11-2019

Data-Informed Duties in AI Development

Frank Pasquale

DATA-INFORMED DUTIES IN AI DEVELOPMENT

*Frank Pasquale**

Law should help direct—and not merely constrain—the development of artificial intelligence (AI). One path to influence is the development of standards of care both supplemented and informed by rigorous regulatory guidance. Such standards are particularly important given the potential for inaccurate and inappropriate data to contaminate machine learning. Firms relying on faulty data can be required to compensate those harmed by that data use—and should be subject to punitive damages when such use is repeated or willful. Regulatory standards for data collection, analysis, use, and stewardship can inform and complement generalist judges. Such regulation will not only provide guidance to industry to help it avoid preventable accidents. It will also assist a judiciary that is increasingly called upon to develop common law in response to legal disputes arising out of the deployment of AI.

INTRODUCTION

Corporations will increasingly attempt to substitute artificial intelligence (AI) and robotics for human labor.[1] This evolution will create novel situations for tort law to address. However, tort will only be one of several types of law at play in the deployment of AI. Regulators will try to forestall problems by developing licensing regimes and product standards. Corporate lawyers will attempt to deflect liability via contractual arrangements.[2] The interplay of tort, contract, and regulation will not

* Piper & Marbury Professor of Law, University of Maryland. I would like to thank the *Columbia Law Review* staff for careful editing of the piece, as well as those commenting on the piece at the *Columbia Law Review* Symposium "Common Law for the Age of AI" (Colleen Chien and Olga Russakovsky). I also wish to thank attendees at the University of Melbourne's "Digital Citizen" conference and participants in a workshop at Data & Society entitled "Algorithms on the Shop Floor." I am particularly grateful to Taylor M. Cruz, Madeleine Elish, and Emanuel Moss for their comments at the Data & Society Workshop. Colleagues at the University of Maryland have also been generous with their time and expertise. Responsibility for errors or omissions, of course, rests with me.

1. See, e.g., James Manyika, Susan Lund, Michael Chui, Jacques Bughin, Jonathan Woetzel, Parul Batra, Ryan Ko & Saurabh Sanghvi, McKinsey & Co., Jobs Lost, Jobs Gained: What the Future of Work Will Mean for Jobs, Skills, and Wages 1 (2017), https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Future%20of%20Organization/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Report-December-6-2017.ashx [https://perma.cc/XCF4-JJPC] (describing the far-reaching impact that automation will have on the global workforce).

2. This is already a common practice in the digital economy. See, e.g., Timothy J. Calloway, Cloud Computing, Clickwrap Agreements, and Limitation on Liability Clauses: A Perfect Storm?, 11 Duke L. & Tech. Rev. 163, 173 (2012) (describing a proliferation of limitation of liability clauses); Aaron T. Chiu, Note, Irrationally Bound: Terms of Use

just allocate responsibility ex post, spreading the costs of accidents among those developing and deploying AI, their insurers, and those they harm. This matrix of legal rules will also deeply influence the development of AI, including the industrial organization of firms, and capital's and labor's relative share of productivity and knowledge gains.

Despite these ongoing efforts to anticipate the risks of innovation, there is grave danger that AI will become one more tool for deflecting liability, like the shell companies that now obscure and absorb the blame for much commercial malfeasance.[3] The perfect technology of irresponsible profit would be a robot capable of earning funds for a firm, while taking on the regulatory, compliance, and legal burden traditionally shouldered by the firm itself. Any proposal to grant AI "personhood" should be considered in this light.[4] Moreover, both judges and regulators should begin to draw red lines of responsibility and attribution now, while the technology is still nascent.[5]

Licenses and the Breakdown of Consumer Rationality in the Market for Social Network Sites, 21 S. Cal. Interdisc. L.J. 167, 195 (2011) (describing the use of "disclaimers of liability" in social media network use agreements). For a practical example of how contracts are used to deflect, allocate, or redirect liability in the construction industry, see generally Patricia D. Galloway, The Art of Allocating Risk in an EPC Contract to Minimize Disputes, Construction Law., Fall 2018, at 26 (discussing risk allocation in engineering, procurement, and construction (EPC) contracts). In the health care context, "hold harmless" clauses can deflect liability from software providers. See Ross Koppel, Uses of the Legal System that Attenuate Patient Safety, 68 DePaul L. Rev. 273, 275–76 ("The 'hold harmless' clause in EHR [Electronic Health Record] contracts functions to prevent vendors from being held responsible for errors in their software even if the vendor has been repeatedly informed of the problem and even if the problem causes harm or death to patients.").

3. As leading AI ethics expert Joanna Bryson has explained:

> Many of the problems we have in the world today come from people trying to evade the accountability of democracies and regulatory bodies. And AI would be the ultimate shell company. If AI is human-like, the argument goes, then you can use human justice on it. But that's just false. You can't even use human justice against shell companies. And there's no way to build AI that can actually care about avoiding corruption or obeying the law. So it would be a complete mistake—a huge legal, moral and political hazard—to grant rights to AI.

Fraser Myers, AI: Inhuman After All?, Spiked-Online (June 14, 2019), https://www.spiked-online.com/2019/06/14/ai-inhuman-after-all/ [https://perma.cc/A26G-YEX4] (conducting an interview with Bryson).

4. See Joanna J. Bryson, Mihailis E. Diamantis & Thomas D. Grant, Of, for, and by the People: The Legal Lacuna of Synthetic Persons, 25 Artificial Intelligence & L. 273, 273 (2017) ("We review the utility and history of legal fictions of personhood, discussing salient precedents where such fictions resulted in abuse or incoherence. We conclude that difficulties in holding 'electronic persons' accountable when they violate the rights of others outweigh the . . . moral interests that AI legal personhood might protect.").

5. Some may argue it is already too late, thanks to the power of leading firms in the AI space. However, there have been many recent efforts to understand and curb the worst effects of such firms. The U.S. government has demonstrated an interest in keeping large tech companies in line. For example, Facebook is currently facing a $5 billion fine from the FTC, a $100 million fine from the SEC, and an FTC antitrust investigation. Ian Sherr,

It may seem difficult to draw such red lines, because both journalists and technologists can present AI as a technological development that exceeds the control or understanding of those developing it.[6] However, the suite of statistical methods at the core of technologies now hailed as AI has undergone evolution, not revolution.[7] Large new sources of data have enhanced its scope of application, as well as technologists' ambitions.[8] But the same types of doctrines applied to computational sensing, prediction, and actuation in the past can also inform the near future of AI advance.[9]

A company deploying AI can fail in many of the same ways as a firm using older, less avant-garde machines or software. This Essay focuses on one particular type of failing that can lead to harm: the use of inaccurate or inappropriate data in training sets for machine learning. Firms using faulty data can be required to compensate those harmed by that data use—and should be subject to punitive damages when such faulty data

---

Facebook's $5 Billion FTC Fine Is Just the Start of Its Problems, CNET (July 25, 2019), https://www.cnet.com/news/facebooks-5-billion-ftc-fine-is-just-the-start-of-its-problems/ (on file with the *Columbia Law Review*). The Department of Justice is also reviewing tech companies for antitrust issues. Brent Kendall, Justice Department to Open Broad, New Antitrust Review of Big Tech Companies, Wall St. J. (July 23, 2019), https://www.wsj.com/articles/justice-department-to-open-broad-new-antitrust-review-of-big-tech-companies-11563914235 (on file with the *Columbia Law Review*). In response, tech companies, such as Facebook and Google, have expanded their lobbying capacity. See Cecilia Kang & Kenneth P. Vogel, Tech Giants Amass a Lobbying Army for an Epic Washington Battle, N.Y. Times (June 5, 2019), https://www.nytimes.com/2019/06/05/us/politics/amazon-apple-facebook-google-lobbying.html (on file with the *Columbia Law Review*).

6. See, e.g., Will Knight, The Dark Secret at the Heart of AI, MIT Tech. Rev. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [https://perma.cc/V3LF-KBLD] (describing Nvidia's experimental autonomous car as having a "mysterious mind" unable to be understood by those designing it); David Weinberger, Our Machines Now Have Knowledge We'll Never Understand, WIRED (Apr. 18, 2017), https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/ [https://perma.cc/FW94-L2BE] ("This infusion of alien intelligence is bringing into question the assumptions embedded in our long Western tradition.").

7. See, e.g., Best Practice AI, Evolution, Not Revolution: What the Bestpractice.ai Library Tells Us About the State of AI (Part 1), Medium (Sept. 17, 2018), https://medium.com/@bestpracticeAI/evolution-not-revolution-what-the-bestpractice-ai-library-tells-us-about-the-state-of-ai-part-1-f488b29add0b [https://perma.cc/VB86-544K] (describing findings from the development of Bestpractice.ai, a library of AI use cases and case studies).

