Utah Law Review

Volume 2021 | Number 4 Article 5

11-2021

Problematic Interactions Between AI and Health Privacy

Nicholson Price
University of Michigan Law School, wnp@umich.edu

Recommended Citation
Nicholson Price, Problematic Interactions Between AI and Health Privacy, 2021 ULR 925 (2021).
https://doi.org/10.26054/0d-th4e-sgvq

PROBLEMATIC INTERACTIONS BETWEEN AI AND HEALTH PRIVACY

W. Nicholson Price II*

The interaction of artificial intelligence ("AI") and health privacy is a two-way street. Both directions are problematic. This Article makes two main points. First, the advent of artificial intelligence weakens the legal protections for health privacy by rendering deidentification less reliable and by inferring health information from unprotected data sources. Second, the legal rules that protect health privacy nonetheless detrimentally impact the development of AI used in the health system by introducing multiple sources of bias: collection and sharing of data by a small set of entities, the process of data collection while following privacy rules, and the use of non-health data to infer health information. The result is an unfortunate anti-synergy: privacy protections are weak and illusory, but rules meant to protect privacy hinder other socially valuable goals. This state of affairs creates biases in health AI, privileges commercial research over academic research, and is ill-suited to either improve health care or protect patients' privacy. The ongoing dysfunction calls for a new bargain between patients and the health system about the uses of patient data.

I. IMPACT OF AI ON MEDICAL PRIVACY

Consider first the impact of artificial intelligence on medical privacy. The advent of artificial intelligence—alongside the big data with which it is trained and on which it operates—weakens mechanisms used to protect medical data privacy in at least two ways. First, AI enables actors with big data and sufficient computing capacity to work around deidentification, a key front-line protection for patient health data. Second, by enabling accurate and sophisticated inferences about health information from large sets of data that are not obviously tied to health, AI reduces the efficacy of trying to protect (or even identify what counts as) "health data."

* © W. Nicholson Price II. Professor of Law, University of Michigan Law School; Core Partner, Centre for Advanced Studies in Biomedical Innovation Law at the University of Copenhagen; and Co-PI, Project on Precision Medicine, AI, and the Law at the Petrie-Flom Center at Harvard Law School. Many thanks to Leslie Francis, Anya Prince, Alexandra Roberts, Kayte Spector-Bagdady, and Charlotte Tschider for thoughtful comments on earlier drafts and to the participants of the 2020 Lee E. Teitelbaum Utah Law Review Symposium for helpful discussion. Thanks also to the editors of the *Utah Law Review* for careful editing. This work was supported by the National Cancer Institute (1 R01 CA214829-01A1) and the Novo Nordisk Foundation (NNF17SA0027784). All errors are my own.

A. Deidentification and Reidentification

Deidentification is a common tool used to protect medical privacy. The Health Insurance Portability and Accountability Act (HIPAA)[1] Privacy Rule is the dominant legal rule governing health data privacy[2] and likely the single most potent federal privacy regime in the United States. The HIPAA Privacy Rule only governs identifiable health information and includes a safe harbor under which information that has been stripped of 18 listed identifiers is defined as not identifiable.[3] What does that mean? Information custodians can remove those identifiers from health data and stop worrying about HIPAA (at least with respect to those data). Deidentification is a popular intervention outside the United States as well; the European Union's General Data Protection Regulation, for instance, does not cover anonymized data.[4]

Artificial intelligence reduces the already-weak power of deidentification[5] to protect health privacy by making it easier to reidentify patients, either individually or at scale.[6] AI enables reidentification by finding patterns in data. Perhaps most

---

[1] Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936.

[2] 45 C.F.R. §§ 160, 164 (2019); The HIPAA Privacy Rule is not the only health privacy law in the United States, of course; state laws may have more restrictive provisions on specific topics or in general, and other federal laws govern subsets of health privacy, such as genetic information. But HIPAA cuts across state lines and structures much of the discussion surrounding health data privacy.

[3] 45 C.F.R. § 164.514(b)(2) (2019).

[4] Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1, 26, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [https://perma.cc/QYF8-WGX3]. Note that the GDPR anonymity standard requires that deidentification be so complete that reidentification is impossible—a standard which the rest of this Section suggests may be nigh impossible. *See* Charlotte A. Tschider, *Regulating the Internet of Things: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age*, 96 DENV. L. REV. 87, 105 (2018) (noting that the anonymization standard renders many data effectively unusable).

