Clearing Opacity Through Machine Learning

W. Nicholson Price II & Arti K. Rai[*]

ABSTRACT: Artificial intelligence and machine learning represent powerful tools in many fields, ranging from criminal justice to human biology to climate change. Part of the power of these tools arises from their ability to make predictions and glean useful information about complex real-world systems without the need to understand the workings of those systems.

But these machine-learning tools are often as opaque as the underlying systems, whether because they are complex, nonintuitive, deliberately kept secret, or a synergistic combination of those three factors. A burgeoning literature addresses challenges arising from the opacity of machine-learning systems. This literature has largely focused on the benefits and difficulties of providing information to lay individuals, such as citizens impacted by algorithm-driven government decisions.

In this Essay, we explore the potential of machine learning to clear opacity —that is, to help drive scientific understanding of the frequently complex and nonintuitive real-world systems that machine-learning algorithms examine. Using examples drawn from cutting-edge scientific research, we argue machine-learning algorithms can advance fundamental scientific knowledge and that deliberate secrecy around machine-learning tools restricts that learning enterprise.

---

[*] Arti Rai is the Elvin R. Latty Professor of Law, Faculty Director of the Center for Innovation Policy at Duke Law School, and Core Faculty, the Duke-Margolis Center for Health Policy. Nicholson Price is a Professor of Law at the University of Michigan Law School and a Core Partner at the Center for Advanced Studies of Biomedical Innovation Law at the University of Copenhagen Faculty of Law. For helpful comments and conversations, we thank Stuart Benjamin, Jamie Boyle, Ana Bracic, Kevin Collins, Greer Donley, Tabrez Ebrahim, Michael Frakes, Kyle Logue, Aisling McMahon, Laura Pedraza-Fariña, Pauline Kim, Govind Persad, Anya Prince, Barak Richman, Ana Santos Rutschman, Jacob Sherkow, Charlotte Tschider, Jeff Ward, Helen Yu, Patti Zettler, and participants at the Faculty Workshop at Duke Law School and Washington University Law School, the RIBS Workshop at the University of Copenhagen, and the Promises and Perils of New Biomedical Technologies conference at Northeastern University. Abigail Lynch, Michael McCarty, Maydha Vinson, and Bennett Wright provided excellent research assistance. Rai's work was supported by the Greenwall Foundation. Price was supported by the Cook Fund at the University of Michigan and Novo Nordisk Foundation grant NNF17SA027784. All errors are our own.

Our argument is more than a general plea for the innovation-related benefits of open science, or even a call for special attention to the unusually strong competitive protection secrecy can provide in the arena of machine learning. Rather, because the counterintuitive results machine learning can produce must be scrutinized particularly closely to distinguish exciting new hypotheses from spurious or otherwise misleading correlations, openness is particularly critical.

Turning to practical questions of law, institutions, and economics, we examine why developers are likely to keep machine-learning systems secret. We then draw on the innovation policy toolbox to suggest ways to reduce secrecy so that machine learning can help us not only to interact with complex, non-intuitive real-world systems but also to understand them.

I. INTRODUCTION

A substantial literature has arisen around worries that machine learning and artificial intelligence are opaque, or "black boxes." Many of these worries are legitimate. But a focus on these worries has led to an underappreciation of the ways that machine learning can help us remove opacity and understand the underlying real-world systems better.

Particularly in its more complex manifestations (e.g., deep-learning convolutional neural nets), machine learning functions quite differently from standard software used to perform analyses and predictions. While standard software applies explicit, human-designed decision rules to data—rules that humans designed with (we hope) some level of understanding of the underlying systems—machine learning algorithms do not, prior to their exposure to data, embody prediction rules.

Instead, machine learning algorithms learn directly from the data. In the common manifestation called "supervised learning" on which this paper focuses,[1] the data scientist exposes the learning algorithm to data that experts in the field have curated with respect to input features and then classified with respect to output labels. To put it more plainly, an algorithm could be given 100,000 X-rays of human lungs of which 5,000 have been labeled by radiologists as showing cancerous tumors. Assuming the radiologists and data scientists have done their job properly, this dataset should represent something close to "ground truth," or the underlying accurate reality.[2] Learning or "training" a model involves a process in which, over a series of iterations, model parameters for translating inputs into outputs are adjusted, and model predictive performance tested, until predictive performance cannot be improved.[3] Once the machine learning model has been trained, it is generally applied to a subset of the training data to which it had not previously been exposed, known as "test" data, to see how well it performs on

1. As contrasted with supervised learning, unsupervised learning that does not need highly curated data is not as far advanced. Because the costs of proper data curation by experts are often nontrivial, unsupervised learning has been described as the "holy grail" of data-based machine learning. Andrew Ng, *Foreword* to WORLD INTELL. PROP. ORG., WIPO TECHNOLOGY TRENDS 2019: ARTIFICIAL INTELLIGENCE 8, 8 (2019) [hereinafter WIPO REPORT]. In addition to supervised and unsupervised learning, other forms of machine learning include reinforcement learning and multi-task learning. For a user-friendly diagram of the different types of techniques that are sometimes called "AI" (including but not limited to machine learning), see *id.* at 42 figs.3.5 & 3.6. For the purposes of this Essay, the taxonomy of the WIPO report is particularly useful because the report gives metrics of patent filings and scientific publications based on its taxonomy. According to WIPO patent data, machine learning is the dominant "AI" technique, "represent[ing] 89 percent of patent families related to an AI technique [and] 40 percent of all AI patent families." *Id.* at 41. The representation of machine learning in the scientific literature on AI methodologies/techniques stands at 64 percent, which constitutes 54 percent of all scientific publications on AI. *Id.*

2. *See generally* GARETH JAMES, DANIELA WITTEN, TREVOR HASTIE & ROBERT TIBSHIRANI, AN INTRODUCTION TO STATISTICAL LEARNING (G. Casella, S. Fienberg & I. Olkin eds., 2013) (aiming to bring statistical learning into the mainstream).

3. *See, e.g.*, David Lehr & Paul Ohm, *Playing With the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 695–701 (2017).

new data.[4] In the best-case scenario, the model is also applied to data from an entirely new source to further validate its performance.[5]

Why use machine learning? It's especially useful to apply to the large number of systems (physical, social, or some combination of the two) where human field experts haven't yet figured out how the systems work.[6] In those cases, machine learning can examine hundreds or thousands of potentially relevant input variables and can, if properly tested and validated, be very helpful for generating accurate predictions.

Biomedicine is an example where human experts often don't understand what is going on due to systemic complexity, and machine learning can be helpful. Consider the WAVE surveillance model that was recently approved by the FDA. This model predicts vital sign instability in hospital patients and triggers alerts so that rapid-response nurse-led teams can intervene to stabilize the patient. The learning algorithm's training data came from the electronic health records of patients who had experienced such instability. When the model and response procedures were implemented, the average duration of instability decreased substantially.[7]

From an epistemological standpoint, it is perhaps not surprising that complex research tools are useful for studying complex systems.[8] Biomedicine is far from the only example. The world is full of complex systems that experts do not yet understand, for which accurate predictions could improve social welfare. Medicine, climate change, traffic patterns, criminal recidivism, and many other fields provide potential use cases for accurate machine-learning tools.

Unfortunately, these models are often opaque. Even when human field experts are given full access to the learning algorithm, training data, training process, and resulting model, the models can be difficult to parse because

4. *Id.* at 684.

5. Adarsh Subbaswamy & Suchi Saria, *From Development to Deployment: Dataset Shift, Causality, and Shift-Stable Models in Health AI*, 21 BIOSTATISTICS 345, 345–46 (2020).

6. Machine learning's distinctive characteristics can be highlighted through comparison with (for example) conventional linear regression analysis. In the latter analysis, the human field expert has—or at least should have—some understanding of the system being studied and can therefore specify a relatively small set of relevant input variables. Only the coefficients attached to the small number of input variables are determined by the software. Moreover, those coefficients are determined by source code that can be understood by the data scientist. And because the human analyst specifies the input variables for a reason, ideally a reason that has a plausible causal interpretation, the human analyst should be able to explain her output results.

7. Ravi B. Parikh, Ziad Obermeyer & Amol S. Navathe, *Regulation of Predictive Analytics in Medicine: Algorithms Must Meet Regulatory Standards of Clinical Benefit*, 363 SCIENCE 810, 811 (2019). To be sure, as the authors point out, reducing vital sign instability may not, in and of itself, be meaningful clinically. *Id.* More generally, the evidence of performance that the FDA required in approving the algorithm was not necessarily optimal. *Id.*

8. Of course, in some cases, complex systems can usefully be probed by simpler tools. We make only the modest claim that, in certain cases, probing by simpler tools has not been fruitful.

they are often complex and nonintuitive. The legal literature is accordingly replete with concern about machine learning's "black box" nature.[9]

In this Essay, we take a position on opacity that is, at best, underrepresented in the legal literature. We begin by noting that opacity in machine learning happens at two related layers: the opacity of the system being studied (what we call "system opacity") and the opacity of the research tool—machine learning—being deployed to study it (what we call "tool opacity"). Both the underlying real-world systems and the machine-learning models built to probe them can be both complex and nonintuitive. We argue that even though the tool may be opaque, concerns about tool opacity should not overshadow the considerable promise associated with using the tool to *clarify* system opacity.

Such clarification will be impeded, however, unless a third type of tool opacity—deliberate secrecy around machine-learning tools—is addressed.[10] In general, the strength of secrecy as a competitive shield is a function of a competitor's ability to reverse engineer or independently invent the information that is being kept secret. In the case of complex machine-learning systems with non-intuitive outputs, challenges associated with reverse engineering may allow secrecy to persist for long periods of time.[11] Meanwhile, it is precisely these non-intuitive outputs that should be most open to the type of robust scrutiny that can help distinguish promising new hypotheses from spurious or otherwise misleading correlations.

Because we focus on clearing system opacity, we consider how secrecy impacts the work and understanding of field experts—for example, life scientists striving to understand and influence human biology[12] or climate

9. *See, e.g.*, Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1829 (2019) ("A recurrent concern about machine learning algorithms is that they operate as 'black boxes.'"); Katherine J. Strandburg, *Rulemaking and Inscrutable Automated Decision Tools*, 119 COLUM. L. REV. 1851, 1858, 1863–64 (2019) (arguing that algorithmic inscrutability is problematic by analogy to rulemaking); Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 841, 842–44 (2018).

10. We do not address fully certain aspects of tool opacity. Some computer scientists have argued, for example, that attempts to use machine learning to "explain" what other machine learning is doing are, at best, statistical approximations and should be avoided, at least in high-risk contexts, in favor of trying to design models that can be understood by humans. Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTEL. 206, 206, 208 (2019); *see also* Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan.–June 2016, at 1, 3–5 (clarifying different forms of opacity in machine learning). We address machine learning tools that explain machine learning primarily to the extent that they prove useful for generating new hypotheses about the opaque physical or social systems being studied.

11. To be sure, this is not an absolute rule and the state of the art in reverse engineering is always improving. *See infra* text accompanying notes 71–72.

12. Diogo M. Camacho, Katherine M. Collins, Rani K. Powers, James C. Costello & James J. Collins, *Next-Generation Machine Learning for Biological Networks*, 173 CELL 1581, 1581 (2018) ("[I]t is becoming imperative to focus our data-analytical approaches on tools and techniques specifically tailored to handle large, heterogeneous, complex datasets. Machine learning . . . aims to address this complexity, providing next-level analyses that allow one to take new perspectives and generate novel hypotheses about living systems.").

scientists seeking to understand weather patterns so as to provide advice to policy makers[13]—rather than how secrecy affects lay individuals who may be influenced by algorithmic decisions, such as patients or populations in areas vulnerable to severe weather. Although the concerns of field experts who hope to clear system opacity are obviously less immediate than those faced by individuals directly impacted by algorithms, the latter set of concerns is —fortunately for our project—already the subject of a substantial literature.[14] Indeed, in the case of human biology, Congress has already determined that "black box" impacts on lay individuals are sufficiently concerning that they provided a statutory foundation for FDA regulatory authority.[15] Throughout this Essay, we draw upon the accountability-related insights of this prior work.[16]

Going beyond this prior work, we argue here that developer secrecy surrounding tools substantially hinders the ability of machine learning to elucidate system opacity. Ours is more than a general argument in favor of open science, or even a call for attention to the unusually potent competitive power of secrecy in machine learning.[17] Rather, openness surrounding non-intuitive outputs produced by machine learning is particularly important, because it sheds light in one of the areas where such light may be needed most.

---

13. MACHINE LEARNING AND DATA MINING APPROACHES TO CLIMATE SCIENCE, at v (Valliappa Lakshmanan, Eric Gilleland, Amy McGovern & Martin Tingley eds., 2015) ("[B]ecause the goal of machine learning in climate science is to improve our understanding of the climate system, it is necessary to employ techniques that go beyond simply taking advantage of co-occurrence and, instead, enable increased understanding."); Nicola Jones, *Machine Learning Tapped to Improve Climate Forecasts*, 548 NATURE 379, 380 (2017); David Rolnick et al., *Tackling Climate Change with Machine Learning*, ARXIV:1906.05433V1, at 1, 53 (2019), https://arxiv.org/pdf/1906.05433v1.pdf [https://perma.cc/DDJ5-P68M] (stating that machine learning has the potential to lead to causal models for "understanding weather patterns, informing policy makers, and planning for disasters").