8. See generally Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz & Zoe Bauer, Artificial Intelligence Index: 2018 Annual Report (2018), http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf [https://perma.cc/3PWE-B7Z8] (presenting data suggesting that the number of patents and academic papers involving AI, among other metrics, have grown rapidly).

9. Notable recent U.S. work in this vein includes Bryan Casey, Robot Ipsa Loquitur, Geo. L.J. (forthcoming 2019) (manuscript at 8–11), https://ssrn.com/abstract=3327673 (on file with the *Columbia Law Review*) (arguing that extant forms of liability should apply to robotics and thus many of the forms of AI that comprise the information processing of such robotics and can address many of the problems posed by such technology).

collection, analysis, and use is repeated or willful. Skeptics may worry that judges and juries are ill-equipped to make determinations about appropriate data collection, analysis, and use. However, they need not act alone—regulation of data collection, analysis, and use already exists in other contexts.[10] Such regulation not only provides guidance to industry to help it avoid preventable accidents and other torts. It also assists judges assessing standards of care for the deployment of emerging technologies. The interplay of federal regulation of health data with state tort suits for breach of confidentiality is instructive here: Egregious failures by firms can not only spark tort liability but also catalyze commitments to regulation to prevent the problems that sparked that liability, which in turn should promote progress toward higher standards of care.[11]

Preserving the complementarity of tort law and regulation in this way (rather than opting to radically diminish the role of either of these modalities of social order, as premature preemption or deregulation might do) is wise for several reasons. First, this hybrid model expands opportunities for those harmed by new technologies to demand accountability.[12] Second, the political economy of automation will only fairly distribute expertise and power if law and policy create ongoing incentives for individuals to both understand and control the AI supply chain and AI's implementation. Judges, lawmakers, and advocates must avoid developing legal and regulatory systems that merely deflect responsibility, rather than cultivate it, lest large firms exploit well-established power imbalances to burden consumers and workers with predictable harms arising out of faulty data.

I. PROBLEMS CAUSED BY INACCURATE AND INAPPROPRIATE DATA

At its best, tort law rectifies wrongs (retrospectively) and enables persons to better plan their lives (prospectively).[13] This Part discusses some classic wrongs addressed by tort law and how the rise of AI, including the rhetoric surrounding it, may unnecessarily complicate adjudication arising out of them. To clarify some critical issues of duty and causation, litigants and courts should begin to focus on questions of inaccurate and inappropriate data, given the importance of data to the development of AI.

The duties of care prescribed by tort are reassuring aspects of a just social order. If a person is injured in a car accident by a negligent driver, courts should ensure some compensatory (and potentially punitive) damages payable by the tortfeasor (or their insurer) to ensure, as well as

10. See infra Part II.

11. See infra Part II.

12. See Mary L. Lyndon, Tort Law and Technology, 12 Yale J. on Reg. 137, 143 (1995) ("The liability system supplements regulation.").

13. See Melvin Aron Eisenberg, The Nature of the Common Law 4–5, 47–48 (1st ed. 1988).

possible, that the plaintiff is returned to the state of financial and physical health they would have enjoyed before the accident.[14] In the medical context, malpractice law is designed to give patients reassurance that if their physician falls below a standard of care, a penalty will be imposed and some portion of it dedicated to the recovery of the patient.[15]

The machines used by drivers and doctors are also subject to forms of tort liability: for example, in case they are negligently manufactured or defective by design.[16] These doctrines should have renewed relevance as new technologies of diagnosis and prediction arise in both general and specialty medical care. While AI applications promise many advances, they also create new risks.

Consider the rise of clinical decision support software for dermatologists. As the *Atlantic* recently reported, "A study that tested machine-learning software in dermatology, conducted by a group of researchers primarily out of Germany, found that 'deep-learning convolutional neural networks,' or CNN, detected potentially cancerous skin lesions better than the 58 dermatologists included in the study group."[17] To the extent such AI is continually validated, it may well become part of the standard of care for many tasks now performed by physicians.[18] However, the mere fact that a technology is better *in general* does not mean that it is optimal for all cases. In the case of facial recognition, there is a well-documented failure of AI systems to recognize the faces of persons of color, relative to its ability to recognize white persons' faces.[19] Many scholars have raised similar concerns with respect to racial disparities in health care in the

---

14. Cf. Stuart M. Speiser, Charles F. Krause, Alfred W. Gans & Monique C. M. Leahy, American Law of Torts § 8:1 (Mar. 2019 Update) (describing the types of redress available to plaintiffs in a tort action).

15. Alex Stein, Toward a Theory of Medical Malpractice, 97 Iowa L. Rev. 1201, 1203, 1209 (2012) ("Under the prevalent doctrine, a doctor commits malpractice when he treats a patient in a way that deviates from the norms established by the medical profession. The applicable norms flow from the accepted, or customary, medical practice: the ways in which similarly situated medical practitioners treat patients."). I introduce the topic with examples from transport and health in part because these fields are among the most affected, or likely to be affected, by advances in AI.

16. See, e.g., Adams v. Toyota Motor Corp., 867 F.3d 903, 917 (8th Cir. 2017) (concluding that the evidence supported a jury verdict finding the manufacturer liable for deaths and injuries of persons involved in the collision in family members' products liability action based on a design defect).

17. Angela Lashbrook, AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind, Atlantic (Aug. 16, 2018), https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/ [https://perma.cc/NLC2-VCFS].

18. A. Michael Froomkin, Ian Kerr & Joelle Pineau, When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning, 61 Ariz. L. Rev. 33, 35, 61–63 (2019).

19. See Tim Simonite, The Best Algorithms Struggle to Recognize Black Faces Equally, WIRED (July 22, 2019), https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/ [https://perma.cc/QQ4J-XBMB].

United States.[20] Physicians and computer scientists are already concerned that skin anomaly–detecting software may fail to work for African Americans and other minority groups in the United States as well as it does for white patients.[21]

Such problems are not new. In many cases, AI is little more than a better-marketed form of statistics, and consulting statistics has long been a part of medical practice.[22] AI is but one of many steps taken over the past two decades to modernize medicine with a more extensive evidence base.[23] Commentators have seized on predictive analytics, big data, artificial intelligence, machine learning, and deep learning as master metaphors for optimizing system performance.[24] Thus literature on each of these areas can illuminate the path forward for identifying problematic data in AI. Moreover, an emerging literature on the limits of AI (including lack of reproducibility, narrow validity, overblown claims, and opaque data) should also inform legal standards.[25]

---

20. See, e.g., Dorothy Roberts, Fatal Invention: How Science, Politics, and Big Business Re-Create Race in the Twenty-First Century 81–103 (2012). See generally Dayna Bowen Matthew, Just Medicine: A Cure for Racial Inequality in American Health Care (2015) (examining racial health disparities through the lens of implicit bias).

21. Adewole S. Adamson & Avery Smith, Machine Learning and Health Care Disparities in Dermatology, 154 JAMA Dermatology 1247, 1247 (2018). A cognate problem has arisen in genomics. See Eric Topol & Kai Fu Lee, It Takes a Planet, 37 Nature Biotechnology 858, 859 (2019) ("AI algorithmic development and validation requires diverse and massive datasets. There is little evidence for saturation but plenty of examples of misleading outputs when the data inputs are limited or venue specific.").

22. See Meredith Broussard, Artificial Unintelligence: How Computers Misunderstand the World 32 (2018) ("Narrow AI is statistics on steroids.").

23. See Inst. of Med. Roundtable on Evidence-Based Medicine, The Learning Healthcare System: Workshop Summary 81 (LeighAnne Olsen, Dara Aisner & J. Michael McGinnis eds., 2007), https://www.ncbi.nlm.nih.gov/books/n/nap11903/pdf/ [https://perma.cc/3VYA-M3S4] ("An essential component of the learning healthcare system is the capacity to continually improve approaches to gathering and evaluating evidence, taking advantage of new tools and methods.").

24. See, e.g., Martin Ford, Architects of Intelligence 4 (2018) (describing deep learning); Viktor Mayer-Schönberger & Kenneth Cukier, Big Data: A Revolution that Will Transform How We Live, Work, and Think 7 (2013) ("Big data marks the beginning of a major transformation."); Nils J. Nilsson, The Quest for Artificial Intelligence 415 (2010) (describing reinforcement learning).

25. Eric Topol, Deep Medicine 94 (2019) (citing concerns about "cherry-picking results or lack of reproducibility"); danah boyd & Kate Crawford, Critical Questions for Big Data, 15 Info., Comm. & Soc'y 662, 666–68 (2012) (describing how claims of objectivity and accuracy in big data can be misleading); Matthew Zook, Solon Barocas, danah boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson & Frank Pasquale, Editorial, Ten Simple Rules for Responsible Big Data Research, PLOS Computational Biology, Mar. 30, 2017, at 1, 2, https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1005399&type=printable (on file with the *Columbia Law Review*) (identifying similar limits).