[5] *See*, *e.g.*, Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716–26 (2010) (noting ways to reidentify data); Yaniv Erlich & Arvind Narayanan, *Routes for Breaching and Protecting Genetic Privacy*, 15 NATURE REV. GENETICS 409, 409–16 (2014) (cataloging ways to reidentify genetic data).

[6] One rebuttal is that AI reidentification is likely to be probabilistic rather than deterministic—that is, while an AI system may think it highly likely that a particular set of data belongs to a particular person, it cannot state that fact with certainty. This seems true but irrelevant, given the inherently probabilistic nature of most data. *Cf.* Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye, *Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models*, 10 NATURE COMMC'NS 3069 (2019), https://www.nature.com/articles/s41467-019-10933-3.pdf [https://perma.cc/8ASB-L2DX] (noting the probabilistic nature of reidentification attacks).

dramatically, researchers have used AI to reidentify a substantial majority of patients from deidentified datasets of physical activity data collected from wearable fitness trackers.[7] AI can also help link deidentified health records with other datasets that include identified information, such as internet searches or other consumer records.[8]

But AI has now appeared on both sides of the health data privacy arms race. Just as AI can be used to decrease the privacy of anonymized datasets, AI can be used to *increase* privacy. AI can deidentify records that are otherwise costly to deidentify, such as textual notes from medical encounters.[9] It can also create fully or partially synthetic datasets—that is, datasets that reflect real data patterns but in which no actual data are real.[10] It's a challenging exercise because for the data to be useful, the patterns must reflect the underlying population, but it's not always easy to know beforehand what patterns are going to be important; simpler patterns are easier to preserve than more complex ones.[11]

---

[7] Liangyuan Na, Cong Yang, Chi-Cheng Lo, Fangyuan Zhao, Yoshimi Fukuoka & Anil Aswani, *Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets from Which Protected Health Information Has Been Removed with Use of Machine Learning*, 1 JAMA NETWORK OPEN e186040 (Dec. 21, 2018), https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2719130 [https://perma.cc/HXF7-9766].

[8] *See* W. Nicholson Price II, Margot E. Kaminski, Timo Minssen & Kayte Spector-Bagdady, *Shadow Health Records Meet New Data Privacy Laws*, 363 SCIENCE 448, 448–49 (2019).

[9] Amber Stubbs, Christopher Kotfila & Özlem Uzuner, *Automated Systems for the De-identification of Longitudinal Clinical Narratives: Overview of 2014 i2b2/UTHealth Shared Task Track 1*, 58 J. BIOMED. INFORMATICS S11, S11 (2015).

[10] *See* Edward Choi, Siddharth Biswal, Bradley Malin, Jon Duke, Walter F. Stewart & Jimeng Sun, *Generating Multi-Label Discrete Patient Records Using Generative Adversarial Networks*, 68 PROC. MACH. LEARNING HEALTHCARE 286 (2017), http://proceedings.mlr.press/v68/choi17a/choi17a.pdf [https://perma.cc/3CH6-AST7] (describing using machine learning to create synthetic datasets); Alexander Watson, *Deep Dive on Generating Synthetic Data for Healthcare*, MEDIUM (May 12, 2020), https://medium.com/gretel-ai/deep-dive-on-generating-synthetic-data-for-healthcare-41acb4078707 [https://perma.cc/RCS6-3MYN] (describing Gretel, a software product for creating synthetic data using machine learning).

[11] Anat Reiner Benaim, Ronit Almog, Yuri Gorelik, Irit Hochberg, Laila Nassar, Tanya Mashiach, Mogher Khamaisi, Yael Lurie, Zaher S. Azzam, Johad Khoury, Daniel Kurnik & Rafael Beyar, *Analyzing Medical Research Results Based on Synthetic Data and Their Relation to Real Data Results: Systematic Comparison from Five Observational Studies*, 8 JMIR MED. INFORMATICS e16492 (2020); Debbie Rankin, Michaela Black, Raymond Bond, Jonathan Wallace, Maurice Mulvenna & Gorka Epelde, *Reliability of Supervised Machine Learning Using Synthetic Data in Health Care: Model to Preserve Privacy for Data Sharing*, 8 JMIR MED. INFORMATICS e18910 (2020).

