTTE A. TSCHIDER, INTERNATIONAL CYBERSECURITY AND PRIVACY LAW IN PRACTICE 12-13 (Wolters Kluwer 2018).

to reconceptualize de-identification as a range of identifiability rather than a false dichotomy.

Data de-identification for AI may not ever be *fully* achievable. Indeed, as the data set grows, de-identification while retaining data usefulness becomes less likely. However, organizations can both retain usefulness and reduce risk to individuals by removing data elements or portions of data elements that are not necessary to collect, use, or retain. This means that *all* activities related to data collection, use, and retention will remain identifiable, albeit *less* identifiable. This means that organizations must still comply with privacy laws like HIPAA, the FTC Act, and the CCPA, and an "easy out" of the statute would no longer exist.

While this might appear overly restrictive on the surface, legitimate interest analysis could offer additional flexibility based on how identifiable the data are. For example, when it is necessary to retain some identifiable data, an organization may pursue expert determination or alternative privacy enhancing technologies to demonstrate low risk to a patient.[241] When there is low risk to a patient, privacy laws could permit more flexible use and data sharing for purposes that benefit the patient following legitimate interest analysis.

Implementing an alternative model like this will benefit organizations in potentially making data more available and open, but it also requires organizations to think more strategically about which data to collect or retain, and when such data should be securely deleted. Therefore, organizations aiming to extensively work with big data implementation will need to develop extensive data lifecycle management strategies that consider the relative costs and benefits to individual patients and to themselves.

C. Notice and Consent

The most commonly used privacy notice in healthcare is the Notice of Privacy Practices, or a privacy notice under the Fair Information Practices and CCPA.[242] The HIPAA Notice of Privacy

---

[241] Cem Dilmegani, *Top 10 Privacy Enhancing Technologies (PETs) & Use Cases*, AI MULTIPLE (July 21, 2020), https://research.aimultiple.com/privacy-enhancing-technologies/; JULES POLONETSKY & ELIZABETH RENIERIS, FUTURE OF PRIVACY FORUM, *PRIVACY 2020 10 PRIVACY RISKS AND 10 PRIVACY ENHANCING TECHNOLOGIES TO WATCH IN THE NEXT DECADE* (Jan. 2020), https://fpf.org/wp-content/uploads/2020/01/FPF_Privacy2020_WhitePaper.pdf.

[242] CAL. CIV. CODE § 1789.100; *see* FTC PRIVACY ONLINE, *supra* note 116.

Practices at the federal level does not require consent, and neither do the Fair Information Practices or CCPA.[243] Despite issues with consent specifically, privacy notices perform a functional role in restricting, to some degree, what an organization may do by binding them to their disclosed purposes for collection.

Most organizations, however, do not disclose when data may be collected for purposes of AI use. Although this does not perform a curative role for patients to enforce their interests, it does enable patients to be aware of AI use and for external parties to hold organizations accountable. Therefore, if organizations intend to engage in AI activities of any kind, including analyzing data for internal operational purposes, to assist in diagnostic or treatment decisions, or when prescribing an AI medical device or using physical, tangible AI machines, the organization should at least disclose that AI systems are used and generally for what purposes.

This approach notifies the patient of AI use, which, for those who are interested in learning more, may prompt a patient to ask more questions of their physicians or to follow-up with the organization.[244] For HIPAA specifically, AI, as described in Part I, could potentially be considered by HHS as part of healthcare treatment, payment, or healthcare operations, which eliminates the need for additional authorizations under some circumstances.

This does not necessarily mean that *all* AI uses would or should be included in the Notice of Privacy Practices. Reliance on third parties to develop these products and potentially analyze or host data could, however, cut in favor of facilitating authorization. HHS has an opportunity to define how will be used to automate and aggregate operational data, to assist in diagnosis or treatment, or to assist the functioning of a human body as primarily within the scope of treatment, payment, or healthcare operations.

An interest-balancing approach undergirds all decisions made regarding data collection and use, and ultimately should be positioned NOT as an alternative to consent, but rather as a separate and

---

[243] *Id.*

[244] It should be noted that although not specified here, the physician may be under an obligation related to informed consent to medical procedures, or may be encouraged due to the potential for a malpractice suit, to disclose when AI are used in health diagnosis or treatment, as well. Despite imperfections in the nature of consent *to* privacy notices generally, notice, if not overly complex, may at least prompt further discussion.

coextensive requirement to other privacy obligations. The cumulative effect of legitimate interest analysis across data uses means that more responsibility rests on the shoulders of organizations collecting data, rather than expecting patients to understand and advocate for their own interests.

It also means that data may be used more flexibly, so long as such uses are consistent with general AI functionality and benefit the patient. For example, benefits may be ranked from most personalized to least: improving the very product the patient uses or was used for the patient (highest weighting), improving product offerings for a class of patients (mid-level weighting), such as patients with Type-1 Diabetes, or generally improving products overall for patients (lower-level weighting).

The Data Protection Network also provides some useful direction for analyzing legitimate interests through the data use lifecycle:

> (a) any link between the original purpose and the intended future processing
>
> (b) the context in which the Personal Data was collected; specifically, the relationship between the Controller and the individual
>
> (c) the nature of the Personal Data
>
> (d) the possible consequences of the change of purpose on individuals
>
> (e) the existence of appropriate safeguards, e.g. encryption or pseudonymization[245]

The cumulative result of patient benefits could be directly compared against organizational benefits, and associated legitimate interest assessment records could be retained by the organization at their discretion in the event an investigation results.

CONCLUSION

This new model centralizes the role of legitimate interest analysis as a key patient risk balancing lens for evaluating data collection, use, retention, and identifiability. It also provides the opportunity to relax

---

[245] *See* DATA PROTECTION NETWORK, *supra* note 238, at 77. Of special note, this approach to evaluating legitimate interest takes into account privacy context.

certain procedural privacy functions that either do not add much value or, worse yet, mislead patients into a false sense of security.

It is tempting to consider rewriting these laws to establish greater restrictions and manifest more comprehensive disclosures to enhance patient choice. Unfortunately, the nature of modern healthcare simply does not provide the scaffolding to animate meaningful choice. Moreover, issues with existing privacy models are not cured by doubling down on ineffective models that restrict access to crucially important data. By integrating the legitimate interest concept into data minimization, identifiability, and notice, privacy laws will consistently enable organizations to create world-class products while protecting patients. Through appropriate data use and reuse, both patients and organizations legitimately stand to benefit.