14. Much of the literature on layperson effects focuses on algorithms developed by social scientists to predict behavior in criminal and civil contexts and addresses not only machine-learning algorithms but also ordinary algorithms protected by secrecy. *See generally, e.g.*, FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015) (discussing the use of algorithms tracking individual behaviors by corporations and how their decisions about how to use that data affect our lives); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018) (developing an "account of trade secret evidence in criminal cases and develop[ing] a framework to address the problems that result").

15. 21st Century Cures Act, 21 U.S.C. § 360j(o) (2018).

16. That said, the level and type of disclosure necessary for appropriate accountability to those who are affected by machine-learning likely differs from that necessary in the scientific context. For example, disclosure sufficient for reproducibility is, at least in principle, required by patent law and by scientific norms of publication, *see infra* Section III.C, but may not be useful or necessary for appropriate accountability to lay persons.

17. *See, e.g.*, W. Nicholson Price II, *Big Data, Patents, and the Future of Medicine*, 37 CARDOZO L. REV. 1401, 1432–36 (2016) [hereinafter Price, *Big Data, Patents, & Medicine*] (describing the power of secrecy to limit competition in medical AI). As we discuss below, cloud-based machine learning models can be particularly difficult to reverse engineer.

Accordingly, we suggest mechanisms for combating secrecy. More specifically, we draw upon innovation policy scholarship to provide a toolkit for overcoming secrecy. This scholarship, which addresses levers for promoting socially beneficial innovation, yields insights that have largely been elided both by law and technology commentators and by commentators focused on competition policy. As to the former, commentators have typically seen secrecy through the lens of due process, privacy, norms of anti-discrimination, or other individual rights.[18] Competition scholars, meanwhile, have often viewed secrecy through the lens of monopoly control over data.[19] In contrast, this Essay invokes innovation policy tools to show how developer secrecy endangers the ability of machine learning to advance basic scientific understanding—that is, its ability to make the "black box" of many real world systems at least a bit grayer.[20]

Ultimately, innovation policy provides a perspective that harnesses machine learning to clarify system opacity over the medium to long term, yet is also grounded in shorter term practical institutional and economic considerations. Innovation policy has a history of addressing opaque systems, including opaque systems (e.g., biological systems) that are probed by research tools or interventions that are themselves opaque (e.g., biomarkers or biologics). Accordingly, it suggests mechanisms for organizing, and grappling with, the full range of interrelated challenges associated with opacity in machine learning. These interrelated challenges include a level of model complexity that precludes full understanding—or easy reverse engineering—by human field experts; secrecy, unpredictability, and a concomitant lack of reproducibility; and ultimately the normative goal that this Essay adopts—providing incentives for production of accountable predictive performance in the short term, robust disclosure in the medium term, and deeper understanding in the medium to long term.

For a number of reasons, we focus in this Essay on the applications of machine learning in the health-related life sciences. First, not only does

18. *See, e.g.*, Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1537–52 (2019) (describing instrumental/error-correction, justificatory/legitimacy, and dignitary reasons for demanding algorithmic accountability).

19. *See infra* text accompanying note 70.

20. Thanks to Kevin Collins for this phrase. A recent, insightful analysis by intellectual property scholar Jeanne Fromer sounds some themes similar to those in our Essay. *See generally* Jeanne C. Fromer, *Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation*, 94 N.Y.U. L. REV. 706 (2019) (discussing computing machines and the effect trade secret law has on data input). Fromer makes the point that, from the standpoint of cumulative innovation, trade secrecy surrounding machine learning may prove excessively protective and offers some suggestions for reducing its potency. *Id.* at 727–36. Our analysis diverges from Fromer's on two principal grounds. First, we focus on what we consider the subset of cumulative innovation for which we believe disclosure is most important—non-intuitive results that may be spurious or may be accurate, with the latter possibility leading to a result that ultimately clears system opacity. Second, we believe the challenges posed by trade secrecy in machine learning are best addressed not by changes in trade secrecy law (absent trade secrecy law, for example, machine learning developers might just use more robust mechanisms to preserve actual secrecy) but through other mechanisms.

conventional health care represent a large part of the U.S. economy (17.7 percent as of 2018),[21] but human health more generally is a critical component of social welfare. Second, the life sciences are rife with opaque systems; human biology is notoriously complex, and many diseases, systems, and challenges lack any clear explanation. To cite just one salient example, we still lack an understanding of the underlying mechanism of Alzheimer's Disease, despite millions afflicted and billions spent on research.[22] Not surprisingly, then, life sciences researchers have embraced machine learning, noting that such approaches are "becoming imperative" for generating new hypotheses from biological data.[23] Third, the health-related life sciences are an area of intense innovation policy interest, in large part for the preceding two reasons; they are the subject of large government expenditures, numerous policy levers, and focused academic attention (including by the two of us). The health-related life sciences, therefore, provide a useful landscape to examine the ways in which policymakers can use the innovation policy toolkit to promote machine learning that clears system opacity.

Although we focus on the example of health-related life sciences, the Essay's analysis should be applicable to contexts where the incentives of key actors can plausibly be aligned with the normative goal of using machine learning to advance scientific knowledge: both accountable prediction and more basic understanding. The climate science community, for example, is primed to benefit from this Essay's analysis as its members become more comfortable with machine learning.[24]

In contrast, in institutional contexts where gaming of, and adversarial attack on, the machine learning model are first-order challenges, the Essay's applicability may be more limited.[25]

---

21. *National Health Expenditure Data: Historical*, CTRS. FOR MEDICARE & MEDICAID SERVS., https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html [https://perma.cc/DBL8-L463] (last updated Dec. 17, 2019, 2:19 PM).

22. *See, e.g.*, Matthew Herper, *One of the World's Best Drug Hunters Went After Alzheimer's. Here's How He Lost*, STAT (June 6, 2019), https://www.statnews.com/2019/06/06/al-sandrock-biogen-alzheimers-aducanumab [https://perma.cc/FYJ2-WQC3] (noting the repeated failure of drug development efforts in the area, in part due to lack of scientific understanding).

23. *See* Camacho et al., *supra* note 12, at 1581.

24. *See* Rolnick et al., *supra* note 13, at 1–2 ("Despite the growth of movements applying [machine learning] and AI to problems of societal and global good, there remains the need for a concerted effort to identify how these tools may best be applied to tackle climate change." (footnote omitted)).

25. *See generally, e.g.*, Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1 (2018) (providing an overview of algorithmic gaming concerns); Ignacio N. Cofone & Katherine J. Strandburg, *Strategic Games and Algorithmic Secrecy*, 64 MCGILL L.J. (forthcoming 2020) (providing a nuanced analysis of how to weigh gaming concerns against the benefits of disclosure); Wachter et al., *supra* note 9, at 851–53 (describing adversarial machine learning algorithms). Gaming is of concern, regrettably, in many areas including health. One recent commentary argues that reimbursement pressures may cause malevolent health care providers to introduce deliberate error into medical machine learning. *See generally* Samuel G. Finlayson et al., *Adversarial Attacks on Medical Machine Learning*, 363 SCIENCE 1287 (2019) (describing the possibility of small deliberate perturbations in medical data to trick medical machine learning

Within the standard innovation policy toolkit, public funding and patents are supposed to be the regimes most associated with disclosure.[26] But theory and practice show the limits of both tools in effectuating disclosure. Recognizing these limitations, the innovation policy literature has substantially expanded the tool kit.

In the health-related life sciences, where resolution of concerns regarding safety and efficacy associated with new products or processes represents a form of innovation,[27] scholars have highlighted risk regulators as data aggregators that could promote disclosure.[28] Some of these risk regulation agencies are moving forward with experiments on regulating machine learning. The FDA, for instance, has started allowing firms with a demonstrated culture of excellence and a willingness to expose themselves to ongoing regulatory scrutiny a chance to reach the market with lighter up-front scrutiny of their software products.[29] Carefully monitored experimentation should be encouraged. However, the lens of innovation policy teaches that regulatory carrots should come with disclosure obligations.

Finally, we advance the idea of promoting disclosure through concerted action by demand side market and public sector actors, principally private and public insurers.

Part II of the Essay provides an introduction to the issue of opacity as seen through the lens of the field expert. We analyze various components of both tool and system opacity. Part III addresses impediments to clearing system opacity created by secrecy, particularly secrecy over training data. Part IV

---

into diagnosing incorrect, but better reimbursed, conditions). Although such malevolence is certainly possible, fraudulent "upcoding" could presumably be achieved through simpler means.

26. Publicly funded grants typically have requirements for disclosure of final results and of research data generated along the way. *See* W. Nicholson Price II, *Grants*, 34 BERKELEY TECH. L.J. 1, 33–34 (2019) (describing disclosure requirements). Patents include disclosure requirements as part of the process of applying for a patent; in fact, the word "patent" is derived from the Latin *patere* (to lay open). *See* 35 U.S.C. § 112 (2018) (enumerating what must be disclosed in a patent application); *see also generally* Jeanne C. Fromer, *Patent Disclosure*, 94 IOWA L. REV. 539 (2009) (providing an analysis of patent law's disclosure requirements). *But see* Fromer, *supra*, at 551 (noting the possibility of inadequate disclosure); W. Nicholson Price II, *Expired Patents, Trade Secrets, and Stymied Competition*, 92 NOTRE DAME L. REV. 1611, 1617–18 (2017) [hereinafter Price, *Expired Patents*] (summarizing how firms pair incomplete disclosure with secrecy to limit competition).

27. *See, e.g.*, Rebecca S. Eisenberg, *The Role of the FDA in Innovation Policy*, 13 MICH. TELECOMMS. & TECH. L. REV. 345, 347 (2007) (describing the FDA's role in innovation, represented by the development of safety and efficacy data about drugs); Rebecca S. Eisenberg & W. Nicholson Price II, *Promoting Healthcare Innovation on the Demand Side*, J.L. & BIOSCIENCES, Apr. 2017, at 3, 4.

28. W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421, 462–65 (2017).

29. *See generally* FDA, DEVELOPING A SOFTWARE PRECERTIFICATION PROGRAM: A WORKING MODEL (2019) [hereinafter FDA, SOFTWARE PRECERTIFICATION PROGRAM], https://www.fda.gov/media/119722/download [https://perma.cc/5Q5F-MQFY] (describing the pilot "Pre-Cert" program for developers with demonstrated cultures of excellence); FDA, PROPOSED REGULATORY FRAMEWORK FOR MODIFICATIONS TO ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) (2019), https://www.fda.gov/media/122535/download [https://perma.cc/7MA4-L4QN] (describing a model for managing changes in artificial intelligence to lower requirements for supplemental regulatory filings).

discusses policy levers, involving both IP and alternatives, that could be used to promote accountable prediction in the short term, disclosure sufficient for reproducibility in the medium term, and reduction in system opacity over the medium to long term.

II. OPACITY AND THE FIELD EXPERT

This Part provides a relatively brisk summary of the now voluminous legal and computer science literature on opacity in machine learning. Although we do not purport to be exhaustive, we highlight points relevant to the normative framework of this Essay. Accordingly, we focus on opacity—both tool opacity and system opacity—not from the standpoint of the layperson, or even a health professional using software enabled by machine learning, but from the standpoint of the field expert.

A. WHAT IS OPACITY?

Both tool and system opacity involve many different aspects. The literature, which has focused on tool opacity, has generally identified three components of tool opacity that are particularly important.[30] These are *complexity*, which can render the number of interdependent input factors involved too high for ready comprehension, even by experts; *non-intuitiveness*, where the decision rules used by an algorithm, even if observable, do not make sense to experts; and *secrecy*, where details of algorithmic development are deliberately concealed.[31]

These three concepts can interact with and enhance each other, but are distinct;[32] the result can range from fully transparent to fully opaque. Furthermore, although the literature has focused on these issues with respect to machine learning tools, two of the issues—complexity and non-intuitiveness—will also (and perhaps more fundamentally) represent features of the *underlying natural or social systems* that the tools are used to study.

---

30. A substantial literature has focused on the related and much-discussed concept of explainability, especially explainability not to field experts but to individuals, and has linked it to concepts like due process, antidiscrimination, or consent. For an overview of legal requirements and scholarly efforts in that area, see, for example, Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1099–117 (2018). *See also generally* Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189 (2019) (describing the explainability requirement as embedded in the European Union's General Data Protection Regulation (GDPR)).

31. For a particularly helpful analysis of these related factors (using, at times, slightly different terminology), see Selbst & Barocas, *supra* note 30, at 1089–99.

32. The interactions between these three concepts are, no surprise, complex. In general, the three concepts will tend to reinforce each other—complex models are more likely to yield non-intuitive results, and the difficulty of reverse-engineering complexity from non-intuitive outputs will tend to strengthen secrecy—but these interactions are not absolutes, and will vary contextually. Complex secret models can still be explained by reverse engineering, and relatively simple models can yield non-intuitive results. A full exploration of these issues is beyond the scope of this Essay.