Of course, when one side pulls ahead in an arms race, the other tries harder[12]—and new AI systems are being developed that can, remarkably enough, extract some identifiable data from purely synthetic datasets.[13] That is, even if no people—no data!—in a synthetic dataset are real, some systems can still glean information about identifiable people from whose data the synthetic dataset was initially created.[14] To be sure, the risk of privacy loss is still much lower with synthetic datasets (at least, for now).[15]

B. Health Inferences

AI can further intrude upon patient privacy by inferring sensitive information about patients, even if that information is never directly shared with anyone. The now-classic example is of Target inferring pregnancy from shopping habits;[16] Anya Prince also persuasively explains how a wealth of health information can be inferred from location data.[17] This pattern—the combination of medical big data and AI enabling the inference of sensitive health data without ever actually accessing sensitive health data—is constantly growing more powerful. As Jeff Skopek puts it, "[The Target] example will soon seem quaint, however, as machine-learning algorithms infer significantly more complex personal traits from seemingly irrelevant data collected across disparate domains of life."[18] Skopek argues that such inferences are not violations of privacy rights as the law understands them,[19] but to the extent that patients, physicians, or others view the acquisition of knowledge about a patient's personal health as sounding in something like privacy, AI nevertheless impacts—and decreases—the strength of that privacy.

---

[12] *See* James Jordon, Daniel Jarrett, Jinsung Yoon, Paul Elbers, Patrick Thoral, Ari Ercole, Cheng Zhang, Danielle Belgrave & Mihaela van der Schaar, Hide-and-Seek Privacy Challenge: Synthetic Data Generation vs. Patient Re-identification with Clinical Time-Series Data (June 30, 2020) (unpublished manuscript), https://www.vanderschaar-lab.com/wp-content/uploads/2020/07/HASPC_overview.pdf [https://perma.cc/8FMF-AKTQ] (describing a literal competition between those generating synthetic datasets based on clinical time-series data and those seeking to reidentify patients based on the synthetic datasets).

[13] Khaled El Emam, Lusy Mosquera & Jason Bass, *Evaluating Identity Disclosure Risk in Fully Synthetic Health Data: Model Development and Validation*, 22 J. MED. INTERNET RSCH. e23139 (2020), https://www.jmir.org/2020/11/e23139/ [https://perma.cc/3NPB-EPZZ].

[14] *Id.*

[15] *Id.*

[16] Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 94–95 (2014).

[17] Anya E.R. Prince, *Location as Health*, 21 HOUS. J. HEALTH L. & POL'Y (forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3767122 [https://perma.cc/2GVG-GTQN].

[18] Jeffrey M. Skopek, *Untangling Privacy: Losses Versus Violations*, 105 IOWA L. REV. 2169, 2223 (2020).

[19] *Id.* at 2223–30.

To make the point broader, protections for health data, HIPAA in particular, take as a given the idea that there is a meaningful category of "health data." Big data and AI show us that "health data" is a nebulous category, and the category of "data that can reveal things about health" contains a much broader set of information.[20] Thus, to the extent that law tries to specially protect health data through privacy regimes, those regimes are likely to be less effective as AI becomes more powerful and more prevalent.

II. IMPACT OF MEDICAL PRIVACY ON AI

In an unfortunate irony, even though AI decreases the strength of health privacy, the rules surrounding health privacy also cause problems for the development of AI used for health and patient care. The use of AI in the health system is rapidly increasing.[21] The Food and Drug Administration has cleared hundreds of AI-powered products for marketing,[22] and many more are being developed and used in-house by hospitals, health systems, and insurers.[23] AI systems are used to diagnose diabetic retinopathy,[24] to identify the risk of brain hemorrhage from CT scans,[25] and to predict the likelihood of patient complications or hospital readmissions,[26] among many other possibilities. But this development faces

[20] W. Nicholson Price II & I. Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 NATURE MED. 37, 39 (2019).

[21] Joachim Roski, Booz Allen Hamilton, Wendy Chapman, Jaimee Heffner, Fred Hutchinson, Ranak Trivedi, Guilherme Del Fiol, Rita Kukafka, Paul Bleicher, Hossein Estiri, Jeffrey Klann & Joni Pierce, *How Artificial Intelligence Is Changing Health and Health Care*, *in* ARTIFICIAL INTELLIGENCE IN HEALTH CARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL 59–79 (Michael Matheny, Sonoo Thadeny Israni, Mahnoor Ahmed & Danielle Whicher eds., 2019), https://nam.edu/artificial-intelligence-special-publication/ [https://perma.cc/D2E9-KLBL].