A. *Inaccurate Data*

In 2012, law professor Sharona Hoffman and computer scientist Andy Podgurski analyzed some common problems in then-emerging uses of big data in healthcare.[26] A great deal of the data that is now set to inform AI applications in healthcare is "generally observational, not experimental, and hence treatments and exposures are not assigned randomly. This makes it much more difficult to ensure that causal inferences are not distorted by systematic biases."[27] Dr. Dhruv Kullar gives a good example of the dangers of these dynamics:

> In medicine, unchecked A.I. could create self-fulfilling prophesies that confirm our preexisting biases, especially when used for conditions with complex trade-offs and high degrees of uncertainty. If, for example, poorer patients do worse after organ transplantation or after receiving chemotherapy for end-stage cancer, machine learning algorithms may conclude such patients are less likely to benefit from further treatment—and recommend against it.[28]

There are several problems with basing treatment on socioeconomic status. A skilled medical practitioner should be interested in *why* poorer patients are doing worse, not simply *that* they are.[29] Perhaps they have a harder time accessing follow-up care or healthy food. The proper response in that case is not to allow poverty to reduce the priority of a patient for a transplant. Rather, it is to invest in transportation, nutritional advice and subsidies, and other social supports that will promote a more

---

26. Sharona Hoffman & Andy Podgurski, Big Bad Data: Law, Public Health, and Biomedical Databases, 41 J.L. Med. & Ethics (Spring Supp.) 56, 56 (2013).

27. Id. at 57.

28. Dhruv Khullar, Opinion, A.I. Could Worsen Health Disparities, N.Y. Times (Jan. 31, 2019), https://www.nytimes.com/2019/01/31/opinion/ai-bias-healthcare.html (on file with the *Columbia Law Review*). As Judea Pearl and Dana MacKenzie have shown, adding accounts of causation via diagrams and other intuitive explanatory tools can help professionals avoid such mistakes. Judea Pearl & Dana MacKenzie, The Book of Why: The New Science of Cause and Effect 13, 39–46 (2018). This is one reason why the European Union has adopted rules designed to promote explainable AI. See High-Level Expert Grp. on Artificial Intelligence, European Comm'n, Ethics Guidelines for Trustworthy AI 21–22 (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 [https://perma.cc/7BKM-VDHP].

29. As Hoffman and Podgurski put it:

> Confounding bias is a systematic error that occurs because there exists a common cause of the treatment/exposure variable and the outcome variable. For example, socioeconomic factors may be confounders because low income may cause individuals to choose sub-optimal, inexpensive treatments and may also separately lead to deteriorated health because of stress or poor nutrition. A failure to account for socioeconomic status may thus skew study results.

Hoffman & Podgurski, supra note 26, at 58 (footnote omitted).

successful transplant.[30] The main problem with the example Khullar gives is that poverty itself is not a direct cause of the bad medical outcomes.[31] Rather, there are intervening causes. AI scholars have long addressed this problem. For example, Judea Pearl and Dana MacKenzie have insisted that a knowledge of causation—*how* an alleged effect generates a cause—is crucial to genuine advances in AI.[32]

Moreover, even if it turns out that, ceteris paribus, poorer individuals simply do not do as well as others after transplants (surviving a shorter period of time, or with worse comorbidities and sequelae of the procedure), that fact alone would not dictate any particular change in their priority for organ transplantation. Society may decide that a thoroughgoing equality of access is the proper baseline for access to scarce organs, even if such allocation rules fail to maximize quality-adjusted life years (QALYs) or similar outcome metrics.[33]

Hoffman and Podgurski also point out the inadequacies of some data, especially those captured on the fly by doctors and nurses who already have more than enough to do on their shifts.[34] Electronic health record (EHR) systems may use different abbreviations: "Different systems may use different terminology to mean the same thing or the same terminology to mean different things. For example, the abbreviation 'MS' can mean 'mitral stenosis,' 'multiple sclerosis,' 'morphine sulfate,' or 'magnesium sulfate.'"[35] At present, the job of correcting (or throwing out) bad data, as well as related tasks of semantic harmonization and standardization, is often treated as secondary or menial.[36] But at a certain level of prevalence, such errors could be disastrous. Researchers must take into account measurement biases, which "are generated by errors in measurement and data collection resulting from faulty equipment or software or from human error."[37] Data are always socially shaped.[38] To

30. See, e.g., Mary Simmerling, Beyond Scarcity: Poverty as a Contraindication for Organ Transplantation, 9 AMA J. Ethics 441, 442–44 (2007) (examining the financial burdens of post-transplant medications on the uninsured, the underinsured, and the poor).

31. See Khullar, supra note 28.

32. Pearl & MacKenzie, supra note 28, at 1–21. See generally Judea Pearl, Causal Inference in Statistics: An Overview, 3 Stat. Surv. 96 (2009) (discussing advances in statistical research that facilitate solving causal questions).

33. See Jon Elster, Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens 22, 35–38 (1992) (discussing diverse normative bases for allocation decisions).

34. Hoffman & Podgurski, supra note 26, at 57.

35. Id. at 57.

36. See Lilly Irani, Justice for "Data Janitors," Pub. Books (Jan. 15, 2015), https://www.publicbooks.org/justice-for-data-janitors/ [https://perma.cc/JLY6-Y6URt] (describing the work done by human "data janitors" to parse information that artificial intelligence systems are not capable of differentiating).

37. Hoffman & Podgurski, supra note 26, at 58.

avoid troubling outcomes downstream, law must incentivize health care providers to ensure that data providers take the time and effort necessary to address well-known biases and shortcomings of data.

In the case of automobiles, similar problems may emerge. There may be certain individuals that a collision avoidance detection system is less likely to identify as persons.[39] Operators of autonomous cars may deploy humans as a backup, to ensure the data a car is reacting to are accurate, but even such a failsafe may itself be blameworthy if improperly applied. Human–computer interaction research has revealed that such "backup" roles are notoriously difficult to perform well, particularly in contexts in which attention is only required rarely and sporadically.[40]

B. *Inappropriate Data*

While earlier versions of AI, such as expert systems, were primarily rules based, data drives modern machine learning.[41] As recent controversies over predictive policing have shown, data can be unfairly unrepresentative: If minority neighborhoods have been overpoliced in the past, more crime will have been found in them than would be found in other neighborhoods, ceteris paribus.[42] Similarly, a firm that primarily hired

---

38. See Lisa Gitelman & Virginia Jackson, Introduction, *in* "Raw Data" Is an Oxymoron 1, 2–6 (Lisa Gitelman ed., 2013) (arguing that data are not inherently neutral but rather constructed and gathered in ways that are shaped by academic disciplines). See generally Taylor M. Cruz, The Making of a Population: Challenges, Implications, and Consequences of the Quantification of Social Difference, 174 Soc. Sci. & Med. 79 (2017) (discussing how the process of gathering population data imposes implicit categorical assumptions on a heterogenous population).

39. Benjamin Wilson, Judy Hoffman & Jamie Morgenstern, Predictive Inequity in Object Detection, arXiv (Feb. 21, 2019), https://arxiv.org/pdf/1902.11097.pdf (on file with the *Columbia Law Review*) (identifying potential for object detection technology to fail to detect people with darker skin tones).

40. See, e.g., David A. Mindell, Our Robots, Ourselves 201–02 (2015) (describing the difficulties and failures associated with human operators serving as a backup in the event of failures by AI-driven systems such as autonomous vehicles); Madeleine Clare Elish, Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction, 5 Engaging Sci., Tech., & Soc'y 40, 52–55 (2019) (noting the difficulty of distributing responsibility and agency between a self-driving car and its safety driver).