1. Complexity

Both machine learning models and the underlying systems typically studied by machine learning are likely to be complex and poorly understood.[33] As a statistical matter, complexity can arise for many reasons, including nonlinearity and discontinuity.[34] In the case of machine learning models, however, the more important additional feature is that they encompass many more input variables than the typical human-designed model.[35] Similarly, the systems studied are often themselves tremendously complex, relying on interrelated networks of features.[36] Indeed, that parallel complexity generates some of the appeal of using machine learning to probe and manipulate systems that are otherwise too complex for traditional scientific tools.[37]

In some cases, particular machine-learning models can attempt to explain underlying systemic complexity, either generally or in a particular case.[38] However, these explanations, which often use machine learning to interpret machine learning, are typically statistical approximations, rather than fully accurate representations.[39]

2. Non-Intuitiveness

Non-intuitiveness of systems and tools can also lead to opacity. Here, the field expert—either by direct analysis of the model or with the assistance of machine learning that helps to explain the model—may be able to hone in on particular correlations between inputs and outputs. Nonetheless, the correlations might appear inexplicable. For example, suppose that a machine-learning model accurately predicts responsiveness to a particular drug. The machine learning that interprets the underlying models (also based on machine learning) "explains" that the prediction was made based in part on the individual's breakfast preferences. Even if such predictions were accurate,

33. *See generally* MELANIE MITCHELL, COMPLEXITY: A GUIDED TOUR (2009) (providing a general analysis of complexity).

34. *See id.* at 22–27. For present purposes, we do not need to dive into complexity theory, nor field-specific definitions.

35. *See* Selbst & Barocas, *supra* note 30, at 1094–96.

36. *See, e.g.*, Gezhi Weng, Upinder S. Bhalla & Ravi Iyengar, *Complexity in Biological Signaling Systems*, 284 SCIENCE 92, 92 (1999); *see also* Selbst & Barocas, *supra* note 30, at 1094 (using the term "inscrutability" to refer to the "situation in which the rules that govern decision-making are so complex, numerous, and interdependent that they defy practical inspection and resist comprehension").

37. *See, e.g.*, W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 429–31 (2015) (describing the advantage of using complex algorithmic models because of underlying biological complexity).

38. *See generally* David Gunning et al., *XAI—Explainable Artificial Intelligence*, 4 SCI. ROBOTICS 1 (2019) (providing an overview of explainable artificial intelligence and arguing that explainability is essential for effective AI management).

39. Rudin, *supra* note 10, at 208.

they would be non-intuitive; it is hard "to weave a sensible story to account for the statistical relationships in the model."[40]

Non-intuitiveness can be a red flag for artifacts. An unknown latent variable may explain both the predicted outcome and the predictor used by the algorithm (say, individuals with certain socioeconomic status tend both to eat a certain type of breakfast and to take medications more consistently). Or the correlation may be purely spurious, resulting from the ability of algorithms to find any number of non-real relationships in datasets.

In contrast with intractable complexity, however, non-intuitiveness can have a bright side. The bright side emerges if the correlation in question arises not as a consequence of some hidden latent variable that provides the actual causal impact but, instead, because the model is picking up on some causal factor in the underlying system that is not intuitive to the human.[41] To put the point another way, machine learning could prove quite fruitful in assisting human hypothesis generation.[42] In the example above, it turns out (really!) that grapefruit consumption mediates the effectiveness of certain drugs.[43] Had we not already known that, the machine learning algorithm might have sparked further investigation, thereby yielding the (current) understanding that the furanocoumarin chemicals in grapefruit juice inhibit a key enzyme involved in drug metabolism.

More generally, machine learning approaches are beginning to achieve widespread use throughout biomedical research precisely because of their ability to yield non-intuitive hypotheses. For example, field experts have used machine learning to develop insights into networks of genetic regulatory activity, protein folding, protein–protein interactions, and many other fundamental biological conundrums.[44] These hypotheses have then been

40. Selbst & Barocas, *supra* note 30, at 1097.

41. As machine learning progresses, an interesting question for patent law scholars will be whether a model's non-intuitive reasoning renders it "nonobvious," and therefore potentially patentable, at least to the extent models are considered patent-eligible subject matter. Of course, to the extent that the "ordinary artisan" that patent law uses as its reference point for determining what is nonobvious, an artisan whose skill is enhanced by machine learning, the artisan-machine learning hybrid might find very little non-intuitive. For an interesting exploration of these possibilities, see generally Ryan Abbott, *Everything is Obvious*, 66 UCLA L. REV. 2 (2019).

42. Alternatively, it could assist in disrupting flawed intuitions. As vast amounts of literature in economics, psychology, and neuroscience have now shown, human intuition is, at best, an incomplete foundation for empirical investigation. *See, e.g.*, Veronika Denes-Raj & Seymour Epstein, *Conflict Between Intuitive and Rational Processing: When People Behave Against Their Better Judgment*, 66 J. PERSONALITY & SOC. PSYCH. 819, 820 (1994).

43. *See generally* David G. Bailey, George Dresser & J. Malcolm O. Arnold, *Grapefruit–Medication Interactions: Forbidden Fruit or Avoidable Consequences?*, 185 CAN. MED. ASS'N J. 309 (2013) (noting that more than 85 drugs are now known to interact with grapefruit, many with "serious adverse effects"); David G. Bailey, J. David Spence, Claudio Munoz & J. Malcom O. Arnold, *Interaction of Citrus Juices with Felodipine and Nifedipine*, 337 LANCET 268 (1991) (identifying the interaction of grapefruit—but not orange—juice with the drugs felodipine and nifedipine, identified accidentally when grapefruit juice was used to mask the taste of alcohol in an earlier study).

44. Travers Ching et al., *Opportunities and Obstacles for Deep Learning in Biology and Medicine*, 15 J. ROYAL SOC'Y INTERFACE 1, 12–22 (2018) (collecting hundreds of studies).

tested experimentally, with the ultimate result being causal identification of important drivers of disease states.[45]

Machine learning is also being used throughout the drug discovery and development process—for instance, to identify and validate novel drug candidates. In addition to the work that is being done by large firms, many startups (180, according to one recent count)[46] are now devoted exclusively to the use of machine learning in drug discovery and development.[47]

Although this drug discovery work is not necessarily directed towards generating novel hypotheses about fundamental questions, some of it is. Moreover, applied work can yield first-principles insight. The use of neural networks—one form of machine learning—to predict cardiovascular disease from retinal scans has yielded a novel indication of strong sex-specific differences in the retinal fundus.[48] Ziad Obermeyer has used machine learning to suggest that standard methods of evaluating knee X-rays miss racial differences and biases diagnoses against non-white patients.[49] Obermeyer found that machine learning analysis of X-rays could predict patient pain scores substantially better than radiologists, suggesting that *something* is objectively wrong in the knees of non-white patients complaining of pain that current radiology has not yet identified.

Thus, although non-intuitiveness contributes to opacity, further work on non-intuitive (to humans) variables identified by machine learning has already proven an important mechanism by which machine learning advances fundamental understanding.

3. Secrecy

A final critical component of opacity is secrecy. Although secrecy directly implicates only tool opacity, addressing tool opacity can be quite important for purposes of reducing system opacity. Specifically, to the extent that the machine learning produces a non-intuitive output, this non-intuitive output may be a reflection of a real but non-intuitive (to humans) feature of the real-world system. Or it could be a misleading correlation. Ultimately, the test of truth will come from rigorous testing of the non-intuitive hypothesis. But before scientists expend resources on hypothesis testing, they should have the

45. *See* Camacho et al., *supra* note 12, at 1582, 1584 (collecting multiple studies that led to identification of causal drivers of disease states as well as characterization of the mechanism of action of existing drugs).

46. GREG REH, DELOITTE, 2020 GLOBAL LIFE SCIENCES OUTLOOK 22 (2020).

47. For a review of machine learning in drug discovery and development, see generally Sean Ekins et al., *Exploiting Machine Learning for End-to-End Drug Discovery and Development*, 18 NATURE MATERIALS 435 (2019).

48. *See* Ryan Poplin et al., *Prediction of Cardiovascular Risk Factors from Retinal Fundus Photographs via Deep Learning*, 2 NATURE BIOMEDICAL ENG'G 158, 161 (2018) (noting that "results show strong gender differences in the fundus photographs and may help guide basic research investigating the anatomical or physiological differences between male and female eyes").

49. Ziad Obermeyer, Machine Learning for Healthcare, *Ziad Obermeyer*: *Algorithms Are as Good as Their Labels*, YOUTUBE (Aug. 3, 2020), https://www.youtube.com/watch?v=xt_pwq4HZWA [https://perma.cc/SR3Q-P6LN] (study description begins at 13:18).

disclosure necessary for thorough investigation into, and robust reproducibility of, the initial output.

For its part, a machine-learning developer may, for reasons of competitive advantage, want to maintain secrecy over one or more of the following aspects of its work product: the learning algorithm's source code, associated parameters, the training data, training process, or the resulting model. The competitive advantage conferred by secrecy increases with the cost of reverse engineering and/or independent invention. For machine learning models, this cost may be quite significant.[50] Notably, because it relates only to the tool, secrecy is the only of the three opacity components that results solely from deliberate choice—that is, the choice of a developer to conceal data, methods, results, or some combination.

In Part III, we discuss the literature's engagement with machine learning secrecy, and add our own contributions. Before turning to that discussion, we address briefly the argument that the relevant normative goal for machine learning should be validated performance, not clarification of opacity (whether system opacity or tool opacity).

B. SHOULD OPACITY MATTER?

One potential response to the foregoing concerns is a shrug. Should field experts view opacity, as contrasted with validated performance, as particularly important?[51]

A utilitarian view might suggest that field experts—whether they are regulatory experts in the public sector with statutory authorization to oversee products enabled by machine learning or field experts in the private sector—should worry primarily about reliable performance, not opacity (tool or system). Indeed, even in high-risk contexts like medicine, those who have criticized the FDA for being too lax in its treatment of machine learning have often focused on lenient performance requirements, not opacity.[52]

As these commentators rightly note, a significant fraction of the rationale for many medical interventions rests (at best) on verified performance, not understanding. For example, the mechanisms of action for many small molecule drugs and biologics are poorly understood.[53] Indeed, in the case of

50. *See* Fromer, *supra* note 20, at 707–08.

51. We note that in contexts that raise first-order legitimacy and dignitary concerns, such as criminal justice, opacity creates other problems. *See, e.g.*, Kaminski, *supra* note 18, at 1545–50 (cataloging dignitary concerns); Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, 52 U.C. DAVIS L. REV. 1067, 1084–85 (2018) (calling for transparency in criminal justice algorithms); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1308–10 (2008) (calling for transparency in algorithmic decisions to satisfy procedural due process); Meg Leta Jones, *The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood*, 47 SOC. STUD. SCI. 216, 231 (2017) (calling for human involvement to protect data subjects).

52. *See, e.g.*, Parikh et al., *supra* note 7, at 810.

53. This lack of understanding has, at times, "led to spectacular failures in medicine." Jacob S. Sherkow & Christopher Thomas Scott, *The Pick-and-Shovel Play: Bioethics for Gene-Editing Vector Patents*, 97 N.C. L. REV. 1497, 1531 (2019).

biologics, particularly complex biologics like monoclonal antibodies ("mAbs," antibodies that bind to a single target), scientists often don't even know precisely what these molecules "are" in terms of structural characterization and fully verifiable reproducibility.[54]

Although the immediate goal of verified performance is clearly important, even a standard utilitarian should, in many cases, favor some clarification of system and tool opacity.[55] Insight into the inputs that influence a model's decision making can itself be important for ensuring reliable, repeated performance, particularly if the model is going to be used on data quite different from its original training data.[56] In health care, where the training data for many machine learning models often comes from relatively ethnically homogeneous, wealthy individuals (for example, those who seek treatment at facilities like Memorial Sloan Kettering), such transferability to different contexts is critical.[57]

Pulling the lens back more fully (and as discussed further in Parts III and IV), an exclusive focus on performance neglects the powerful role that basic scientific understanding plays in promoting social welfare. As economists have long noted, such understanding generally produces significant positive externalities for society and is an important part of the innovation ecosystem.[58] In the case of machine learning, an additional significant positive externality of basic understanding (albeit one not typically accounted for by economists) might be greater trust by the public at large.

Machine learning also underscores the nonlinearity of scientific progress now recognized by most science policy analysts. Contrary to Vannevar Bush's post-World War II vision of science as linear movement from basic understanding to applied work that produces social welfare benefits,[59] analysts now recognize that the flow of knowledge is bidirectional. That is, applied

54. For an extended discussion of this issue, and its implications for price competition and fundamental understanding, see generally W. Nicholson Price II & Arti K. Rai, *Manufacturing Barriers to Biologics Competition and Innovation*, 101 IOWA L. REV. 1023 (2016).

55. Greater challenges arise where there is a tradeoff between performance and explainability. To look at the extreme example of such a tradeoff, a linear model of two variables may have little predictive performance but be quite explainable. Whether, when, and to what extent such explanation/performance tradeoffs exist in various areas of machine learning are hotly debated questions. The interventions we propose below focus on the secrecy-related aspects of opacity rather than the complexity or non-intuitiveness-based aspects to which this tradeoff might be more salient.