[22] Casey Ross, *Explore STAT's Database of FDA-Cleared AI Tools*, STAT (Feb. 3, 2021), https://www.statnews.com/2021/02/03/fda-artificial-intelligence-clearance-products/ [https://perma.cc/EXM4-NXVD].

[23] W. Nicholson Price II, Rachel E. Sachs & Rebecca S. Eisenberg, *New Innovation Models in Medical AI*, WASH. U. L. REV. (forthcoming 2022) (manuscript at 4–5) (on file with authors), https://ssrn.com/abstract=3783879 [https://perma.cc/Q963-PKUX].

[24] *History of Digital Diagnostics*, DIGIT. DIAGNOSTICS, http://digitaldx.wpengine.com/about/history/ [https://perma.cc/DXB3-BM3V] (last visited Mar. 8, 2021) (describing the IDx-DR system).

[25] *See* Mohammad R. Arbabshirani, Brandon K. Fornwalt, Gino J. Mongelluzzo, Jonathan D. Suever, Brandon D. Geise, Aalpen A. Patel & Gregory J. Moore, *Advanced Machine Learning in Action: Identification of Intracranial Hemorrhage on Computed Tomography Scans of the Head with Clinical Workflow Integration*, 1 NPJ DIGIT. MED. 9 (2018) (demonstrating a positive impact from applying machine learning to workflow optimization in radiology based on automated CT analyses).

[26] *See, e.g.*, Ben J. Marafino, Alejandro Schuler, Vincent X. Liu, Gabriel J. Escobar & Mike Baiocchi, *Predicting Preventable Hospital Readmissions with Causal Machine Learning*, 55 HEALTH SERVS. RSCH. 993, 993 (2020) (suggesting that machine learning can be used to identify preventable hospital readmissions).

substantial hurdles in terms of privacy rules around health data (closely paralleled by requirements for informed consent in some contexts). For now, I'll set aside whether those hurdles are justified and focus instead on their effects. Principally, privacy protections around health data AI make it more challenging to collect datasets, and in particular make it harder to collect broad, representative, diverse datasets.[27] This results in datasets, and health AI, that reflect and encode problematic biases.

A. Privacy Hurdles for Health AI Development

AI needs to be trained with large amounts of data, whether patient medical records, pharmacy data, insurance claims information, or other health-related data. And therein lies the challenge: the privacy protections for health data, though vulnerable in the ways accounted for above, still raise substantial hurdles for the use of health data to train AI. Dealing with those hurdles raises its own challenges for health AI.

Take HIPAA. The HIPAA Privacy Rule prohibits most health-care providers, health insurers, and health information clearinghouses (collectively, "covered entities") and their business associates from using or disclosing identifiable health information[28] absent any of several specified exceptions, including the authorization of the patient or use for quality improvement—but not for research aimed at developing generalizable knowledge.[29] If a hospital wishes to share patient information useful to develop, for instance, a predictor of the risk of stroke, it must typically either obtain limited-duration individual patient authorization (and consent) to share the information or deidentify the patient data. The first approach is costly, because obtaining patient authorization and meaningfully informed consent[30] takes time.[31] The second approach, deidentifying, can remove information

---

[27] *See generally* Charlotte A. Tschider, *AI's Legitimate Interest: Towards a Public Benefit Privacy Model*, 21 HOUST. J. HEALTH L. & POL'Y (forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3725933 [https://perma.cc/WYW4-4W2Z] (describing this tension).

[28] 45 C.F.R. §§ 160.103, 160.502 (2019).

[29] *Id.* § 160.501.

[30] *See generally* Charlotte A. Tschider, *The Consent Myth: Improving Choice for Patients of the Future*, 96 WASH. U. L. REV. 1505 (2019) (identifying the difficulties of obtaining meaningful consent).