41. Pedro Domingos, The Master Algorithm 7 (2015).

42. See Angèle Cristin, Predictive Algorithms and Criminal Sentencing, *in* The Decisionist Imagination: Sovereignty, Social Science, and Democracy in the 20th Century 272, 279–80 (2019) ("When predictive algorithms identify 'hot spot' crime zones (usually low-income African American neighborhoods), policemen are more likely to patrol in these neighborhoods and arrest people who will later be convicted. . . . This data will later be entered into the algorithm, thus producing a feedback loop.").

male managers in the past may end up developing AI hiring mechanisms that correlate success with gender, as opposed to actual job performance.[43]

Activists and authors are now exposing numerous examples of problematic data sets. For example, Caroline Criado Perez has explained how data sets often do not adequately represent women, with very troubling results.[44] In too much medical research and pedagogy, for instance, maleness is assumed as a default. As Perez asks, "There are still vast medical gender data gaps to be filled in, but the past twenty years have demonstrably proven that women are not just smaller men: male and female bodies differ down to a cellular level. So why aren't we teaching this?"[45]

Data may also be illegally obtained and therefore inappropriate for use. For example, an AI hiring algorithm might incorporate breached medical records that help it predict an applicant's health issues. Even if such health issues would impair the applicant's job performance, this data use is suspect. Thanks to trade secrecy, it may be difficult to detect or litigate.[46] Nevertheless, litigants are becoming increasingly sophisticated at unearthing the true bases of decisionmaking, and no firm should be entitled to hide the use of illegally obtained data.[47]

---

43. See, e.g., Gideon Mann & Cathy O'Neil, Hiring Algorithms Are Not Neutral, Harv. Bus. Rev. (Dec. 9, 2016), https://hbr.org/2016/12/hiring-algorithms-are-not-neutral [https://perma.cc/BA6V-492D] ("When humans build algorithmic screening software, they may unintentionally determine which applicants will be selected or rejected based on outdated information—going back to a time when there were fewer women in the workforce, for example—leading to a legally and morally unacceptable result."); see also Miranda Bogen & Aaron Reike, Upturn, Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias 8–9 (2018), https://www.upturn.org/static/reports/2018/hiring-algorithms/files/Upturn%20–%20Help%20Wanted%20-%20An%20Exploration%20of%20Hiring%20Algorithms,%20Equity%20and%20Bias.pdf [https://perma.cc/5T6U-4QL4] (describing examples of potential bias in predictive hiring tools).

44. See generally Caroline Criado Perez, Invisible Women: Data Bias in a World Designed for Men (2019) (examining the "gender data gap").

45. Id. at 199.

46. See Sonia K. Katyal, The Paradox of Source Code Secrecy, 104 Cornell L. Rev. (forthcoming 2019) (manuscript at 104–05) (footnote omitted), https://ssrn.com/abstract=3409578 (on file with the *Columbia Law Review*) ("At their core, these automated systems often implicate central issues of due process, criminal (and civil) justice, and equal protection. Yet, because their inner workings are often protected as trade secrets, they can remain entirely free from public scrutiny."); Frank Pasquale, Digital Star Chamber, Aeon (Aug. 18, 2015), https://aeon.co/essays/judge-jury-and-executioner-the-unaccountable-algorithm [https://perma.cc/56VN-M3AT] ("Protected by trade secrecy, many algorithms remain impenetrable to outside observers.").

47. Concededly, the Supreme Court has offered a First Amendment imprimatur for reuse of illegally obtained information in some contexts. See, e.g., Bartnicki v. Vopper, 532 U.S. 514, 517–18 (2001) (finding the First Amendment protects "speech that discloses the contents of an illegally intercepted communication"). However, that defense is conditioned on a "public interest" finding, id. at 540 (Breyer, J., concurring), and secret categorization or ranking of applicants should not qualify. See Frank Pasquale, Reforming the Law of Reputation, 47 Loy. U. Chi. L.J. 515, 529–30 (2015) (discussing the limits of *Bartnicki*).

Finally, certain inferences can become data that are extraordinarily suspect.[48] Consider, for instance, the rise of efforts to correlate persons' facial features and voices with illness, risk, or aptitude. Machine learning researchers have stirred controversy by claiming that our faces may reveal our sexual orientation and intelligence.[49] Using a database of prisoners' faces, some have even developed stereotypes of criminal features, reprising long-discredited physiognomy and phrenology.[50] A firm has claimed that it can deploy facial recognition to spot pedophiles and terrorists.[51] These inferences are deeply troubling. When such methods of pattern recognition are used to classify persons, they overstep a fundamental boundary between objective analysis and moral judgment. And when such moral judgments are made, persons categorized by the judgements deserve a chance to understand and contest them.

When a data set is not representative of the group it is used to classify, any results based on it should be clearly qualified. For example, a machine learning classifier may properly be said to succeed in classifying some percentage of faces *in its data set* in certain ways. But it should not be deployed as a potential classifier for all persons unless and until we have some sense of how the training set maps to the full set of persons it ostensibly classifies. As Dan McQuillan warns, machine learning often makes powerful predictions, "prompting comparisons with science. But rather than being universal and objective, it produces knowledge that is irrevocably entangled with specific computational mechanisms and the data used for training."[52] Both lawmakers and policymakers should hold users of such data sets responsible for making predictable errors based

48. For a fuller account of the problem of troubling or inappropriate inferences, see Sandra Wachter & Brent Mittelstadt, A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI, 2019 Colum. Bus. L. Rev. 494, 499–505 (2019).

49. See, e.g., Sam Levin, Face-Reading AI Will Be Able to Detect Your Politics and IQ, Professor Says, Guardian (Sept. 12, 2017), https://www.theguardian.com/technology/2017/sep/12/artificial-intelligence-face-recognition-michal-kosinski [https://perma.cc/X4HD-KNAK].

50. Sam Biddle, Troubling Study Says Artificial Intelligence Can Predict Who Will Be Criminals Based on Facial Features, The Intercept (Nov. 18, 2016), https://theintercept.com/2016/11/18/troubling-study-says-artificial-intelligence-can-predict-who-will-be-criminals-based-on-facial-features/ [https://perma.cc/X3SN-QAEU]. It was later suggested that the sources of images used for the study may have been a key factor explaining its results. @davidjayharris, Twitter (Mar. 7, 2019), https://twitter.com/davidjayharris/status/1103636069180993537 [https://perma.cc/AKD5-TGPT].

51. Matt McFarland, Terrorist or Pedophile? This Start-Up Says It Can Out Secrets by Analyzing Faces, Wash. Post (May 24, 2016), https://www.washingtonpost.com/news/innovations/wp/2016/05/24/terrorist-or-pedophile-this-start-up-says-it-can-out-secrets-by-analyzing-faces/ (on file with the *Columbia Law Review*).

52. Dan McQuillan, People's Councils for Ethical Machine Learning, Soc. Media + Soc'y, Apr.–June 2018, at 1, 1, https://doi.org/10.1177/2056305118768303 [https://perma.cc/9CS7-4AAK].

on defective data sets, particularly if they fail to disclose the limitations of the data used.

II. COMPLEMENTARY TORT AND REGULATORY REGIMES

Tort law has evolved to handle the changing risks and affordances of new technologies.[53] However, judges alone cannot adequately respond to the new challenges posed by AI. Objective sources of information on best practices in data science are necessary as well. Expert agencies are particularly well positioned to analyze and articulate emerging industry standards, which should inform judicial determinations of standards of care. This Part describes emerging doctrinal and regulatory approaches that suggest data-driven duties for the developers of artificial intelligence. This type of data stewardship serves two purposes: ex ante, to ensure that the training data for machine learning adequately reflects the domain it governs or affects, and ex post, to detect anomalies and remedy them before they cause great harm.[54] Developing and maintaining these duties will be crucial to promoting just and humane advances in AI.

As Professors Dan Dobbs, Paul Hayden, and Ellen Bublick explain, "A tort is conduct that constitutes a legal wrong and causes harm for which courts will impose civil liability."[55] Negligence, vicarious liability, strict liability, and product liability regimes all may be relevant to future torts attributable to AI.[56] In the realm of negligence, the plaintiff generally must prove that the defendant caused the plaintiff's injury, owed a duty of care to the plaintiff, and breached that duty.[57] There are also diverse vicarious liability doctrines, each hinging on factors that include the degree of control an entity has over the direct cause of harm.[58] As

---

53. Donald G. Gifford, Technological Triggers to Tort Revolutions: Steam Locomotives, Autonomous Vehicles, and Accident Compensation, 11 J. Tort L. 71, 143 (2018).

54. Cf. Kristin Madison, Health Regulators as Data Stewards, 92 N.C. L. Rev. 1605, 1607–09 (2014) (arguing that regulators, as data stewards, bear a duty to serve as both an aggregator and editor of big health care data in order to ensure both the integrity of data collection and informed, continuous evaluation of regulation).

55. Dan B. Dobbs, Paul T. Hayden & Ellen Bublick, Hornbook on Torts 3 (2d ed. 2016). This basic tort definition is consistent even in civil law countries around the world. See, e.g., Principles of European Tort Law: Text and Commentary tit. 1, art. 1:101 (Eur. Grp. on Tort Law 2005) ("Basic Norm (1) A person to whom damage to another is legally attributed is liable to compensate that damage."); Tort Law of the People's Republic of China, Ministry of Commerce of China (Dec. 26, 2009), http://english.mofcom.gov.cn/article/policyrelease/Businessregulations/201312/20131200432451.shtml [https://perma.cc/5LMK-YLDC] ("Those who infringe on civil rights and interests shall be subject to tort liability according to this Law.").