56. *See, e.g.*, Ariel Dora Stern & W. Nicholson Price II, *Regulatory Oversight, Causal Inference, and Safe and Effective Health Care Machine Learning*, 21 BIOSTATISTICS 363, 364–65 (2020) (noting the importance of causal understanding for a trans-contextual application of AI).

57. *See generally* W. Nicholson Price II, *Medical AI and Contextual Bias*, 33 HARV. J.L. & TECH. 65 (2019) (discussing the importance of transferability in health care AI systems).

58. For a recent, highly accessible summary of this literature (coupled with an argument for greater public funding of basic science), see generally JONATHAN GRUBER & SIMON JOHNSON, JUMP-STARTING AMERICA: HOW BREAKTHROUGH SCIENCE CAN REVIVE ECONOMIC GROWTH AND THE AMERICAN DREAM (2019).

59. *See* VANNEVAR BUSH, SCIENCE: THE ENDLESS FRONTIER 19 (Nat'l Sci. Found. 1960) (1945) ("Basic research . . . creates the fund from which the practical applications of knowledge must be drawn.").

science can yield fundamental understanding. Thus, institutional structures that encourage disclosure of applied science, and not just basic science, should be encouraged.

The case for clearing both system and tool opacity becomes even stronger, of course, to the extent that one's perspective is non-utilitarian. For example, if the field expert has non-utilitarian duties to promote the autonomy of those whom her work will ultimately affect, then she should clearly work towards explainability.[60] An important goal should be to avoid a scenario where people subject to an unexplainable predictive model are essentially held captive by unknown variables that they cannot seek to change.[61] And these non-utilitarian concerns may, of course, be much more substantial when machine-learning is applied in areas outside our focus here, such as predictive policing, political campaigning, or allocation of social resources.[62]

On the assumption that system and tool opacity are worth attempting to reduce, we posit that the last contributor—secrecy associated with tools—is most amenable to policy interventions. Accordingly, Part III analyzes the implications of secrecy in machine learning tools.

To keep the analysis tractable, we focus only on innovation-related harms and benefits associated with secrecy. We therefore assume that disclosure can be made in a manner that retains appropriate safeguards against harm to any individuals from whom the data was derived.[63]

III. SECRECY, COMPETITIVE ADVANTAGE, AND DISCLOSURE

The central obstacle to reducing tool opacity—and to reproducibility by competitors—is secrecy with respect to the learning algorithm, training data, training process, and associated parameters. Thus, the most straightforward approach to clearing opacity derived from secrecy is disclosure of the learning algorithm, associated parameters, and training data. Such disclosure could be fully public—that is, a general disclosure—or more focused, available to field experts only.[64]

---

60. *See* Selbst & Barocas, *supra* note 30, at 1118–19 (describing dignitary justifications for explainability).

61. *Id.*

62. *See generally* Kaminski, *supra* note 30 (describing the requirements of the European Union's General Data Protection Regulation (GDPR)); Selbst & Barocas, *supra* note 30 (exploring the difference between machine learning and other forms of decision-making).

63. *See* Roger Allan Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOMMS. & TECH. L. REV. 1, 29–39 (2016) (describing ways to protect patient privacy while sharing health-related big data). *But see* Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1841–45 (2011) (describing the ability to re-identify de-identified data).

64. For example, data from pharmaceutical clinical trials has traditionally been kept secret; in the wake of recent efforts to increase disclosure, results are now more frequently available, but often only to vetted experts in the field for predetermined legitimate purposes. *See, e.g.*, *Policies & Procedures to Guide External Investigator Access to Clinical Trial Data*, YODA PROJECT, https://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data [https://perma.cc/49PY-USCM] (stating that proposals for access to clinical trial data from partner

This Part considers what disclosure means in the context of machine learning. We begin by describing institutional motivations behind secrecy —why firms keep information secret, how powerful secrecy is, and what competitive advantages it confers. We start with competitive advantage because, particularly in the absence of other innovation incentives, this competitive advantage can be an important incentive. Next, we consider disclosure's benefits, whether directly to field experts, to downstream users and innovators, or to scientific reproducibility and progress more generally. Finally, we illustrate concretely how disclosure in machine learning could work by looking to the issue in a specific incentive context that has been rigorously studied by legal scholars and economists: the patent system. There, disclosure is mandated—but is widely agreed to be largely ineffective.

A. *SECRECY AND COMPETITIVE ADVANTAGE*

As intellectual property scholars have long noted,[65] secrecy can, depending on context, be a relatively weak or strong form of protection. On the one hand, trade secrecy[66] is a relatively low *cost* form of protection—it attaches so long as the information in question confers competitive advantage by virtue of "not being generally known" and "is the subject of [reasonable] efforts . . . to maintain . . . secrecy."[67] On the other hand, trade secrecy does not protect against independent invention or reverse engineering. Therefore, strength of protection can vary considerably depending on the cost of independent invention or reverse engineering (or some combination of the two) faced by competitors.

In general, the source code for the learning algorithm (that is, the program that does the learning) confers only limited competitive advantage. In many cases, the competitor may be able to find open source options —underlying the lack of competitive advantage, some prominent commercial firms make versions of their learning algorithms freely available.[68] Alternatively, depending on resources, the competitor may be able to independently invent and/or reverse engineer the source code.

---

pharmaceutical companies will be reviewed for scientific merit and cannot be used for commercial or litigation purposes).

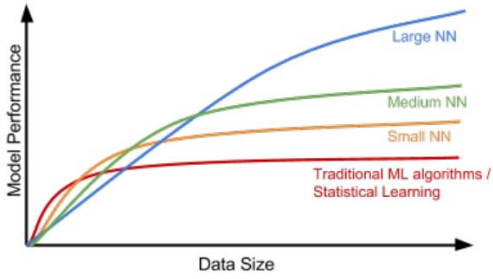
65. *See generally* Jerome H. Reichman, *How Trade Secrecy Law Generates a Natural Semicommons of Innovative Know-How*, *in* THE LAW AND THEORY OF TRADE SECRECY 185 (Rochelle C. Dreyfuss & Katherine J. Strandburg eds., 2011) (summarizing the applications of trade secrecy law); Pamela Samuelson & Suzanne Scotchmer, *The Law and Economics of Reverse Engineering*, 111 YALE L.J. 1575 (2002) (surveying the historical practice of reverse engineering and its application to intellectual property secrecy).

66. For the moment, we elide the concepts of actual secrecy (that is, keeping information from others via locks or other non-legal forms of protection), and trade secrecy (that is, the body of law that allows suits for misappropriation of trade secrets). We discuss some of these distinctions below.

67. UNIF. TRADE SECRETS ACT § 1(4) (NAT'L CONF. OF COMM'RS ON UNIF. STATE L. 1985).

68. *See* MARTÍN ABADI ET AL., TENSORFLOW: A SYSTEM FOR LARGE-SCALE MACHINE LEARNING 265 (2016), https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf [https://perma.cc/MU6Y-RU6T]; *Tools*, FACEBOOK AI, https://ai.facebook.com/tools/pytorch [https://perma.cc/Y55U-E65Q].

Figure 1. Data Size and Model Performance



Adapted from Andrew Ng Talk: https://youtu.be/F1ka6a13S9I

More than learning algorithm source code, well-curated and labelled training data represents a very significant hurdle. As Figure 1, taken from a talk by Andrew Ng shows, the data needs of large neural networks may be particularly significant.[69] Competition policy commentators have, therefore, rightly focused on this latter hurdle as a potentially significant entry barrier.[70]

Reverse engineering training data from trained machine learning models is sometimes possible but can be quite difficult. To be sure, for relatively simple machine learning models available as a service, reverse engineering may be possible. In one case where the machine learning model revealed confidence values associated with its outputs, researchers' repeated query and response engagement with the model resulted in a few thousand query/response pairs. This data was then used as training data for the reverse engineer's own machine learning, and the result was a near-clone that performed similarly to the target machine learning.[71]

However, commentators are doubtful that similar results will emerge for more complex learning models.[72] Thus, at least for the moment, collecting, curating, and concealing reliable training data for these more complex models will continue to serve as a barrier to entry.

69. Figure 1 shows the increase of model performance with dataset size; large neural networks especially benefit from larger datasets.

70. *See, e.g.*, Hal Varian, *Artificial Intelligence, Economics, and Industrial Organization*, *in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE: AN AGENDA 399, 402–04 (Ajay Agrawal, Joshua Gans & Avi Goldfarb eds., 2019); Judith Chevalier, *Comment on Artificial Intelligence, Economics, and Industrial Organization*, *in* THE ECONOMICS OF ARTIFICIAL INTELLIGENCE, *supra*, at 419–21.

71. FLORIAN TRAMÈR, FAN ZHANG, ARI JUELS, MICHAEL K. REITER & THOMAS RISTENPART, STEALING MACHINE LEARNING MODELS VIA PREDICTION APIS 601 (2016), https://www.usenix.org/system/files/conference/usenixsecurity16/sec16_paper_tramer.pdf [https://perma.cc/GCB9-6J7Z].

72. *See* Andy Greenberg, *How to Steal an AI*, WIRED (Sept. 30, 2016, 11:06 AM), https://www.wired.com/2016/09/how-to-steal-an-ai [https://perma.cc/UH47-Z3RP].

Beyond issues of market entry, competition, and price is the issue of disclosure outside the developer firms themselves. Even if firms are competing vigorously over data, and perhaps even innovating incrementally based on this data—and thus there is no problem from a short-term competition policy perspective—the pace of disclosure outside those firms may be quite slow. Below we discuss the benefits of disclosure, particularly to field experts. Some of these benefits are benefits of open science generally. But robust disclosure may be particularly useful for machine learning. Ordinary processes of reverse engineering may be less likely to generate competition and public domain training data at scale. Most importantly, robust disclosure (and interrogation of that disclosure) provides a relatively low-cost mechanism for an initial vetting of machine learning's interesting, but non-intuitive, outputs.

B. BENEFITS OF DISCLOSURE

Disclosure brings many benefits.[73] And while these are described as benefits from disclosure, they could as easily be described in reverse as the harms of secrecy, as the presence of secrecy tends to limit the availability of these benefits.[74] Although we focus on the benefits of disclosure, we are of course mindful of harms as well. First, to the extent other protections are not available, secrecy may provide an important investment incentive.[75] Second, disclosure of low-quality information can lead others down problematic paths[76] and disclosure of accurate information can lead others to follow the

73. We focus here on the potential benefits that arise from disclosure to field experts, rather than to the public generally, to lay actors, or to regulators. That said, the benefits of direct disclosure to non-field-experts may also be substantial. *See, e.g.*, Price, *supra* note 28, at 462–65 (extolling the benefits of disclosure of medical algorithm development information to regulators); Kaminski, *supra* note 18, at 1578–80 (describing the benefits of algorithmic disclosure to lay users); *cf.* Sherkow & Scott, *supra* note 53, at 1544–47 (describing the benefits to users of patent disclosures about gene editing vector technology); *id.* at 1538–40 (noting that secrecy for vectors may decrease the ability of patients to give truly informed consent to their use).

74. Secrecy has other potential harms as well. Physical efforts at secrecy, such as building fences and security systems, can incur suboptimal social costs. *See, e.g.*, Mark A. Lemley, *The Surprising Virtues of Treating Trade Secrets as IP Rights*, 61 STAN. L. REV. 311, 334–36 (2008) (arguing that one benefit of trade secrecy is reducing the need for investment in actual secrecy). Other efforts at secrecy, such as nondisclosure agreements, have been suggested to limit knowledge flows, worker movement, innovation, and economic productivity more generally. *See generally* ORLY LOBEL, TALENT WANTS TO BE FREE: WHY WE SHOULD LEARN TO LOVE LEAKS, RAIDS, AND FREE RIDING (2013) (collecting evidence on the harms of secrecy and trade secrecy). These harms, however, are not our focus here.

75. *See* Price, *Big Data, Patents, & Medicine*, *supra* note 17, at 1432–33.

76. There are myriad examples of disclosure of incorrect information leading to substantial problems down the road. Especially salient today is a 1980 letter to the New England Journal of Medicine—this five-sentence letter concluding (erroneously, and with little data) that opioid treatment for chronic pain carried a low risk of addiction was cited hundreds of times and likely contributed substantially to the opioid epidemic and the lack of physician and scientist belief that opioid addiction was an important risk and worth studying. *See* Pamela T.M. Leung, Erin M. Macdonald, Matthew B. Stanbrook, Irfan A. Dhalla & David N. Juurlink, *A 1980 Letter on the Risk of Opioid Addiction*, 376 NEW ENG. J. MED. 2194, 2194–95 (2017) (criticizing Jane Porter & Hershel Jick, *Addiction Rare in Patients Treated with Narcotics*, 302 NEW ENG. J. MED. 123 (1980)).

same path rather than branching out to new areas.[77] Third, disclosure can also lull policymakers into a false sense of security under the assumption that transparency substitutes for regulation, even when it should not.[78] Fourth, certain types of data, particularly health data, may contain information that could be traced to a particular individual and be used to harm the individual in question.[79] Finally, accurate disclosure could enable sophisticated actors to game the system.[80] Particularly because we focus on scientific research contexts where gaming should not be a first-order concern, we view these harms as relatively minor compared to the benefits: benefits to field experts in terms of understanding, downstream benefits that accrue to others in the system as a result of that understanding, and benefits to science as a whole in terms of promoting the norm of reproducibility. Nevertheless, we return to disclosure harms in Part IV where we discuss policy options.