[31] Sharona Hoffman & Andy Podgurski, *Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research*, 65 SMU L. REV. 85, 123 (2012) (reviewing empirical studies of informed consent costs). To be sure, hospitals can find ways around these hurdles by obtaining (arguably unethical) pro forma authorization and consent without meaningfully engaging patients, getting waivers of consent requirements from Institutional Review Boards, using limited datasets with less onerous requirements, or other mechanisms. But all these avenues have their own costs. *See, e.g.*, I. Glenn Cohen, Ruben Amarasingham, Anand Shah, Bin Xie & Bernard Lo, *The Legal and Ethical Concerns that Arise from Using Complex Predictive Analytics in Health Care*, 33 HEALTH AFFS. 1139, 1141 (2014) (noting

that is useful for prediction, such as zip code or age (particularly for elderly patients). Trying to scrub information that isn't one of HIPAA's listed identifiers but is relatively unique or rare can also decrease AI performance.[32] It can also make it hard to reconnect patient records from different parts of a fragmented health system; if records that track a patient for years are most useful, and the patient moves from state to state in that time, deidentifying patient records makes it substantially more difficult to rejoin those records.[33] Different methods of deidentification can shift conclusions drawn from deidentifed data.[34] And finally, deidentification itself is expensive, especially with free-text data such as physician encounter notes.[35]

To be sure, privacy hurdles are just that—hurdles, not walls. They can be surmounted. For instance, well-resourced developers may be able to simply buy partial or deidentified datasets and then use the sort of tactics described above to reconnect disconnected records, reidentify deidentified records, or infer additional information from non-health data to health data to obtain a more complete picture.[36] Commercial developers are not bound by HIPAA's rules (unless as business associates of otherwise covered entities). Once they obtain or create datasets through whatever workarounds are available, they need not follow HIPAA strictures going forward.[37] And commercial developers are also typically not bound by informed-

---

that, to comply with HIPAA requirements, the pro forma method is used, and describing the method as "a highly legalistic form requiring the patient's signature" while lacking patient understanding).

[32] *See, e.g.*, Xing Song, Lemuel R. Waitman, Yong Hu, Bo Luo, Fengjun Li & Mei Liu, *The Impact of Medical Big Data Anonymization on Early Acute Kidney Injury Risk Prediction*, 2020 AMIA JOINT SUMMITS ON TRANSLATIONAL SCI. PROC. 617, 623 (2020) (describing the tradeoff between leaving information in datasets used to predict early acute kidney injury and the performance of those predictors); Charlotte A. Tschider, *Deus ex Machina: Regulating Cybersecurity and Artificial Intelligence for Patients of the Future*, 5 SAVANNAH L. REV. 177, 184 n.49 (2018).

[33] W. Nicholson Price II, *Risk and Resilience in Health Data Infrastructure*, 16 COLO. TECH. L.J. 65, 69–74 (2017).

[34] Heng Xu & Nan Zhang, *Implications of Data Anonymization on the Statistical Evidence of Disparity*, MGT. SCI. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3662612 [https://perma.cc/4VDU-SGSF].

[35] In fact, deidentification is enough of a burden that significant effort has gone into training AI to, yes, deidentify patient medical records. *See generally* Stephane M. Meystre, F. Jeffrey Friedlin, Brett R. South, Shuying Shen & Matthew H. Samore, *Automatic De-identification of Textual Documents in the Electronic Health Record: A Review of Recent Research*, 10 BMC MED. RSCH. METHODOLOGY 70 (2010) (reviewing such efforts); *see supra* Section I.A.

[36] *See, e.g.*, Price et al., *supra* note 8, at 448–49; *see supra* Sections I.A., I.B.

[37] Price et al., *supra* note 8, at 448.

consent requirements that make data sharing and use costlier, and that do apply to most academic institutions and other health-care providers.[38] These factors can make commercial data sources comparatively attractive, even for academic researchers.[39]

B. Privacy Hurdles Bias Dataset Collection and AI Development

The existence of substantial privacy hurdles to health data collection creates the opportunity for biases in the resulting data and AI trained on them. These possibilities arise from multiple sources, including the identity of the entities most able to shoulder the cost of dealing with privacy protections and the processes for addressing or working around those protections. To be very upfront, these processes are not the *only* sources of bias in health data—completely unbiased data collected equally across the entire system would still reflect biases embedded in the underlying health-care itself[40]—but these sources of bias are closely related to the privacy protections described here.

Consider which hospitals and health systems (collectively, "hospitals") can collect and share patient data that can be used to train AI. The process of obtaining patient authorization and consent is expensive (or sometimes impossible, for prior patients), as is the process of reliably deidentifying data.[41] So is the process of trying to make sure that the data to be shared are well-formatted, accurate, and reliable.[42] And so, unsurprisingly, the hospitals that are most likely to share data for the development of health AI are a small subset of all hospitals—a subset with substantial resources.[43] Academic medical centers with more substantial resources are best positioned to gather and collect health data so that it can be used to develop

---

[38] *See*, *e.g.*, Kayte Spector-Bagdady, Raymond Hutchinson, Erin O'Brien Kaleba & Sachin Kheterpal, *Sharing Health Data and Biospecimens with Industry — A Principle-Driven, Practical Approach*, 382 NEW ENG. J. MEDICINE 2072, 2072–75 (2020) (describing the applicable strictures on university data sharing and the University of Michigan's approach to sharing patient data with private industry).