56. For a useful typology of torts, see the table of contents of Dobbs et al., supra note 55, at xv–xxxi.

57. Dan B. Dobbs, The Law of Torts 269 (2000).

58. Harry Shulman, Fleming James Jr., Oscar S. Gray & Donald G. Gifford, Law of Torts: Cases and Materials 112–30 (5th ed. 2010).

services become more complex, one of the most promising developments in tort law is corporate liability for failure to maintain adequate safety standards.

For example, in one of the leading cases in medical corporate liability, *Thompson v. Nason Hospital*, the Pennsylvania Supreme Court did not allow the responsibility for a bad outcome to dissolve into a mist of contractual relationships among a hospital, its staff, doctors, and the manufacturers of devices that its doctors and staff used.[59] Rather, the *Thompson* court articulated a general duty of a hospital "to ensure the patient's safety and well-being while at the hospital."[60] The court went on to articulate four nonexhaustive dimensions of this general duty to protect safety and well-being:

> (1) a duty to use reasonable care in the maintenance of safe and adequate facilities and equipment; (2) a duty to select and retain only competent physicians; (3) a duty to oversee all persons who practice medicine within its walls as to patient care; and (4) a duty to formulate, adopt and enforce adequate rules and policies to ensure quality care for the patients.[61]

This standard of corporate negligence has much to offer outside of the healthcare setting. One classic theoretical foundation of health law as a distinctive field is the great difference between ordinary consumer markets, on the one hand, and the healthcare field, where information asymmetries and power differentials routinely arise between patients and healthcare providers, on the other.[62] The rise of software and cyber-physical

---

59. 591 A.2d 703, 709 (Pa. 1991).

60. Id. at 707.

61. Id. (citations omitted).

62. As Donald Cohodes argues, medical care can be differentiated from "most other products" in six general ways:

> 1. *Demand for health.* Medical care services are not purchased from any desire for such services in themselves . . . [but instead are] derived from the "demand" for good health.
>
> 2. *Medical care and health.* Medical care is only one determinant of health status, and for most people at most times it is not even a very important determinant. . . .
>
> 3. *Risk.* The need for medical care is unpredictable, requiring expenditures that are irregular and of uncertain magnitude.
>
> 4. *Immediacy.* The need for medical care is often immediate, allowing little time for shopping around and seeking advice or alternatives.
>
> 5. *Lack of Information.* Consumers are usually ignorant of their medical care needs. They cannot possibly obtain the knowledge and training to diagnose their own medical care needs . . . .
>
> 6. *Uncertainty.* Physicians, though highly trained and better able to diagnose needs and prescribed treatment, also are often uncertain about the appropriate services to provide.

Donald R. Cohodes, Where You Stand Depends on Where You Sit: Musings on the Regulation/Competition Dialogue, 7 J. Health Pol. Pol'y & L. 54, 56 (1982).

systems portends a similar increase in complexity, power differentials, and information asymmetry reminiscent of the highly scientific and professionalized medical milieu.[63] Doctrines and approaches developed in the medical setting have already been proposed for other aspects of data governance. For example, health privacy law can serve as a model for the regulation of other data.[64] Jack Balkin and Jonathan Zittrain have proposed that a law of fiduciary duties, itself heavily reliant on the model of doctors' duties to patients, should bind large technology firms with respect to their treatment of data collected from users.[65]

Thompson has been cited many times, and its factors helpfully articulate theories of liability.[66] An elaboration of the corporate negligence standard in a complex environment can illuminate the roles and responsibilities of the developers of artificial intelligence. For example, the first duty (to use reasonable care in the maintenance of safe facilities and equipment) suggests a similar obligation to exercise due care in the selection of sources of data. *Thompson* also reflects in law the conclusions of a larger quality-improvement movement: that it is less important to find particular persons to blame in the case of accidents, than to identify malfunctioning sociotechnical systems of human–computer interaction.[67]

The third *Thompson* factor, regarding adequate supervision, also raises important questions in the context of automation developed in corporate labs and its testing outside of controlled settings. Surveillance techniques are widespread and well-developed.[68] Such technology could

---

63. On the rise of software in ordinary products, see Paul Ohm & Blake Reid, Regulating Software When Everything Has Software, 84 Geo. Wash. L. Rev. 1672, 1676–79 (2016); see also James Grimmelmann, Note, Regulation by Software, 114 Yale L.J. 1719, 1723–24 (2005) (giving "four patterns [that] provide a general methodology for assessing the use of software in a given regulatory context").

64. Frank Pasquale, The Black Box Society: The Secret Algorithms that Control Money and Information 150–51 (2015) (discussing HIPAA standards for consent, security, and accounting of disclosures of health data as a model for other forms of data).

65. Jack M. Balkin, Information Fiduciaries and the First Amendment, 49 U.C. Davis L. Rev. 1183, 1221–25 (2016). But see Lina Khan & David Pozen, A Skeptical View of Information Fiduciaries, 133 Harv. L. Rev. (forthcoming 2019) (manuscript at 6–8), https://ssrn.com/abstract=3341661 (on file with the *Columbia Law Review*) (arguing that creating new fiduciary duties based on information custody is fundamentally incompatible with existing corporate law of fiduciary duties and therefore impossible to implement in the form proposed by Balkin and Zittrain).

66. As of March 15, 2019, Thompson v. Nason Hospital, 591 A.2d 703 (Pa. 1991), has been cited in 198 cases and 273 secondary sources on Westlaw Edge.

67. Lucian L. Leape, Error in Medicine, 272 JAMA 1851, 1853 (1994) (describing the importance of system-level analysis in attribution of blame and prevention of future harms).

68. See, e.g., Karen E.C. Levy, The Contexts of Control: Information, Power, and Truck-Driving Work, 31 Info. Soc'y 160, 160, 164 (2015) (describing how trucking firms have extensively deployed telematics to monitor truck drivers with regard to performance and timekeeping); Steve Kolowich, Behind the Webcam's Watchful Eye, Online Proctoring Takes Hold, Chron. Higher Educ. (Apr. 15, 2013), https://www.chronicle.com/article/

help reduce bias in data collection and promote vigilance among those tasked with overseeing the deployment of AI in sensitive settings. On the other hand, privacy activists may raise concerns if the common law of tort promotes excessive surveillance of workers.[69] Once again, the health care industry has been at the forefront, developing balanced frameworks for the inclusion of surveillance technology in workplaces in which human life is routinely at risk.[70]

III. REGULATORY STANDARDS FOR DATA USE AND REPORTING

If a large proportion of cases involving AI went to trial, reported opinions would serve as a prominent source of guidance for AI vendors and users concerned about safety and effectiveness. However, we can expect that here, as with data security, the prevalence of settlements of disputes will frustrate such evolutionary clarification of duties.[71] In this vacuum, regulators should play a vital role in setting (or at least informing)

Behind-the-Webcams-Watchful/138505 [https://perma.cc/CQ5T-2C74] (describing online proctors that watch students through a webcam to detect cheating); Natasha Singer, Online Test-Takers Feel Anti-Cheating Software's Uneasy Glare, N.Y. Times (Apr. 5, 2015), https://www.nytimes.com/2015/04/06/technology/online-test-takers-feel-anti-cheating-softwares-uneasy-glare.html (on file with the *Columbia Law Review*) (describing software developed to detect cheating during online and computer exam taking).

69. See, e.g., Lewis Maltby, Can They Do That?: Retaking Our Fundamental Rights in the Workplace 16–17 (2009) (describing an example of intrusive surveillance of workers); Ifeoma Ajunwa, Kate Crawford & Jason Schultz, Limitless Worker Surveillance, 105 Calif. L. Rev. 735, 735–36, 772–73 (2017) (describing the trend of increased worker surveillance and exploring possible remedies to protect worker privacy).

70. See generally Clara Berridge, Jodi Halpern & Karen Levy, Cameras on Beds: The Ethics of Surveillance in Nursing Home Rooms, 10 AJOB Empirical Bioethics 55 (2019) (examining survey data on the use of "family-provided cameras" in nursing homes and their legal and ethical implications); Karen Levy, Lauren Kilgour & Clara Berridge, Regulating Privacy in Public/Private Space: The Case of Nursing Home Monitoring Laws, 26 Elder L.J. 323, 326–27 (2019) (comparing "state laws and regulations governing resident-room cameras in nursing homes . . . focus[ing] on how such rules approach and balance the privacy concerns of the multiple relations involved in such contexts, and how legal protections do—and do not—address relationship-specific interests").