1. Direct Benefits from Disclosure to Field Experts

In the context of machine learning, a key benefit of disclosure to field experts is to enable further work on penetrating system opacity. As described above, the non-intuitiveness that characterizes many machine-learning models has the silver lining of generating hypotheses which, by their very non-intuitiveness, can open new avenues of scientific exploration.

In the simplified drugs/breakfast/grapefruit example above, imagine that the algorithm simply recommended a dosage to an individual based on an "explanatory" statement that their breakfast preferences were correlated with drug reaction. A bare statement along those lines would make it difficult to learn more, including whether (for example) the algorithm was picking up a real effect or some correlation between eating particular foods and adherence to drug prescriptions. But if the developer disclosed the data on which the algorithm was trained, and how it was trained, field experts might be able to factor out the possibility of such confounds. They might recognize a potential biochemical relationship between particular foods and dosage response/dosage relationship and then test that hypothesis more rigorously. Eventually, experts would presumably identify the true, biochemical relationship between grapefruit consumption and drug efficacy, increasing

77. *See* THOMAS S. KUHN, THE STRUCTURE OF SCIENTIFIC REVOLUTIONS 23–34 (2d ed. 1970); *cf.* Joseph P. Fishman, *Creating Around Copyright*, 128 HARV. L. REV. 1333, 1339 (2015) (arguing that inaccessible areas of inquiry—in his case, because of intellectual property, in ours, secrecy—can promote innovation around those areas).

78. *See generally* OMRI BEN-SHAHAR & CARL E. SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE (2014) (describing in detail how disclosure fails to perform the oversight and regulatory functions for which it is prescribed).

79. *See generally* W. Nicholson Price II & I. Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 NATURE MED. 37 (2019) (outlining the challenges and potential solutions surrounding patient privacy and big data).

80. *See, e.g.*, Bambauer & Zarsky, *supra* note 25, at 44 (describing gaming and countergaming in the context of dynamic algorithmic systems); Arti K. Rai, *Machine Learning at the Patent Office: Lessons for Patents and Administrative Law*, 104 IOWA L. REV. 2617, 2639–40 (2019) (discussing applicant gaming of administrative agency algorithms).

our knowledge in the area. Without disclosure, the development of a properly vetted hypothesis would be much less likely.

Disclosure of information about model training, development, and validation can also allow field experts to probe the behavior of the models themselves; if those models are in use in real-world situations, such probing can serve as a parallel form of oversight to that practiced by regulators.[81]

Field experts involved in such endeavors can come in different flavors. Some may be academics and nonprofit entities engaged mainly in knowledge-development enterprises. For those experts, the availability of disclosure from algorithms which have already been developed and perhaps deployed may be an essential source of data for follow-on efforts. Similarly, field experts that are themselves developers, but with lesser resources, may find disclosure a key part of their own development efforts, lowering barriers to entry into the field and expanding the scope of potential developers. This broadening of entrants increases the potential for the development of different models—with the potential both for greater practical performance,[82] but also for clearing system opacity. A diversity of approach limits the likelihood that any one paradigm becomes fully entrenched, increasing the chance of knowledge development.[83]

Field experts could also come from other, larger entities with the capacity to develop their own algorithms. For experts in large entities, equipped with substantial resources, disclosure from other developers can still create substantial benefits. Most straightforwardly, more data increases the size of potential training and validation sets. As described below, larger datasets can improve performance.[84]

Potentially more important, to the extent that the collectors of large datasets do so from different populations and with different collection strategies, different datasets are likely to incorporate different limitations and different biases.[85] When such datasets are combined as a result of disclosure,

81. *See, e.g.*, Price, *supra* note 28, at 465–73 (suggesting collaborative governance oversight via FDA-mediated disclosure of information about medical algorithms).

82. *See* Camacho et al., *supra* note 12, at 1584 (arguing that one machine learning "rule[] of thumb" in network biology is that combinations of different models produce the most robust results).

83. *See* KUHN, *supra* note 77, at 43–51 (describing entrenched paradigms); Laura G. Pedraza-Fariña & Ryan Whalen, *A Network Theory of Patentability*, 87 U. CHI. L. REV. 63, 98–100 (2020) (describing recombination from diverse fields as the path to groundbreaking innovation). This outcome is not definitive; one could also imagine that disclosure increases the odds that later developers follow similar paths as those trod earlier, decreasing the eventual overall variety of approaches. *See supra* notes 76–77 and accompanying text. Given the resource constraints imposed by big data requirements for machine learning, we think this outcome less likely, but the point is contestable.

84. *See infra* Section IV.A (describing performance benefits from scale).

85. *See, e.g.*, Ruha Benjamin, *Assessing Risk, Automating Racism*, 366 SCIENCE 421, 421–22 (2019) (describing bias in health data and algorithms); Effy Vayena, Alessandro Blasimme & I. Glenn Cohen, *Machine Learning in Medicine: Addressing Ethical Challenges*, 15 PLOS MED. 1, 1–4 (2018) (describing bias, among other challenges); Price, *supra* note 57, at 66 (describing the potential for population-based bias in medical big data and artificial intelligence); Han Liu & Mihaela Cocea, *Granular Computing-Based Approach for Classification Towards Reduction of Bias in*

these differences in representativeness may at least partially counter one another, leading to improved performance. In terms of understanding, the tensions and discrepancies between models trained in different fashions and on different datasets—but with similar goals—can also point to fruitful avenues for future exploratory work.

2. Downstream Benefits of Disclosure to Field Experts

Though we focus on disclosure *to* field experts, those experts are not the only ones benefited by such disclosure. Greater understanding that arises from disclosure to field experts—whether of the machine learning systems themselves or of the complex real-world systems that provide the underlying data—can help other downstream users. For example, system regulators stocked with a greater understanding can better manage the tools that rely on those systems—FDA can evaluate medical algorithms, the Securities and Exchange Commission can monitor trading and markets, and the National Highway Traffic Safety Administration can oversee self-driving vehicles, and all can do so better if they can understand what is going on.

Similarly, system users may be more willing to trust algorithms if they (or others they trust) can understand more about how the systems work, or at least know that such understanding is possible. For medical algorithms, patients might be more willing to follow algorithmic recommendations if at least some facets of those algorithms are disclosed and potentially understood—and, eventually, if we come to learn more about the underlying biological systems on which their recommendations are based.[86]

More broadly, as the public is asked to put its trust in algorithmic decision-making in an increasing variety of spheres, disclosure to field experts may help increase that trust. In part, this would result from the ability of field experts to interrogate the procedural methods by with algorithms are developed, validated, and deployed. The literature on algorithmic accountability has emphasized these issues.[87]

But disclosure to field experts also keeps open the possibility that system opacity, and perhaps even tool opacity, doesn't have to be permanent. In the future, complexity and non-intuitiveness may be better understood and thus become more transparent—that is, decreasing secrecy may eventually

Ensemble Learning, 2 GRANULAR COMPUTING 131, 131 (2017) (discussing potential computational approaches to reducing bias).

86. This claim is speculative. It is also possible that patients—or doctors—don't much care if anyone understands how medical algorithms work, as long as they do work. One of us has assumed this to be the case. *See, e.g.*, W. Nicholson Price II, *Big Data and Black-Box Medical Algorithms*, 10 SCI. TRANSLATIONAL MED. 1, 2–3 (2018). But, he may very well have been wrong; the field awaits empirical evidence to clarify the point.

87. *See, e.g.*, Kaminski, *supra* note 18, at 1564–77; Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473, 526–27 (2016); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 61 (2019); Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633, 684 –88 (2020) (lamenting the lack of sufficient private partners for a collaborative governance approach).

decrease other aspects of opacity, leading to greater interrogability and accountability.

3. Promotion of Reproducibility

Finally, disclosure promotes scientific reproducibility. Core scientific norms view such disclosure and reproducibility as essential for producing verified knowledge.[88] For physical experiments, reproducibility requires that an independent researcher could obtain the same results using the disclosed information about an experiment's conditions, parameters, and equipment —and know that the same results were obtained.[89] Computational reproducibility requires that the same results be obtained from the data and code (ideally executable code) used in the original study.[90]

C. DISCLOSURE IN THEORY AND PRACTICE; THE EXAMPLE OF PATENT LAW

Machine learning is hardly the first technological tool to raise difficult issues regarding secrecy and disclosure that are important for law. To the contrary, the patent system has long grappled with questions of complexity, secrecy, and disclosure.[91] Patent law seeks to advance the "[p]rogress of [s]cience and useful [a]rts,"[92] at least in part through a bargain that involves disclosure enabling others of "ordinary skill" in the field both to practice the disclosed invention and to build on its underlying advance. In patent law, the level of detail demanded is a function of how well-understood the field is. For systems and tools that are opaque from the standpoint of complexity and non-intuitiveness, disclosure requirements are supposed to be high.

Unfortunately, as we also discuss below, the theory of reproducibility through disclosure is not as easily achieved in practice.

---

88. *See generally* Bruce Alberts et al., *Self-Correction in Science at Work*, 348 SCIENCE 1420 (2015) (providing an argument, authored by the former President of the National Academy of Sciences and colleagues, that the norm of reproducibility is necessary for implementing self-correction).

89. If an independent researcher can know that she obtained the same results as the original researcher, the research also has verifiability. A particular challenge arises when tacit knowledge—often know-how—is also needed for reproducibility. *See generally* Laura G. Pedraza-Fariña, *Spill Your (Trade) Secrets: Knowledge Networks as Innovation Drivers*, 92 NOTRE DAME L. REV. 1561 (2017) (describing the importance of know-how to innovation and the emergence of informal networks sharing such know-how).

90. Roger D. Peng, *Reproducible Research in Computational Science*, 334 SCIENCE 1226, 1226 –27 (2011). One step beyond reproducibility is replicability—the true gold standard. Replicability requires not only independent research but use of independent inputs. In physical experiments, reproduction will often take place on different physical equipment in any event so the difference between reproducibility and replicability may not be large. In computational science, by contrast, generating independent input code and data may not be the norm.

In the case of machine learning that relies on an element of randomness (e.g., a random seed), perfect reproducibility may be difficult to achieve. But the greater the disclosure, the greater the likelihood of reproduction.

91. As for non-intuitiveness, the patentability requirement of nonobviousness could be seen as quite similar.

92. U.S. CONST. art. I, § 8, cl. 8.

1. Theory

In recognition of the scientific value of reproducibility, the patent statute requires its own version: it requires the applicant to disclose sufficient information to show the artisan in the scientific or technological field how "to make and use the" invention.[93] Moreover, patent doctrine specifically provides for those cases where the end product invention is opaque and can only be reproduced by a very precise description of the process by which it is made. In those cases, the patent applicant is supposed to apply for a product-by-process patent, which defines and thus claims the covered invention not by describing the actual product, but rather the process used to create it.[94] Indeed, in that subset of product-by-process cases where materials used in the process are not readily reproducible, the applicant is supposed to deposit those materials in an appropriate, publicly accessible repository.[95] In the case of biological materials, this requirement also applies more generally: "Where the invention involves a biological material and words alone cannot sufficiently describe how to make and use the invention in a reproducible manner, access to the biological material may be necessary for the satisfaction of the statutory requirements for patentability under 35 U.S.C. 112."[96]

2. Practice

Regrettably, patent law does not live up to its idealized articulation, in part because the underlying science doesn't always either. NIH Director Francis Collins has famously opined "that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring."[97] Collins notes that in their publications scientists often either do not report basic elements of experimental design "or describe them only vaguely [in order] to retain a competitive edge."[98]

Perhaps even more problematic is irreproducibility in studies that rely heavily on statistical analysis. Here, irreproducibility is linked to pervasive problems of poor statistical design such as small sample sizes, small effect sizes, and suppression of data that does not support the researcher's preferred result; these underlying problems are exacerbated when details of statistical

93. 35 U.S.C. § 112 (2018).

94. *See generally* Dmitry Karshtedt, Note, *Limits on Hard-to-Reproduce Inventions: Process Elements and Biotechnology's Compliance with the Enablement Requirement*, 3 HASTINGS SCI. & TECH. L.J. 109 (2011) (explaining the overarching law concerning product-by-process claims).

95. *See, e.g.*, Price, *Big Data, Patents, & Medicine*, *supra* note 17, at 1428–31 (suggesting algorithmic deposition as a way to satisfy the § 112 enablement and written description requirements).