[39] *See*, *e.g.*, Kayte Spector-Bagdady, Amanda Fakih, Chris Krenz, Erica E. Marsh & J. Scott Roberts, *Genetic Data Partnerships: Academic Publications with Privately Owned or Generated Genetic Data*, 21 GENETICS MEDICINE 2827, 2827–29 (2019) (finding a significant increase in academic publications relying on privately held genetic data).

[40] *See generally* DAYNA BOWEN MATTHEW, JUST MEDICINE: A CURE FOR RACIAL INEQUALITY IN AMERICAN HEALTH CARE (2015) (discussing racial bias); DONALD A. BARR, HEALTH DISPARITIES IN THE UNITED STATES: SOCIAL CLASS, RACE, ETHNICITY, AND THE SOCIAL DETERMINANTS OF HEALTH (3d. ed. 2019) (discussing multiple sources of bias); Alexandra D. Lahav, *Medicine Is Made for Men*, N.Y. REV. (Feb. 11, 2021), https://www.nybooks.com/articles/2021/02/11/medicine-is-made-for-men/ [https://perma.cc/3UH9-FE52] (discussing gender bias and reviewing CAROLINE CRIADO PEREZ, INVISIBLE WOMEN: DATA BIAS IN A WORLD DESIGNED FOR MEN (2019)).

[41] *See* Hoffman & Podgurski, *supra* note 31.

[42] W. Nicholson Price II, *Big Data, Patents, and the Future of Medicine,* 37 CARDOZO L. REV. 1401, 1411–15 (2016).

[43] W. Nicholson Price II, *Medical AI and Contextual Bias,* 33 HARV. J.L. & TECH. 65, 79–80 (2019) [hereinafter Price, *Contextual Bias*].

health AI, in contrast to institutions with fewer resources, such as community health centers or rural hospitals.[44]

To take the prime example: undoubtedly, the most important freely accessible set of health data used to train AI is MIMIC (the "Medical Information Mart for Intensive Care"), which includes records from ICU patients seen at a single center in Boston: Beth Israel Deaconess Medical Center.[45] An enormous number of papers and conferences have been based on MIMIC data—over 500 in 2019.[46] But MIMIC hasn't been cheap to create—among other things, Beth Israel Deaconess spends time and resources to carefully remove all identifiable health information so that the data are not subject to the HIPAA Privacy Rule's constraints or informed-consent requirements.[47]

But the use of data from a small subset of high-resource settings creates the chance of biases and limitations in the datasets and resulting AI. MIMIC, after all, is based on data from just one high-resource hospital in Boston. IBM's Watson for Oncology, a much-maligned AI tool aimed at improving cancer care by learning from experts, was trained on data from the high-resource Memorial Sloan Kettering Cancer Center in New York.[48] When researchers noticed that health AI datasets "seemed to be coming from the same sorts of places: the Stanfords and UCSFs and Mass Generals,"[49] their follow-up study found that health AI algorithms "were disproportionately trained on cohorts from California, Massachusetts, and New York, with little to no representation from the remaining 47 states."[50]

---

[44] *Id.* at 79–90.

[45] Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi & Roger G. Mark, *MIMIC-III, A Freely Accessible Critical Care Database*, 3 SCI. DATA 160035 (2016).

[46] Rebecca Robbins, *How Patient Records from One Boston Hospital Fueled an Explosion in AI Research in Medicine*, STAT (July 12, 2019), https://www.statnews.com/2019/07/12/boston-hospital-records-fuel-artificial-intelligence-research/ [https://perma.cc/9PRK-6E55].

[47] *Requesting Access*, MIMIC, https://mimic.physionet.org/gettingstarted/access/ [https://perma.cc/GP3C-UUHV] (last visited Mar. 8, 2021) (describing deidentification and the lack of HIPAA requirements); Robbins, *supra* note 46 (noting the waiver of IRB requirements based on deidentification and other factors).