71. See William McGeveran, The Duty of Data Security, 103 Minn. L. Rev. 1135, 1144 (2019) ("There are numerous lawsuits about data security, which raise claims under tort, contract, or consumer protection law, among other theories. Courts considering these cases offer hardly any insight into the *content* of the duty of data security, however, because they almost never reach the merits." (footnote omitted)); cf. Owen M. Fiss, Against Settlement, 93 Yale L.J. 1073, 1075, 1078–85 (1984) (complaining of the problems caused by this avoidance). Instead, in the data security context, the Federal Trade Commission has taken the lead. See Woodrow Hartzog and Daniel J. Solove, The FTC and the New Common Law of Privacy, 114 Colum. L. Rev. 583, 585–86 (2014) ("Despite over fifteen years of FTC enforcement, there are hardly any judicial decisions to show for it. The cases have nearly all resulted in settlement agreements. . . . Thus, in practice, FTC privacy jurisprudence has become the broadest and most influential regulating force on information privacy in the United States . . . .").

standards.[72] Though the current Congress is unlikely to establish a new agency, existing statutory authorities already grant extant agencies the power to gather, analyze, and disseminate data that would aid courts' assessments of the proper standard of care in disputes related to AI-informed and AI-performed services.[73] Some of these agencies have also established standards that have informed tort cases in data-related fields, such as privacy law.[74]

A. *Ensuring the Integrity of Inputs*

One purpose of the Health Insurance Portability and Accountability Act's (HIPAA) security requirements is to protect data from hackers or other corrupting influences.[75] A logical extension of this duty is for agencies to set standards for AI vendors and users to verify the quality and accuracy of the data they use.[76] These standards may start at an elementary level. For example, HIPAA best practices dictate that a covered entity both record any source of data it receives and record its transfer of data to other covered entities or business associates.[77] Those recipients of data must in turn do the same.[78] This creates a set of links that makes it easier to trace and then minimize the impact of inaccurate, unrepresentative,

---

72. For a general account of the government role in promoting standardized data, see generally Michal S. Gal & Daniel L. Rubinfeld, Data Standardization, 94 N.Y.U. L. Rev. (forthcoming 2019), https://ssrn.com/abstract=3326377 (on file with the *Columbia Law Review*).

73. See Andrew F. Popper, Gwendolyn M. McKee, Anthony E. Varona, Philip J. Harter, Mark C. Niles & Frank Pasquale, Administrative Law: A Contemporary Approach 1067–134 (3d ed. 2016) (describing the power, and the limits of such power, of U.S. agencies to demand information).

74. See infra sections III.A–.B.

75. See Frank Pasquale, Redescribing Health Privacy: The Importance of Information Policy, 14 Hous. J. Health L. & Pol'y 95, 105–09 (2014) (describing the range of security measures prescribed by HIPAA).

76. See, e.g., Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Schultz & Oscar Schwartz, AI Now Report 2018, at 4–7 (2018), https://ainowinstitute.org/AI_Now_2018_Report.pdf [https://perma.cc/4J3T-TCTR] (discussing the importance of sectoral regulation); cf. Frank Pasquale, Private Certifiers and Deputies in American Health Care, 92 N.C. L. Rev. 1661, 1668–69, 1671–73, 1692 (2014) (describing a broad array of public and private actors that have cooperated in highly technical areas to promote data quality and interoperability in the health care industry).

77. Bill Becker, HIPAA Compliance Best Practices: Questions and Answers to Improve Security and Avoid Penalties, HIPAA J. (May 16, 2017), https://www.hipaajournal.com/hipaa-compliance-best-practices-8809/ [https://perma.cc/GZZ5-HM4L]; Office for Civil Rights, How Are Covered Entities Expected to Determine What Is the Minimum Necessary Information that Can Be Used, Disclosed, or Requested for a Particular Purpose?, HHS: Health Info. Privacy (Dec. 19, 2002), https://www.hhs.gov/hipaa/for-professionals/faq/207/how-are-covered-entities-to-determine-what-is-minimum-necessary/index.html [https://perma.cc/R788-UZRW] (last updated Mar. 14, 2006).

78. See Becker, supra note 77 (discussing best practices for improving data security); Office for Civil Rights, supra note 77 (setting out requirements for minimum data sharing).

or otherwise compromised data.[79] Similar standards should inform the stewardship of data used for machine learning and AI. Federal standards for data protection may, in turn, become part of the standard of care for torts like breach of medical confidentiality.[80]

For a concrete example of why such practices matter, consider how voice recognition software may be more or less accurate with respect to persons with different voices or accents.[81] As of 2020, databases may have a certain level of inclusiveness;[82] by 2025, this is likely to have improved markedly.[83] An AI vendor using the 2020 database in 2025 for mission-critical applications may rightly be faulted for failing to update in light of new knowledge about the limitations of the database. But we would not even know where to look for such a problem if the source of the firm's data was not recorded adequately.[84]

---

79. See Woodrow Hartzog, Chain Link Confidentiality, 46 Ga. L. Rev. 657, 677 (2012) ("The HIPAA Privacy Rules provide that, although only covered entities such as healthcare providers are bound to confidentiality, these entities may not disclose information to their business associates without executing a written contract that places the business associate under the same confidentiality requirements as the healthcare providers."). These protections have been strengthened even further by the Health Information Technology for Economic and Clinical Health Act (HITECH) (and the HIPAA Omnibus Rule of 2013), which impose statutory and regulatory duties on business associates and even their downstream contractors. See Frank Pasquale & Tara Adams Ragone, Protecting Health Privacy in an Era of Big Data Processing and Cloud Computing, 17 Stan. Tech. L. Rev. 595, 609–15 (2014) (describing these duties).

80. See Barry R. Furrow, Thomas L. Greaney, Sandra H. Johnson, Timothy Stoltzfus Jost, Robert L. Schwartz, Brietta R. Clark, Erin C. Fuse Brown, Robert Gatter, Jaime S. King & Elizabeth Pendo, Health Law: Cases, Materials and Problems 201 (8th ed. 2018) (noting that courts have held that "despite the absence of a private right of action under HIPAA, it can inform the applicable standard of care in common law tort cases"); see also Bonney v. Stephens Mem'l Hosp., 17 A.3d 123, 128 (Me. 2011) ("HIPAA standards, like state laws and professional codes of conduct, may be admissible to establish the standard of care associated with a state tort claim . . . ."); Acosta v. Byrum, 638 S.E.2d 246, 253 (N.C. Ct. App. 2006) (describing HIPAA "providing evidence of the duty of care owed . . . with regards to the privacy of plaintiff's medical records"). But see Young v. Carran, 289 S.W.3d 586 (Ky. Ct. App. 2008) (declining to adopt a negligence per se standard); Sheldon v. Kettering Health Network, 40 N.E.3d 661, 664 (Ohio Ct. App. 2015) (same).

81. See Sonia Paul, Voice Is the Next Big Platform, Unless You Have an Accent, WIRED (Mar. 20, 2017), https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/ [https://perma.cc/78TL-PSC9] (reporting on the difficulties associated with creating software that recognizes different accents).

82. See id. (reporting tech companies' efforts to improve the inclusiveness of their accent data); Kyle Wiggers, These Companies Are Shrinking the Voice Recognition 'Accent Gap,' Venture Beat (Aug. 11, 2018), https://venturebeat.com/2018/08/11/using-ai-and-big-data-to-address-the-accent-gap-in-voice-recognition-systems/ [https://perma.cc/F96Z-9FH4] (same).

83. See Paul, supra note 81; Wiggers, supra note 82.

84. This is not a mere hypothetical; I recently had to take to Twitter to learn where the voices for a Google Assistant feature (Duplex) came from. The source was not clearly labeled on the corporate website trumpeting the feature.

Such standards will be resisted. AI vendors will likely push for another approach, simply disclosing potential problems with their data in advance in disclaimers.[85] Perhaps it is the responsibility of the person using the AI, rather than the vendor of AI, to correct for error-prone datasets. However, courts may also find ample precedent for holding vendors responsible. For example, in lawsuits over food poisoning, consumers' "reasonable expectation" of purity and appropriateness of ingredients has been recognized.[86]

Some AI-driven devices may also need to be subjected to the certification and testing now applied (albeit minimally) to electronic health records.[87] Thanks to the HITECH Act of 2009, the Department of Health and Human Services must assure that EHRs meet basic functionality requirements.[88] Failures of EHR vendors to comply with federal health standards have already led to litigation.[89] Given the False Claims Act's

---

85. For entertaining examples of the rhetoric one can expect, see Chris Jay Hoofnagle, Denialists' Deck of Cards: An Illustrated Taxonomy of Rhetoric Used to Frustrate Consumer Protection Efforts (Feb. 9, 2007) (unpublished manuscript), https://ssrn.com/abstract=962462 (on file with the *Columbia Law Review*) (illustrating "a taxonomy of arguments used in denialism" by using "a deck of playing cards to make it more interesting and to emphasize that denialists are engaged in a predictable game to 'do little and delay.'").