96. MPEP § 2402 (9th ed. Rev. 10, June 2020).

97. Francis S. Collins & Lawrence A. Tabak, *NIH Plans to Enhance Reproducibility*, 505 NATURE 612, 612 (2014).

98. *Id.*

analyses are not shared.[99] Both problems—poor design and inadequate reporting—appear in patent practice.[100]

The problems go beyond statistics and reporting. In 2015, Iain Cockburn, Tim Simcoe, and Leonard Freedman estimated that U.S. researchers spent about $28 billion per year on irreproducible preclinical studies.[101] As they note, a significant percentage of the problem in preclinical studies can be traced to the unpredictable behavior of antibodies and cell lines.[102]

These opaque biological products also create problems for reproducibility and disclosure in patent law practice. For instance, in order for a complex biologic protein such as a monoclonal antibody to be verifiably reproduced, the cell line and culture conditions in which it was originally produced typically need to be precisely known.[103] Varying those conditions too much can foil attempts to reproduce the biologic, because cell lines and culture conditions interact with the protein's DNA sequence in complex ways that biologists have not fully characterized.[104] For purposes of meeting the disclosure and reproducibility requirements of patent law, complex biologics, therefore, should probably be claimed in product-by-process form.[105] However, the Patent Office has not generally enforced this requirement.[106]

On the positive side, problems with patent practice in biologics yield lessons for machine learning models. Like biological proteins, machine learning models are highly dependent on the specific process and input materials (learning algorithm plus training data) by which they are produced.[107] Thus, at least in theory, the Patent Office has the authority to require learning algorithm, training data, and training process disclosure for patents on machine learning models. (In contrast, for standard rules-based software, at least software that models well-understood systems, disclosure of a basic algorithm should suffice[108]).

---

99. *See generally* John P.A. Ioannidis, *Why Most Published Research Findings Are False*, 2 PLOS MED. 696 (2005) (discussing the issues with published research findings).

100. *See generally* Janet Freilich, *The Replicability Crisis in Patent Law*, 95 IND. L.J. 431 (2020) (finding that many experiments disclosed in patents lack key indicia of reproducibility); Jacob S. Sherkow, *Patent Law's Reproducibility Paradox*, 66 DUKE L.J. 845 (2017) (describing the failure of many patented inventions, especially pharmaceuticals, to actually work in practice).

101. Leonard P. Freedman, Iain M. Cockburn & Timothy S. Simcoe, *The Economics of Reproducibility in Preclinical Research*, PLOS BIOLOGY, June 9, 2015, at 1, 3.

102. *Id.* at 5–6.

103. Price & Rai, *supra* note 54, at 1035–36.

104. *Id.* at 1034–36.

105. Karshtedt, *supra* note 94, at 139.

106. *See* Price & Rai, *supra* note 54, at 1051.

107. Subbaswamy & Saria, *supra* note 5, at 345–46. If one were inclined to take the parallel between biologics and machine learning to its logical extreme, the DNA sequence that codes for the biologic would be analogous to the learning algorithm's source code. The cell line would be analogous to training data.

108. *See* Mark A. Lemley, *Software Patents and the Return of Functional Claiming*, 2013 WIS. L. REV. 905, 926 (discussing the requirements of patent claim language).

In principle, then, the advent of machine learning could provide the Patent Office an opportunity to revisit product-by-process claiming and require such claiming for both complex biologics and complex machine learning models. In that regard, the Patent Office's recent request for information on artificial intelligence patenting, which includes questions about whether such patenting poses unique disclosure issues, is a promising sign.[109]

We turn to this possibility in Part IV, along with other innovation policy levers that could potentially produce disclosure, and ultimately first-principles scientific knowledge, at a somewhat faster pace.

IV. DISCLOSURE AND UNDERSTANDING: INNOVATION POLICY LEVERS

Although secrecy is the aspect of opacity most amenable to policy interventions, the interventions we discuss below don't necessarily take as their primary aim the goal of reducing secrecy. Indeed, simply requiring unilateral disclosure by firms entails complex substantive and political-economy challenges,[110] suggesting that a broader range of possibilities should be considered. Equally important, the vigorous policy conversation around data governance has already suggested, or even put into place, a variety of different policy levers for achieving substantively important goals *other than* secrecy reduction. These policy levers could do double duty by working not only to achieve their primary goal but also to reduce secrecy.

Accelerated disclosure of data and methodology underlying machine-learning algorithms could emerge from a number of different institutional settings. These include publicly funded "big data" projects (e.g., the Human Genome Project, the Cancer Genome Atlas, or All of Us), decentralized data generation in publicly funded academic institutions, patent filings, FDA regulatory processes, or demand side private sector actors like insurers that have custody of large amounts of electronic health records ("EHR") data. The freshest, and potentially most promising, context for disclosure may be the regulatory approach of "pre-certification" currently being considered by the FDA. Part IV presents each of these possibilities independently, though the optimal policy strategy is likely a combination of different approaches.

A. PUBLIC FUNDING FOR LARGE SCALE DATA GENERATION

One possibility to increase disclosure—at least of the data enabling machine learning—is public funding for the generation of large-scale data that can then be used by many actors for both basic and applied tasks. For instance, the NIH's All of Us cohort study, created as part of the Precision Medicine Initiative, aims to generate and capture large amounts of data, including genome sequences, on at least one million Americans from varying

109. Request for Comments on Patenting Artificial Intelligence Inventions, 84 Fed. Reg. 44,889 (Aug. 27, 2019).

110. *See* Price & Rai, *supra* note 54, at 1054–55 (describing industry resistance to mandated disclosure).

racial and ethnic categories as well as socioeconomic backgrounds.[111] The project envisions creating a high-quality dataset that is broadly available.[112]

Although not all of this data will necessarily have been curated and labeled in a manner necessary to represent machine-learning-ready training data for a given research project, some will be. Additional curation and labeling will produce additional training data. Machine-learning algorithms trained on such data could point to (possibly non-intuitive) scientific hypotheses, and the public nature of the dataset would allow various actors to explore those hypotheses.

Some public data-generation efforts have explicit safeguards to drive public disclosure. The Human Genome Project, for example, included guidance that researchers involved not seek patents on the genes they sequenced,[113] and that sequence data be immediately disclosed.[114] Heidi Williams has found that publicly disclosed data generated by the Human Genome Project generated more downstream scientific and commercial output than parallel genomic sequence data produced by Celera Genomics —which was protected by trade secrecy and by contractual nondisclosure language.[115]

B. IMPROVING DISCLOSURE IN THE PATENT SYSTEM

As a threshold matter, patent lawyers may object to the patent lever on the grounds that recent, vaguely worded Supreme Court decisions that may make software harder to patent[116] will spur developers of machine learning

111. *See About*, NIH: ALL OF US RSCH. PROGRAM, https://allofus.nih.gov/about [https://perma.cc/X84E-3XJB].

112. *Opportunities for Researchers*, NIH: ALL OF US RSCH. PROGRAM, https://allofus.nih.gov/get-involved/opportunities-researchers [https://perma.cc/7DPD-4N5R].

113. Jorge L. Contreras, *Leviathan in the Commons: Biomedical Data and the State*, *in* GOVERNING MEDICAL KNOWLEDGE COMMONS 19, 28 (Katherine J. Strandburg, Brett M. Frischmann & Michael J. Madison eds., 2017). Note that under the Supreme Court's decision in *Ass'n for Molecular Pathology v. Myriad Genetics, Inc.*, isolated genes are no longer patentable subject matter. Ass'n for Molecular Pathology v. Myriad Genetics, Inc., 569 U.S. 576, 596 (2013).

114. Jorge L. Contreras, *Bermuda's Legacy: Policy, Patents, and the Design of the Genome Commons*, 12 MINN. J.L. SCI. & TECH. 61, 85 (2011).

115. Heidi L. Williams, *Intellectual Property Rights and Innovation: Evidence from the Human Genome*, 121 J. POL. ECON. 1, 14 (2013). Williams and her co-author Bhaven Sampat elsewhere note that some forms of IP can be important for future work, finding that *patented* genes appear, ex ante, to be more valuable than unpatented genes. Bhaven Sampat & Heidi L. Williams, *How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome*, 109 AM. ECON. REV. 203, 214–15 (2019).

116. *See generally* Alice Corp. v. CLS Bank Int'l, 134 S. Ct. 2347 (2014) (expanding the scope of the "abstract idea" exception to patentable subject matter); *Myriad Genetics, Inc.*, 569 U.S. 576 (expanding the scope of the "products of nature" exception to patentable subject matter); Mayo Collaborative Servs. v. Prometheus Lab'ys, Inc., 566 U.S. 66 (2012) (expanding the scope of the "law of nature" exception to patentable subject matter). Indeed, because the Supreme Court decisions take aim at both the natural sciences and software, machine learning in health care may be particularly vulnerable. *See* Price, *Big Data, Patents, & Medicine*, *supra* note 17, at 1420–26 (arguing that patent protection for artificial intelligence in medicine is weak under § 101).

models to avoid the system.[117] Indeed, some recently published law firm data indicate that grant rates for patents on the use of machine learning in certain areas, including health care, appear relatively low.[118] Thus far, however, this law firm data also show that machine learning patent *applications* in many areas, including health care, continue to rise.[119]

To the extent developers of machine-learning models and tools continue to pursue patents, the patent system can help promote meaningful disclosure. A simple mechanism for improving disclosure could be a Patent Office rule mandating product-by-process disclosure, at least for certain categories of complex machine learning.[120] Thus, developers seeking to patent machine learning products could be required not only to give detailed information about the training process but also to deposit training data and the trained machine learning models into a depository which would make the information publicly available. Although the Patent Office does not have rulemaking authority over the core requirements of patentability,[121] patent law already provides for this type of disclosure.[122] Such a mandate should, at least in principle, be an option for the Patent Office.

To be sure, the political economy of the Office implementing such a requirement would be tricky. But the substantive case for disclosure is much stronger for machine learning algorithms than for standard software, for the

---

117. Kate Gaudry & Samuel Hayim, *Artificial Intelligence Technologies Facing Heavy Scrutiny at the USPTO*, IPWATCHDOG (Nov. 28, 2018), https://www.ipwatchdog.com/2018/11/28/artificial-intelligence-technologies-facing-heavy-scrutiny-uspto/id=103762 [https://perma.cc/D4BJ-45TN] (discussing the impact of the Supreme Court's *Alice* decision).

118. KILPATRICK TOWNSEND & GREYB SERVS., INDUSTRY-FOCUSED PATENTING TRENDS 17 (2019), https://apps.kilpatricktownsend.com/files/Patent%20Trends%20Study.pdf [https://perma.cc/48MG-EU5M] (finding in 2013, grant rates for artificial intelligence-related applications in health care and financial technology were around 60 percent and slightly less in digital marketing and education); *see also* Mateo Aboy, Cristina Crespo, Kathleen Liddell, Timo Minssen & Johnathon Liddicoat, Mayo*'s Impact on Patent Applications Related to Biotechnology, Diagnostics and Personalized Medicine*, 37 NATURE BIOTECHNOLOGY 513, 515 (2019) ("[While] *Mayo* has had a substantial impact on patent prosecution in the life sciences . . . our results also show that the impact of *Mayo* may not be as devastating for biotech, diagnostics and personalized medicine patent applications as many commentators have stated.").

119. KILPATRICK TOWNSEND & GREYB SERVS., *supra* note 118, at 28. Continued patent application in the face of persistent rejection by the Patent Office is a puzzle that Colleen Chien and one of the authors are currently exploring.

120. It is worth noting that the scope of resulting patents, like the scope of product-by-process patents more generally, would likely be relatively narrow. W. Nicholson Price II, *Describing Black-Box Medicine*, 21 B.U. J. SCI. & TECH. L. 347, 355 (2015).

121. *E.g.*, Jonathan S. Masur, *Regulating Patents*, 2010 SUP. CT. REV. 275, 276 ("[T]he Patent and Trademark Office (PTO) has never had substantive rule-making authority."); Arti K. Rai, *Growing Pains in the Administrative State: The Patent Office's Troubled Quest for Managerial Control*, 157 U. PA. L. REV. 2051, 2053 (2009) ("[T]he PTO lack[s] substantive rulemaking authority . . . ."); Arti K. Rai, *Patent Validity Across the Executive Branch: Ex Ante Foundations for Policy Development*, 61 DUKE L.J. 1237, 1270 (2012) ("[T]he PTO . . . does not have rulemaking authority over questions of patent validity . . . .").

122. Enzo Biochem, Inc. v. Gen-Probe Inc., 323 F.3d 956, 965 (Fed. Cir. 2002) (holding that a written reference to publicly deposited biological material can satisfy the written description requirement of 35 U.S.C. § 112); USPTO Rules of Practice in Patent Cases, 37 C.F.R. § 1.802 (2019) (implementing the depository requirement).

reasons described above. Moreover, even in the latter case, sustained pressure on the Office and the courts by politically powerful large information technology firms impeded by "bad patents" did lead to legal requirements that software patents disclose a basic algorithm.[123]

One substantive challenge with a product-by-process requirement would be that U.S. law allows patents to be filed early in the research and development process.[124] And, as a practical matter, patents are indeed filed quite early.[125] This institutional structure is not optimal for machine learning models, which are often engineered for improvement based on learning from new data. The optimal disclosure regime might therefore be quite different from the current static regime and involve continual updating of disclosure during the life of the patent.[126]

Indeed, in the context of other inventions that continually improve, scholars have suggested changing the patent statute to require disclosure updates. Jeanne Fromer has proposed that patentees be required to disclose all commercial embodiments of a particular patented invention, which among other goals would "reveal[] helpful technological information."[127] One of us has argued that patent law should incorporate an "economic enablement" requirement.[128] Patentees would be required to disclose enough information—likely after initial patenting—to enable competitors to market competing products once the patent term has expired.[129] Disclosed

123. Kevin Emerson Collins, *Patent Law's Functionality Malfunction and the Problem of Overbroad, Functional Software Patents*, 90 WASH. U. L. REV. 1399, 1451–60 (2013); *see* Supplementary Examination Guidelines for Determining Compliance with 35 U.S.C. 112 and for Treatment of Related Issues in Patent Applications, 76 Fed. Reg. 7162, 7162–7172 (Feb. 9, 2011).