[48] Casey Ross & Ike Swetlitz, *IBM's Watson Supercomputer Recommended 'Unsafe and Incorrect' Cancer Treatments, Internal Documents Show*, STAT (July 25, 2018), https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/ [https://perma.cc/T6LZ-FN44].

[49] Rebecca Robbins, *Medical AI Systems Are Disproportionately Built with Data from Just Three States, New Research Finds*, STAT (Sept. 25, 2020), https://www.statnews.com/2020/09/25/medical-ai-diagnostic-geographic-diversity/ [https://perma.cc/B377-7UBA] (quoting Amit Kaushal).

[50] Amit Kaushal, Russ Altman & Curt Langlotz, *Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms*, 324 JAMA 1212, 1212–13 (2020).

These high-resource contexts are not representative, and the data from them aren't either. High-resource contexts see different patients and engage in different patterns of care than other contexts.[51] The data from those contexts reflect only certain patients and care patterns, and AI trained on those data reflect those patterns as well—and are correspondingly likely to encounter problems or perform more poorly when translated into other contexts that look different.[52]

In addition to *which* institutions collect data, the process of *how* those institutions collect data can create additional biases. Different patient populations are differently willing to have their data used for future research.[53] These differences are understandable, given the long history of systemic racism and prejudice that exists within the health system that demonstrates a lack of trustworthiness with respect to minority patients.[54] The causes and solutions to systemic racism in the health system are beyond the scope of this piece, but the bias in patient consent is not. Obtaining patient authorization and consent for data sharing can accordingly bias both the resulting datasets and the AI created based on those datasets. Deidentification—and avoiding the patient consent process—avoids these issues, though it may raise its own trustworthiness concerns, and as noted above, it does create separate challenges for dataset and AI quality.

Finally, the triangulation of health information from non-health data, such as shopping patterns, fitness trackers, or internet searches, can circumvent privacy protections but also introduce the possibility for bias. To take a simple example, Apple products typically have more restrictive privacy protections than Google products, including Android-powered phones[55]—and Apple products are often more expensive, and the user communities are demographically different.[56] Data collected

---

[51] *Id.* at 1213.

[52] *Id.*; Price, *Contextual Bias*, *supra* note 43, at 90–98.

[53] Hoffman & Podgurski, *supra* note 31, at 114–19 (discussing bias). *But see* Julie H. T. Dang, Elisa M. Rodriguez, John S. Luque, Deborah O. Erwin, Cathy D. Meade & Moon S. Chen Jr., *Engaging Diverse Populations About Biospecimen Donation for Cancer Research*, 5 J. CMTY. GENETICS 313, 322 (2014) (finding that, across racial and ethnic groups, "[o]nce participants . . . understood the meaning, use, and intent of collecting biospecimens for future research, the majority of individuals demonstrated ample willingness to consider participation").

[54] *See generally* Jan M. McCallum, Dhananjaya M. Arekere, B. Lee Green, Ralph V. Katz & Brian M. Rivers, *Awareness and Knowledge of the U.S. Public Health Service Syphilis Study at Tuskegee: Implications for Biomedical Research*, 17 J. HEALTH CARE FOR POOR & UNDERSERVED 716 (2006) (highlighting the legacy of racism in medical research and its impact on current research); Ruha Benjamin, *Race for Cures: Rethinking the Racial Logics of 'Trust' in Biomedicine*, 8 SOCIO. COMPASS 755 (2014) (discussing the need for medical research to strive to become more trustworthy in light of past failures).

[55] *See* DOUGLAS C. SCHMIDT, GOOGLE DATA COLLECTION 24 (Aug. 2018), https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf [https://perma.cc/RGD3-XZSD] (comparing Google's data collection on Android phones to its collection on iPhones).

[56] *See*, *e.g.*, Jim Edwards, *Here's Why Developers Keep Favoring Apple over Android*,

from smartphones, and the health information that can be inferred (or gleaned directly) from those data, then, are likely to contain at least some bias. Data from internet searches, electronic transactions, and fitness trackers may similarly incorporate bias into resulting datasets and AI trained on them.