86. See Gail Kachadurian McCallion, Note, From the Source to the Mouth: What Can You Reasonably Expect to Find in Your Food?, 5 Fordham Envtl. L.J. 189, 212 (1993) ("The reasonable expectation test asserts that regardless of whether a substance in a food product is natural to an ingredient, liability will lie for injuries caused by the substance where the consumer of the product would not have reasonably expected to find the substance in the product."); see also Richard E. Kaye, Foreign Substance in Food or Beverage, 124 Am. Jur. Proof Facts 3d 91, § 2 (2018) (citing Restatement (Third) of Torts: Prod. Liab. § 7 (Am. Law Inst. 1998)).

87. See, e.g., ONC—Authorized Testing Laboratories (ONC-ATLs), HealthIT.gov (Sept. 24, 2018), https://www.healthit.gov/node/95011 [https://perma.cc/63XV-7A97] (listing the five Authorized Testing Laboratories "accredited by NVLAP and authorized by ONC to test Health IT Modules under the ONC Health IT Certification Program").

88. Health Information Technology Standards, 45 C.F.R. §§ 170.202–170.210 (2019) (providing a detailed set of standards for the use and storage of electronic health information including, for example, encryption and hashing algorithm requirements).

89. See, e.g., Complaint at 1–4, United States ex rel. Delaney v. eClinicalWorks, No. 2:15-CV-00095-WKS (D. Vt. May 1, 2015) (alleging that the defendant failed to comply with federal requirements and that it misrepresented information and failed to disclose flaws in its EHR system in violation of the False Claims Act). The defendant later settled the claim for $155 million. See Press Release, Dep't of Justice, Electronic Health Records Vendor to Pay 155 Million to Settle False Claims Act Allegations (May 31, 2017), https://www.justice.gov/opa/pr/electronic-health-records-vendor-pay-155-million-settle-false-claims-act-allegations [https://perma.cc/59B3-6M5E]; see also Jessica Davis, eClinicalWorks Sued for Nearly $1 Billion for Inaccurate Medical Records, Healthcare IT News (Nov. 17, 2017), https://www.healthcareitnews.com/news/eclinicalworks-sued-nearly-1-billion-inaccurate-medical-records [https://perma.cc/3FFF-NNQV] ("EHR vendor eClinicalWorks has been hit with a class-action lawsuit that alleges . . . that millions of patients have compromised patient records, as eClinicalWorks' software didn't meet meaningful use and certification requirements laid out by the Office of the National Coordinator."); Heather Landi, $1 Billion Class Action Lawsuit Filed Against eClinicalWorks, Healthcare Innovation (Nov. 20, 2017),

(FCA) role in assuring that healthcare providers are treating patients with valid and effective forms of care, this form of liability should be a bellwether specifically for AI vendors contracting with governmental authorities. Consumer protection authorities should also take note.

B. *Ensuring the Transparency of Outputs*

Health regulators have long considered data stewardship a critical role under their statutory mandate.[90] When the federal government began funding EHRs in earnest in 2011, it not only demanded certain basic recordkeeping but also set providers on an ambitious path toward "meaningful use" of information technology—including potentially AI-driven tools like clinical decision support.[91] In 2015, Congress promoted interoperability in the Medicare Access and CHIP Reauthorization Act (MACRA).[92] This drive for interoperability continues to this day, as the Office for the National Coordinator of Health Information Technology and the Centers for Medicare and Medicaid Services have recently announced rulemakings designed to help promote data liquidity.[93]

One key rationale for interoperability is supporting the massive disclosure and reporting requirements mandated pursuant to healthcare finance reforms (covering Advanced Payment Models (APMs) such as Accountable Care Organizations (ACOs), as well as readmissions penalties and bundled payments).[94] It may be very difficult for networks like

---

https://www.hcinnovationgroup.com/clinical-it/news/13029475/1-billion-class-action-lawsuit-filed-against-eclinicalworks [https://perma.cc/EKR7-KPKJ] ("The class action lawsuit alleges ECW falsely represented to its certifying bodies that its software complied with the requirements for certification and the payment of incentives under the MU program, and therefore, caused its users to falsely attest to using a certified EHR technology.").

90. See generally Madison, supra note 54, at 1607–28 (discussing ways the federal government has taken on "the responsibility for protecting the integrity and confidentiality of data" in the health care sector).

91. Frank Pasquale, Grand Bargains for Big Data: The Emerging Law of Health Information, 72 Md. L. Rev. 682, 710–11 (2013) (explaining how the law promotes patient health).

92. Pub. L. No. 114-10, § 106(b), 129 Stat. 87, 138–40 (2015) (codified at 42 U.S.C. § 1395w-4 (2018)) ("The term 'interoperability' means the ability of two or more health information systems or components to exchange clinical and other information . . . to provide access to longitudinal information for health care providers in order to facilitate coordinated care and improved patient outcomes.").

93. Interoperability and Patient Access for Medicare Advantage Organization and Medicaid Managed Care Plans, State Medicaid Agencies, CHIP Agencies and CHIP Managed Care Entities, Issuers of Qualified Health Plans in the Federally-Facilitated Exchanges and Health Care Providers, 84 Fed. Reg. 7610 (proposed Mar. 4, 2019); 21st Century Cares Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program, 84 Fed. Reg. 7424 (proposed Mar. 4, 2019).

94. See, e.g., 2019 Program Requirements Medicare, Ctrs. for Medicare & Medicaid Servs., https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/2019ProgramRequirementsMedicare.html [https://perma.cc/GB7L-PK34] (last modified May 8, 2019) (describing reporting requirements for Medicare programs to comply with Promoting Interoperability measures); MIPS Overview, Quality Payment Program, Ctrs. for

ACOs to accurately report on quality standards without a common infrastructure of EHRs that can aggregate data on key performance indicators and benchmarks.[95] A common indicator of nosocomial infection, for instance, may be critical to ensuring the integrity of performance assessment.

AI applications are already playing a role in promoting health-related interventions and should be subject to similar performance assessments. For example, as Natasha Singer has reported, Facebook has deployed an algorithm to flag users that may be so suicidal that police should be called by Facebook employees to intervene.[96] Mason Marks has documented numerous other examples of "social suicide prediction" programs, which use machine learning to generate risk scores for individuals.[97] There are long-term risks to privacy and autonomy that such scores could create—for example, if unregulated and shared beyond their source, they may affect the marketing a person experiences, or even job or insurance opportunities.[98]

They also raise important concerns about immediate risks to safety caused by false positives. What are the stigmatic concerns raised by being falsely accused of extreme suicidality, or of a suicide attempt? What do first responders think of the interventions they have been prompted to carry out? Ensuring that there are standard ways of reporting positive and negative interventions here could help policymakers better determine which AI to fund in this critical area. It could also nip in the bud problematic interventions, like the Samaritans' Radar App, which shut

---

Medicare & Medicaid Servs., https://qpp.cms.gov/mips/overview [https://perma.cc/BWX7-YNEM] (last visited June 28, 2019) (describing the Merit-based Incentive Payment System (MIPS) and four areas of reporting: "Quality, Improvement Activities, Promoting Interoperability (formerly Advancing Care Information), and Cost"); Promoting Interoperability (PI), Ctrs. for Medicare & Medicaid Servs., https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRincentiveprograms [https://perma.cc/CH85-HQ5Q] (last modified Aug. 14, 2019) (describing CMS' Promoting Interoperability program).

95. On the role of such indicators and benchmarks in AI-driven medical practice, see Frank Pasquale, Professional Judgment in an Era of Artificial Intelligence and Machine Learning, 46 boundary 2, at 73, 85 (2019) (exploring the role that metrics play in ACO performance assessments and compensation under the Affordable Care Act).

96. Natasha Singer, In Screening for Suicide Risk, Facebook Takes on Tricky Public Health Role, N.Y. Times (Dec. 31, 2018), https://www.nytimes.com/2018/12/31/technology/facebook-suicide-screening-algorithm.html (on file with the *Columbia Law Review*).

97. See generally Mason Marks, Artificial Intelligence Based Suicide Prediction, Yale J. Health Pol'y L. & Ethics (forthcoming), https://ssrn.com/abstract=3324874 (on file with the *Columbia Law Review*) (discussing the contours and unforeseen consequences of programs initiated by companies such Facebook, Crisis Text Line, and Operation Zero that "collect [consumers'] digital traces and analyze them with AI to infer [consumers'] health information").