124. In this respect, U.S. patent law, whether by accident or design, adopts a "prospect" theory of patents. On this view, broad patents are similar to ordinary property rights and should be granted early because they can then serve as a focal point for efficient development and commercialization of the inventive "prospect." Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265, 276 (1977); F. Scott Kieff, *The Case for Registering Patents and the Law and Economics of Present Patent-Obtaining Rules*, 45 B.C. L. REV. 55, 66 (2003) (explaining that, while "the prospect and rent dissipation theories provide important insights about how the patent system can both increase and decrease rent dissipation-type social costs," they fall short in other areas). Not surprisingly, the literature arguing for and against this approach is voluminous. John F. Duffy, *Rethinking the Prospect Theory of Patents*, 71 U. CHI. L. REV. 439, 441–42 (2004) (listing scholars who praise and scholars who criticize prospect theory).

125. *See* Christopher A. Cotropia, *The Folly of Early Filing in Patent Law*, 61 HASTINGS L.J. 65, 71 (2009) (describing early filing practice and exploring the problems associated with it, including "too many applications, too many patents, underdevelopment of patented technology, and increased assertion of patent rights"); Mark A. Lemley, *Ready for Patenting*, 96 B.U. L. REV. 1171, 1195 (2016) (noting that laws encouraging early patent filing "encourag[e] ideas at the expense of those who take the time to develop and test their inventions").

126. *Cf.* Price & Rai, *supra* note 54, at 1043 (arguing for continuing disclosure in the context of opaque biologic manufacturing processes).

127. Jeanne C. Fromer, *Dynamic Patent Disclosure*, 69 VAND. L. REV. 1715, 1716 (2016). As noted earlier, *see generally* Fromer, *supra* note 20, Fromer has also suggested, on the subject of this Essay, that trade secrecy law protects machine learning too strongly and should be modified accordingly.

128. Price, *Expired Patents*, *supra* note 26, at 1632–40.

129. *See id.*

information would include manufacturing processes, interchangeability standards, or other knowledge that carries through the patent bargain of limited monopoly followed by free competition.[130] Finally, Jacob Sherkow, responding specifically to concerns of irreproducibility of inventions disclosed in patent applications, argues that courts and the PTO should consider evidence that arises after patent filing in determining when the invention is enabled.[131] A patent that discloses irreproducible information serves no teaching function,[132] and irreproducibility is often knowable only well after the patent has been filed.[133]

A requirement of dynamic disclosure would, however, be well beyond the scope of the Patent Office's current authority. Congress would have to confer such authority on the Office, and the political economy of that option probably makes dynamic patent disclosure a dead letter.[134]

If a risk regulator has jurisdiction over the machine-enabled model, a potentially greater level of disclosure through the risk regulator is an option. The next Section considers data aggregation and disclosure through the risk regulator, in this case the FDA.

C. AGGREGATION AND DISCLOSURE THROUGH THE RISK REGULATOR

In the 21st Century Cures Act of 2016, Congress explicitly gave the FDA regulatory authority over most clinical interventions enabled by machine learning.[135] In essence, Congress stated that FDA can regulate software that recommends, or makes, clinical decisions in situations where the scientific basis for the recommendation is not likely to be independently understood by the physician.[136] FDA could use this power to require submission of training data and model source code for machine learning products that recommend or make clinical interventions without sufficient transparency.[137]

More broadly, by virtue of its role as the pre-market regulator of biological risk and benefit introduced by therapies (both small molecules and biologics), FDA has custody of troves of human biological data that could

130. *See id.*

131. Sherkow, *supra* note 100, at 907–11.

132. *Id.* at 903–05.

133. *Id.* at 908.

134. On the other hand, to the extent that dynamic disclosure results in continuing fees to the PTO, there might be more institutional interest than would be initially suspected.

135. *See* 21st Century Cures Act § 3060, 21 U.S.C. § 360j(o) (2018).

136. *See id.*

137. Research by one of us indicates that FDA has not, thus far, been asking for disclosure of actual training data or model code. *See* Arti K. Rai, Isha Sharma & Christina Silcox, *Accountability, Secrecy, and Innovation in AI-Enabled Clinical Decision Software*, J.L. & BIOSCIENCES, Nov. 14, 2020, at 1, 5, 24. In an interesting recent article, Andrew Tutt invokes the power and reputation that the FDA has historically enjoyed and calls for an FDA-type federal agency to monitor performance of machine-learning algorithms. *See generally* Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83 (2017) (calling for a federal agency to oversee performance of machine learning algorithms). As a consequence of the 21st Century Cures Act, we already have an FDA for certain machine learning algorithms.

serve as training data for machine learning. Well before recent advances in machine learning, FDA's failure to disclose this data when it was authorized to do so as a matter of statute—that is, after statutory IP exclusivities held by the data originator had expired—was heavily criticized.[138]

As with the Patent Office's failure to use its power to impose product-by-process requirements, FDA's position can largely be traced to political economy. And as with the Patent Office, perhaps the advent of machine learning will shift the calculus.

To FDA's credit, it responded to pressure on data disclosure by setting up a pilot program for originators of clinical trial data on branded therapeutics that wish to disclose data voluntarily; although the pilot has concluded, FDA continues to explore data sharing possibilities.[139] As with private sector efforts that have also been launched to disclose data,[140] this voluntary approach may be able to capitalize on the desire of drug makers to cultivate an image of trustworthiness with patients, physicians, and the general public.

In the specific case of machine learning models, a voluntary "Pre-Certification" program that the FDA is piloting for software products may provide a test bed for further experimentation with disclosure. The program recognizes the reality that software products, particularly software products enabled by machine learning, should be encouraged to change relatively rapidly as they incorporate new data. Thus, the current system of pre-market approval, together with requirements that the product manufacturer apply for updated approval for every subsequent change, is not necessarily an optimal structure.

Under the regime, firms can get pre-certified as "organizations that perform high-quality software design and testing."[141] FDA discussion drafts indicate that pre-certification would provide a streamlined path to initial market entry, and would allow for frequent updates, in exchange for continuous monitoring by the FDA of real-world distribution and use. Based on these promises by the FDA, multiple major software development firms have signed up for the program.[142]

---

138. For a detailed history of the FDA's failure to exercise its statutory authority, and resulting criticisms, see generally Arti K. Rai, *Risk Regulation and Innovation: The Case of Rights-Encumbered Biomedical Data Silos*, 92 NOTRE DAME L. REV. 1641 (2017). That article does not, however, address machine learning.

139. Press Release, Janet Woodcock, Dir. of the Ctr. for Drug Evaluation & Rsch., FDA, FDA Continues to Support Transparency and Collaboration in Drug Approval Process as the Clinical Data Summary Pilot Concludes (Mar. 26, 2020), https://www.fda.gov/news-events/press-announcements/fda-continues-support-transparency-and-collaboration-drug-approval-process-clinical-data-summary [https://perma.cc/M5JR-Q66K].

140. *See* Rai, *supra* note 138, at 1652, 1652 nn.57–58 (discussing efforts by multiple drug makers in the Clinical Study Data Repository and by the collaboration between Johnson & Johnson and Yale ("YODA")); *see also supra* note 64 (describing YODA's access policies).

141. FDA, SOFTWARE PRECERTIFICATION PROGRAM, *supra* note 29, at 7.

142. Danielle Kosecki, *How Apple, Fitbit, Samsung and More Are Helping to Modernize the FDA*, CNET (Sept. 5, 2019, 5:30 AM), https://www.cnet.com/health/how-apple-fitbit-samsung-and-more-are-helping-to-modernize-the-fda [https://perma.cc/KQ82-J6QJ].

One additional obligation that the FDA could impose (presumably within its statutory power since the program is voluntary) is data disclosure after a specified period of time, including training data, algorithmic change (based on incorporation of new data), and performance attributes. Since the FDA would already be actively monitoring data on real-world distribution and use, it should be able to enforce that obligation without much additional administrative burden. Meanwhile, the pre-certifying firm might be able to monetize the mantle of trustworthiness data disclosure could provide, either directly or as a defense in any potential product liability action.

All that said, though we and others have suggested FDA-mediated disclosure in different contexts before,[143] we should note that not all are sanguine about the prospect. Jacob Sherkow and Christopher Scott, while recognizing that the FDA does serve an information disclosure role, are skeptical of expanding that brief, at least in the context of manufacturing vectors for gene therapy: "One can . . . imagine a regime where the FDA is both statutorily authorized and administratively willing to mandate maximum disclosure regarding inputs for therapeutic manufacture. But that is purely imaginative. The FDA is both legally prohibited from requiring the disclosure of confidential business information from clinical trials and culturally unwilling . . . ."[144] One need not agree with their legal conclusion[145] to agree that the political economy of enhanced regulatory disclosure represents a real challenge.[146]

D. DEMAND-SIDE DISCLOSURE

Payers present another important opportunity for disclosure that is underdeveloped in the literature that applies specifically to health innovations. For health technologies, payers—whether insurers, integrated health systems, or public payers like Medicare or Medicaid—serve gatekeeping roles by determining which technologies will be reimbursable.[147] One of us has suggested, in work with Rebecca Eisenberg, that insurers thus have an important role in innovation.[148] Some payers are already playing this

143. *See, e.g.*, Eisenberg, *supra* note 27, at 380–84 (clinical trial data); W. Nicholson Price II, *Regulating Secrecy*, 91 WASH. L. REV. 1769, 1802–12 (2016) (biopharmaceutical product details generally); Price, *supra* note 28, at 465–73 (information about medical algorithms); Price & Rai, *supra* note 54, at 1053–56 (biologic manufacturing processes); Rai, *supra* note 138, at 1666 (diagnostic test data); *see also* Rachel E. Sachs & Thomas J. Hwang, *Increasing the Transparency of FDA Review to Enhance the Innovation Process*, *in* TRANSPARENCY IN HEALTH AND HEALTH CARE IN THE UNITED STATES 185, 185 (Holly Fernandez Lynch, I. Glenn Cohen, Carmel Shachar & Barbara J. Evans eds., 2019) (existence and details of investigational new drug clinical trials).

144. Sherkow & Scott, *supra* note 53, at 1544.

145. *See, e.g.*, Eisenberg, *supra* note 27, at 380 ("[T]he statutory language invoked in support of [the FDA's] position [on withholding clinical data from public disclosure] is ambiguous."); Rai, *supra* note 138, at 1655 ("[T]he FDA's position [on disclosing trial data] appears overly conservative.").

146. *See, e.g.*, Rai, *supra* note 138, at 1657–58 (noting the FDA's continued resistance to releasing data, even with greater statutory authority).

147. Eisenberg & Price, *supra* note 27, at 26.

148. *Id.* at 5.

technology-evaluation role, including by analyzing the troves of health data they already possess.[149] Payers could extend this function by, for instance, disclosure of the sort we have been describing. Although large private insurers have substantial power on their own,[150] Medicare is a particularly influential payer in this space, because where it leads, many private insurers follow.[151] Thus, policymakers could exert potentially substantial influence on the disclosure of machine-learning tools by changing Medicare policy to require some forms of disclosure as a condition of payment—with the potential that private payers might follow suit.[152]

Such disclosure could, at a minimum, be to field experts within the payers themselves. That is to say, payers can employ experts in health big data and machine learning that can evaluate disclosures by algorithm developers. This role would likely fit more neatly within a parallel oversight role (that is, ensuring performance and efficiency of the purchased technology) than a public knowledge-development role, but could still be useful.[153]

Payers could also require more public disclosure, enabling the work of field experts more generally. As a condition of agreeing to reimburse a technology, payers could require that technology developers make available their data, development practices, and validation practices to either the general public or to a subset of qualified field experts.[154] Such a requirement could align with payers' own goals in several ways. It could outsource part of the parallel oversight role, enabling academics, nonprofits, or other field experts to find problems or to validate algorithmic development and performance.[155] The development of fundamental knowledge could also help promote further technological development that might be of value to payers. And if understanding does indeed increase user trust and consequent adoption, such increased adoption of money-saving technologies could redound to payer benefit.[156]

Finally, payers could also potentially advance knowledge by disclosing the data they already have, including algorithmic performance data and

149. *Id.*

150. *See id.* at 27 (discussing the role of market concentration among health insurers).

151. Rachel E. Sachs, *Prizing Insurance: Prescription Drug Insurance as Innovation Incentive*, 30 HARV. J.L. & TECH. 153, 196–200 (2016).

152. *See id.* at 201–08 (suggesting deliberate use of Medicaid policy to set innovation incentives).

153. *See supra* text accompanying note 81 (discussing the potential for disclosure to enable private parallels to regulatory oversight).

154. *See supra* note 64 and accompanying text.

155. *See, e.g.*, Price, *Big Data, Patents, & Medicine*, *supra* note 17, at 1451–52 (suggesting "bounties" for entities that validate the performance of (or expose flaws in) algorithms developed by others).