To be sure, there are approaches that try to minimize the challenges described here, some of them technological. For instance, federated machine learning techniques involve training models on data from many different institutions while leaving those data in place rather than collecting them.[57] Alongside such technical solutions, an awareness of the problems that can arise from limited datasets can prompt rethinking both the training applied to those datasets and the rules for validation that respond to the potential for bias.[58] And after training, auditing of AI can help reveal biases that have become incorporated, whether they arose from privacy-related workarounds or other sources—though privacy rules may, unsurprisingly, make after-the-fact auditing itself harder to undertake by limiting data sharing.[59]

III. IMPLICATIONS AND CONCLUSIONS

What are we to make of all of this? The current intersection of health privacy and AI seems deeply problematic: AI weakens protections for health privacy, and health privacy weakens the AI used in health.

One reaction might be that everything is fine: Privacy is a value worth protecting, and if there are chinks in the armor, well, that is to be expected. And AI to improve health is important, but if protecting privacy degrades its capacity somewhat, well, that's okay too.

This reaction seems wrong, though; the status quo is hard to defend. In particular, if health privacy is worth defending, then why limit those defenses to the narrow set of actors and data covered by HIPAA, as the United States largely does? HIPAA's outdated focus on covered entities and its safe harbor for "deidentified" data leave too much for manipulation, if health privacy protection is the goal.

SLATE (Apr. 4, 2014, 1:23 PM), https://slate.com/business/2014/04/apple-vs-android-developers-see-a-socioeconomic-divide.html [https://perma.cc/3UFC-X56L] ("The rich, it seems, use iPhones while the poor tweet from Androids.").

[57] *See*, *e.g.*, Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletarì, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N. Galtier, Bennett A. Landman, Klaus Maier-Hein, Sébastien Ourselin, Micah Sheller, Ronald M. Summers, Andrew Trask, Daguang Xu, Maximilian Baust & M. Jorge Cardoso, *The Future of Digital Health with Federated Learning*, 3 NPJ DIGIT. MED. 119 (2020) (describing federated efforts as a way to train models without centralizing data); Fadila Zerka, Samir Barakat, Sean Walsh, Marta Bogowicz, Ralph T. H. Leijenaar, Arthur Jochems, Benjamin Miraglio, David Townend & Philippe Lambin, *Systematic Review of Privacy-Preserving Distributed Machine Learning from Federated Databases in Health Care*, 4 JCO CLINICAL CANCER INFORMATICS 184, 184–94 (2020) (reviewing the field of federated machine learning).

[58] Price, *Contextual Bias*, *supra* note 43, at 110–13.

[59] Roger Allan Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOMM. & TECH. L. REV. 1, 29–31 (2016).

The status quo also problematically privileges commercial entities over academic and nonprofit researchers. As Kayte Spector-Bagdady has pointed out, academic medical centers in particular face substantially more restrictions than other collectors of big data or AI developers: They are typically "covered entities" subject to HIPAA privacy requirements and are also recipients of federal grant funding subject to informed consent requirements.[60] Making research and development harder for noncommercial entities represents an odd policy position that is difficult to justify.

Right now, individuals theoretically have some control over their health data and some privacy protections—but those protections and control are largely illusory given the many possible avenues of compromise. Rather, privacy protections have perverse and unequal effects in determining who gets seen by the system and how data can be used to develop new understanding and improve that system. Tinkering with HIPAA to smooth out its inequalities and patchiness is a first step, but only a first step.

More generally, getting privacy right while tapping the power of big data and AI to improve the health system requires a broader bargain between patients and the health system. The right approach may be a communitarian one, rather than the individualistic focus largely dominant today. Giving up some level of individual-centered explicit control over data demands the assurance that those data will be used to improve the health system and that those improvements will be available for everyone, not just for a select few. This new bargain will be complex to shape and implement; academic medical centers attempting to develop responsible learning health systems are beginning to encounter the challenges involved.[61] The relationship between health privacy and the development of big data and health AI is dysfunctional now, but the rewards to getting it right are potentially immense.

---

[60] Kayte Spector-Bagdady, *Governing Secondary Research Use of Health Data and Specimens: The Inequitable Distribution of Regulatory Burden Between Federally-Funded and Industry Research*, J.L. & BIOSCIENCES (forthcoming) (manuscript at 1) (on file with author), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3786853 [https://perma.cc/CW6G-TUPR].

[61] Nancy E. Kass & Ruth R. Faden, *Ethics and Learning Health Care: The Essential Roles of Engagement, Transparency, and Accountability*, 2 LEARNING HEALTH SYS. e10066 (2018), https://onlinelibrary.wiley.com/doi/full/10.1002/lrh2.10066 [https://perma.cc/R4KM-72R8].