98. Danielle Keats Citron & Frank Pasquale, The Scored Society: Due Process for Automated Predictions, 89 Wash. L. Rev. 1, 4 (2014) (describing spread of scoring technologies).

down its simple program for automated detection of suicidality after public complaints.[99]

CONCLUSION

Futurists envision AI programs that effectively act of their own accord, without direction or control by their developers (or any other person). Such entities could be quite dangerous.[100] However, advocates for such AI believe that law should effectively step out of the way of its development. How, the question goes, can the creators or owners of such general-purpose technology anticipate all the potential legal problems their AI might generate or encounter? No one wants to hold Microsoft responsible for ransom notes written with MSWord—it is a blank slate. Nor are parents responsible for the crimes of their children—they are independent entities.

Leading developers of AI, at present, benefit from both the "blank slate" and "independent entity" intuitions of nonresponsibility for their creations. But neither should immunize such firms, given a decade of research on algorithmic accountability. As Jack Balkin has observed, we all now know that algorithms can "(a) construct identity and reputation through (b) classification and risk assessment, creating the opportunity for (c) discrimination, normalization, and manipulation, without (d) adequate transparency, accountability, monitoring, or due process."[101] Moreover, we are well aware of their ability to malfunction, dating back at least to the Therac-25 debacle of the 1980s.[102] These factors all counsel in favor of discouraging the development of any AI whose actions are not directly attributable to a person or persons that can be held responsible for them.[103]

---

99. Jamie Orme, Samaritans Pulls 'Suicide Watch' Radar App over Privacy Concerns, Guardian (Nov. 7, 2014), https://www.theguardian.com/society/2014/nov/07/samaritans-radar-app-suicide-watch-privacy-twitter-users [https://perma.cc/342U-99KP].

100. See, e.g., Lynn M. Lopucki, Algorithmic Entities, 95 Wash. U. L. Rev. 887, 951 (2018) ("[Algorithmic entities] constitute a threat to humanity because the only limits on their conduct are the limits the least restrictive human creator imposes.").

101. 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 St. L.J. 1217, 1239 (2017). Algorithmic information processing is in effect the "brain" of robotics and AI agents. See generally Domingos, supra note 41, at 93–119.

102. See generally Edmond W. Israelski & William H. Muto, Human Factors Risk Management as a Way to Improve Medical Device Safety: A Case Study of the Therac 25 Radiation Therapy System, 30 Joint Commission J. Quality & Safety 689 (2004); Nancy G. Leveson & Clark S. Turner, An Investigation of the Therac-25 Accidents, Computer, July 1993, at 18, 18 (presenting an accident investigation of overdoses caused by the Therac-25 radiation therapy machine).

103. Frank Pasquale, Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society, 78 Ohio St. L.J. 1243, 1252–55 (2017) (arguing that very high levels of autonomy should be illegal if any harm is foreseeable, given the problems of attribution they can give rise to).

However appealing dreams of artificial general intelligence may be, the dominant version of AI now prevalent in commerce and government is only a few steps removed from algorithmic systems we are all now familiar with. For example, "AI hiring" based on voice parsing is not a substitute for a Director of Human Resources.[104] Nor is it an all-purpose assessment of character. Rather, it is a method of translating data (a voice) into an output (an assessment of likely success at a job) based on computational analysis of how past employees with similar voices have fared at the job. True, the concept of "similarity" here may have far more dimensions than a simple linear relationship; contemporary machine learning is premised on advances in computational power that not only allow various, granular hypotheses to be tested, but also combine potentially relevant variables in myriad ways.[105] However, the collection, analysis, and use of data is foundational to the process, and presents several opportunities for imposing duties on AI developers, given possibly inaccurate or inappropriate data.

Advocates for legal technology (including legaltech, regtech, and fintech) have promoted a "duty of technological competence" for lawyers.[106] In many cases, an attorney cannot properly serve a client without knowing how to use certain databases or search engines. Nor can a lawyer competently advise a modern business on a topic like document retention without a clear sense of how computers store data. Rules of

---

104. Stephen Buranyi, How to Persuade a Robot that You Should Get the Job, Guardian (Mar. 3, 2018), https://www.theguardian.com/technology/2018/mar/04/robots-screen-candidates-for-jobs-artificial-intelligence [https://perma.cc/5GCW-JN5E]. For further descriptions of such analytics, see Matthew T. Bodie, Miriam A. Cherry, Marcia L. McCormick & Jintong Tang, The Law and Policy of People Analytics, 88 U. Colo. L. Rev. 961, 963, 1032–38 (2017).

105. See The Royal Soc'y, Machine Learning: The Power and Promise of Computers that Learn by Example 19–20 (2017), https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf [https://perma.cc/YUE9-87MZ].

106. See, e.g., Model Rules of Prof'l Conduct 1.1 cmt. 8 (Am. Bar Ass'n 2018) (including the duty to "maintain the requisite knowledge and skill . . . including [keeping up-to-date on] the benefits and risks associated with relevant technology"); see also Anthony E. Davis & Steven M. Puiszis, An Update on Lawyers' Duty of Technological Competence: Part 1, N.Y. L.J. (Mar. 1, 2019), https://www.law.com/newyorklawjournal/2019/03/01/an-update-on-lawyers-duty-of-technological-competence-part-1/ [https://perma.cc/3XU5-6P4P] [hereinafter Davis & Puiszis, Update Part 1] (arguing that it is necessary for today's lawyers to maintain data security and become familiar with the technology used to run a law firm and practice law); Anthony E. Davis & Steven M. Puiszis, An Update on Lawyers' Duty of Technological Competence: Part 2, N.Y. L.J. (May 3, 2019), https://www.law.com/newyorklawjournal/2019/05/03/an-update-on-lawyers-duty-of-technological-competence-part-2/ [https://perma.cc/P7Z5-BWX4] [hereinafter Davis & Puiszis, Update Part 2] (using social media, electronic discovery, client technology, and technology to present information in court).

professional responsibility, as well as tort doctrines of legal malpractice,[107] enforce a duty of technological competence on many attorneys.[108]

In numerous fields, there is a parallel duty for technology providers to have some basic understanding of the law as they serve their clients. A video hosting service in the United States, for example, needs to understand the fundamentals of copyright law.[109] Firms developing electronic health record software unaware of the requirements of HIPAA[110] (and many other laws governing health privacy) cannot serve their clients well. In these cases, and many others, the onus is not simply on the buyer of the technology to vet what it is buying or leasing. Rather, principles of secondary liability effectively impose what might be called a duty of legal competence—of a basic understanding of what law requires—on technologists.[111] Some popular understandings of artificial intelligence pose a threat to the duty of legal competence by mystifying the bases of decisions. However, law and policy can require basic safeguards be taken in its development, can standardize public reporting on its effectiveness and safety, and can impose liability on the developers of unsafe, biased, or otherwise defective AI.

The promise of AI law and policy is to ensure that the owners and developers of algorithms are more accountable to the public.[112] Without imposing legal duties on the developers of AI, there is little chance of ensuring accountable technological development in this field. By focusing on data, the fundamental input for AI, both judges and policymakers can channel the development of AI to respect, rather than evade, core legal values.

---

107. See, e.g., James v. Nat'l Fin. LLC, No. 8931-VCL, 2014 WL 6845560, at *12 (Del. Ch. Dec. 5, 2014) (citing Del. Rules of Prof'l Conduct 1.1 cmt. 8).

108. See Model Rules of Prof'l Conduct r. 1.1 cmt. 8; Katherine Medianik, Note, Artificially Intelligent Lawyers: Updating the Model Rules of Professional Conduct in Accordance with the New Technological Era, 39 Cardozo L. Rev. 1497, 1512, 1514–16 (2018); Davis & Puiszis, Update Part 1, supra note 106; Davis & Puiszis, Update Part 2, supra note 106.

109. For an example of such copyright law, see Digital Millennium Copyright Act of 1998, 17 U.S.C. §§ 1202–1332 (2012).

110. Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (codified in scattered sections of 26, 29, and 42 U.S.C.).

111. In the case of HIPAA, the secondary liability would be imposed on the vendor via a business associate agreement. See Pasquale & Adams Ragone, supra note 79, at 609–15.

112. See Robyn Caplan, Joan Donovan, Lauren Hanson, & Jeanna Mathews, Algorithmic Accountability: A Primer 10 (2018), https://datasociety.net/output/algorithmic-accountability-a-primer/ [https://perma.cc/UTW2-62M9] ("Algorithmic accountability ultimately refers to the assignment of responsibility for how an algorithm is created and its impact on society; if harm occurs, accountable systems include a mechanism for redress."). Edward Rubin has defined accountability as "the ability of one actor to demand an explanation or justification of another actor for its actions, and to reward or punish the second actor on the basis of its performance or its explanation." Edward Rubin, The Myth of Accountability and the Anti-Administrative Impulse, 103 Mich. L. Rev. 2073, 2073 (2005).