156. If new technologies are more expensive, this could have deleterious consequences for payers as users demanded more expensive treatments. *See, e.g.*, Eisenberg & Price, *supra* note 27, at 5 (discussing this point).

other real-world evidence.[157] While these are not data about algorithmic development, they could shed light on performance and places where algorithmic results are in tension with ground truth, opening possibilities for further probing. Such disclosure raises challenges endemic to those of health data: Electronic health records and claims data often have substantial quality problems because of the clunkiness of health data systems,[158] incentives to record skewed data for payment purposes,[159] and the general fragmentation of health data.[160]

It is worth noting that the role of insurers is different outside of the health sphere. In health, insurers can exercise control by deciding whether or not to pay for a technology (and paying for technology absent insurance is often impossible). In other fields, insurers can already exercise control by deciding whether or not to issue a policy that may be a legal or practical requirement for a certain activity.[161] Auto insurers, for instance, currently constrain individual driving behavior through underwriting and rate-setting of individual drivers.[162] But auto insurers could similarly exercise their

157. Ross Koppel & David Kreda, *Health Care Information Technology Vendors' "Hold Harmless" Clause: Implications for Patients and Clinicians*, 301 JAMA 1276, 1277 (2009).

158. Sharona Hoffman & Andy Podgurski, *Big Bad Data: Law, Public Health, and Biomedical Databases*, 41 J.L. MED. & ETHICS 56, 57 (2013) (describing quality issues in big data in medicine).

159. *See generally* ROBERT WACHTER, THE DIGITAL DOCTOR: HOPE, HYPE, AND HARM AT THE DAWN OF MEDICINE'S COMPUTER AGE (2015) (discussing the integration of technology into the medical field and its different effects on the profession, particularly the doctor-patient relationship).

160. *See* W. Nicholson Price II, *Risk and Resilience in Health Data Infrastructure*, 16 COLO. TECH. L.J. 65, 69–73 (2017) (describing fragmentation of health data systems). That said, the recent rush of technology firms to collaborate with hospital systems to secure access to EHR documents, *see generally* I. Glenn Cohen & Michelle M. Mello, *Big Data, Big Tech, and Protecting Patient Privacy*, 322 JAMA 1141 (2019), does suggest that problems with quality are not insuperable. *See, e.g.*, Emily Schweich, *The University of Chicago Medicine Collaborates with Google on Machine Learning Research*, AM.'S ESSENTIAL HOSPS. (Aug. 14, 2017), https://essentialhospitals.org/university-chicago-medicine-collaborates-google-machine-learning-research [https://perma.cc/7Z8D-PNPD] ("Researchers from UChicago Medicine . . . are teaming with Google to use machine learning to find patterns in electronic health records (EHRs) and use those patterns to predict readmissions, complications, and other hospital-acquired conditions."); Natasha Singer & Daisuke Wakabayashi, *Google to Store and Analyze Millions of Health Records*, N.Y. TIMES (Nov. 11, 2019), https://www.nytimes.com/2019/11/11/business/google-ascension-health-data.html [https://perma.cc/LGL7-38H2] ("In a sign of Google's major ambitions in the health care industry, the search giant is working with [the Ascension] hospital system to store and analyze the data of millions of patients in an effort to improve medical services . . . .").

161. *See generally* Omri Ben-Shahar & Kyle D. Logue, *Outsourcing Regulation: How Insurance Reduces Moral Hazard*, 111 MICH. L. REV. 197 (2012) (discussing the general phenomenon of regulation by insurers and providing examples); Kyle D. Logue, *Encouraging Insurers to Regulate: The Role (If Any) for Tort Law*, 5 U.C. IRVINE L. REV. 1355 (2015) (describing why insurers often regulate indirectly rather than mandating behavior); John Rappaport, *How Private Insurers Regulate Public Police*, 130 HARV. L. REV. 1539 (2017) (discussing how private insurers regulate the conduct of police agencies). Health insurers cannot typically exercise this sort of influence over individual insured behavior because federal law requires community rating, whereby individuals in a certain group are all charged the same rate, and guaranteed issue, whereby insurers are required to issue policies if sought. Patient Protection and Affordable Care Act § 1201, 42 U.S.C. §§ 300gg–300gg-1 (2018).

162. Kyle D. Logue, *The Deterrence Case for Comprehensive Automaker Enterprise Liability*, 2019 J.L. & MOBILITY 1, 15–17.

influence to require disclosure of information about self-driving vehicle algorithms in much the same way as described above—whether to the insurers themselves or to field experts more broadly—by, for instance, refusing to insure a car with autonomous capabilities unless data about its algorithms was disclosed.[163] Such disclosures could inform not only the development of the tools (autonomous driving algorithms) but also knowledge about the underlying systems (traffic dynamics, the prevalence of road hazards, and other risk elements).

To the extent that insurers are concerned about systemic risk,[164] disclosure requirements could also reduce that risk and might therefore be of interest. Notably, these insurer actions are often less amenable to policy action because non-health insurance markets may lack a parallel to the massive, public, market-leading Medicare program. There is no massive federal auto insurer that the market follows, for instance. Thus, innovation *policy* interventions aiming to facilitate disclosure by non-health insurers would likely be more indirect. Nevertheless, the possibility of disclosure-forcing behavior by private, non-health insurers may be an important piece of the disclosure picture for machine learning tools.

E. OBJECTIONS AND INTERNATIONAL IMPLICATIONS

We address briefly a set of objections to our proposals. Most of these objections flow from the potential harms from disclosure that we noted earlier. These harms may have particular salience when placed in an international context.

First, in certain contexts, data disclosure may include disclosure of information that is traceable to an individual or could, if combined with other sources of data, be made traceable to an individual.[165] Although federal law

163. The economics of such demands need study. Perhaps drivers would flock to non-disclosure-requiring insurers, creating competitive pressure against requiring disclosure. But in the story suggested here, insurers gain enough information from disclosure to price more accurately or to reduce risk better, such that they receive a competitive advantage from that disclosure. The growing presence of systems to record driving behavior suggests at least the possibility of insurer-mandated disclosure. Lilia Filipova-Neumann & Peter Welzel, *Reducing Asymmetric Information in Insurance Markets: Cars with Black Boxes*, 27 TELEMATICS & INFORMATICS 394, 394 (2010).

The story of insurers for fully autonomous vehicles, in a future system, is likely to be governed by product liability—but that's a different story. *See generally* Kenneth S. Abraham & Robert L. Rabin, *Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era*, 105 VA. L. REV. 127 (2019) (describing the current landscape, near future, and far future and proposing a new liability regime).

164. U.S. health insurers seem to be less concerned with overall systemic risk and costs, perhaps because the fragmentation of the American health-care system means that the costs of risks for any given patient are likely to be borne by someone else if those risks manifest in the future—and that someone else will be the government through Medicare if the risks happen far enough in the future. Eisenberg & Price, *supra* note 27, at 22 n.119.

165. *See* W. Nicholson Price II, Margot E. Kaminski, Timo Minssen & Kayte Spector-Bagdady, *Shadow Health Records Meet New Data Privacy Laws*, 363 SCIENCE 448, 448–50 (2019) (detailing how third parties have developed "shadow health records"); Cohen & Mello, *supra* note 160, at 1141 (noting that, in today's world, the quantity of information about individuals gained from

in the United States has, at least thus far, viewed removal of key identifiers from data, combined with various anti-discrimination safeguards, as providing sufficient protection against (respectively) traceability and harm,[166] numerous commentators have criticized federal law.[167]

More concretely, various regimes (including certain states with the United States, such as California) have moved beyond U.S. federal law.[168] The European Union's General Data Protection Regime, for instance, creates strong privacy protections.[169] Disclosure of underlying data from regimes with strong individual privacy protections faces additional hurdles; machine-learning developers with models trained on EU and U.S. data would have a harder time meeting disclosure mandates than those with models trained on U.S. data alone.[170] This disparity could bias the types of disclosure and the contours of resulting scientific endeavors.

A second group of concerns arise from competition among jurisdictions for machine-learning developers. If disclosure regimes are considered onerous, we might see machine-learning developers move to jurisdictions with weaker disclosure regimes, potentially promoting a race to the bottom or preventing disclosure efforts from getting off the ground. To the extent that disclosure is tied to the locus of the *market*, this concern may be ameliorated; if developers want to use machine-learning algorithms in products sold in the United States, moving the firm to a less-disclosure-promoting jurisdiction will not relieve them of obligations tied to U.S. markets.

The third set of concerns arises from what one might consider data protectionism. If the United States were to implement policies designed to force or encourage disclosure, and other jurisdictions did not—or, indeed,

---

their internet activity, their smartphone geolocation data, and "highly penetrant and often interlinked" EHRs and the technology that compiles this information "mean that individuals can often be identified in deidentified data by triangulating data sources").

166. Health Insurance Portability and Accountability Act of 1996, 45 C.F.R. § 164.514(b)(2)(i)(A)–(R) (2017) (listing the 18 key identifiers to be removed).

167. *See* Price & Cohen, *supra* note 79, at 39; Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, AM. J. BIOETHICS, Sept. 1, 2010, at 3, 9 (concluding that the deidentification requirements of 45 C.F.R. § 164.514 and the Federal Policy for the Protection of Research Subjects (Common Rule) are "insufficient to protect privacy and respect autonomy in research" and that "[i]t is indefensible from technical, ethical, and policy standpoints to continue drawing a bright-line regulatory distinction between identifiable and deidentified health information").

168. Price et al., *supra* note 165, at 448.

169. *See* Paul M. Schwartz & Karl-Nikolaus Peifer, *Transatlantic Data Privacy Law*, 106 GEO. L.J. 115, 128–29 (2017) ("The EU's recourse to a regulation follows from its recognition of privacy as a human right and the high status of the data subject. . . . [Thus], the GDPR provides directly binding statutory protection in EU law for her."); Kaminski, *supra* note 30, at 192–93 ("The GDPR contains a significant set of rules on algorithmic accountability, imposing transparency, process, and oversight on the use of computer algorithms to make significant decisions about human beings," and Kaminski understands it "to create a broader, stronger, and deeper algorithmic accountability regime than what existed under the EU's Data Protection Directive (DPD).").

170. As noted above, *supra* note 66, we elide privacy issues in this Essay by assuming disclosure mechanisms that sufficiently protect individuals from harm, but the difficulty in designing such mechanisms will vary based on jurisdiction.

adopted policies that explicitly *discouraged* disclosure—resulting dynamics would be complex, since disclosure can easily travel across international borders. We might expect firms in non-disclosing jurisdictions to gain a competitive advantage because they can use both their proprietary data and also data disclosed under policies like those we have suggested here. These concerns are obviously quite salient to U.S. policymakers, particularly in light of the strong interest expressed by competitors like China in machine learning.[171]

As noted earlier, our levers for promoting disclosure do not extend to contexts such as national security where adversarial attack and gaming are first-order concerns. Particularly outside those contexts, we could imagine that a disclosure-promoting environment might create localized benefits that could counteract competitive advantages from data protectionism, looking to the rich literature on trade secrecy, noncompete and nondisclosure agreements, and innovation clusters.[172] These complex issues provide rich topics for future work.

V. CONCLUSION

The various potential policy levers described in Part IV have their own strengths and weaknesses. Promoting disclosure to field experts will likely involve some combination of those levers, depending on the context. For instance, regulator-mediated disclosure, while potentially very powerful in the area of life sciences, will be unavailable in fields without strong regulatory gatekeepers. For the life sciences, an area where we have suggested that algorithmic disclosure is particularly salient, all the levers we mention are available, and the right combination depends on some mixture of political economy, technocratic efficiency, and redundancy to ensure effectiveness. We aim to begin that conversation, not to complete it.

At the end of the day, our argument here is simple. There is an ongoing conversation about the extent to which machine-learning algorithms need to be disclosed to those impacted directly by the technology, whether users of the algorithms or people about whom decisions are made. This is a valuable conversation, but it is incomplete.

There exists a *separate* set of reasons to promote disclosure about algorithms, on many levels, to field experts. These reasons emerge from, but also go beyond, traditional arguments in favor of open science. They relate to the need to sift out from the universe of non-intuitive correlations that machine learning can generate the most promising new hypotheses for

171. Sarah O'Meara, *China's Ambitious Quest to Lead the World in AI by 2030*, 572 NATURE 427, 427 (2019) (describing China's goal "to lead the world when it comes to artificial intelligence" and noting that China's 2017 "New Generation Artificial Intelligence Development Plan[] has spurred myriad policies and billions of dollars of investment in research and development").

172. *See generally* LOBEL, *supra* note 74 (advocating for a new view of what creates successful innovation ecosystems); Orly Lobel, *Noncompetes, Human Capital Policy & Regional Competition*, 45 J. CORP. L. 931 (2020) (outlining the current state of research on human capital and economic competition and anticipating future areas of research in the field).

further investigation. To the extent this happens, machine learning models may help us not only with prediction but with fundamental understanding of the opaque real-world systems that the models are meant to probe. Those reasons should be part of the conversation—and should, we argue, drive us to adopt a set of innovation policy levers to promote robust disclosure of algorithms' datasets, methods, and parameters to experts in the field.