We Robot 2018 Conference Draft

When AIs Outperform Doctors:
The dangers of a tort-induced over-reliance on machine learning and what (not) to do about it

A. Michael. Froomkin*, Ian Kerr† & Joëlle Pineau‡

* Laurie Silvers & Mitchell Rubenstein Distinguished Professor of Law, University of Miami. Member, University of Miami Center for Computational Science; Fellow, Yale ISP. Thanks to Peter Asaro, Jack Balkin, Caroline Bradley, Ryan Calo, Kate Crawford, Brad DeLong, Ed Felten, Jonathan Frankle, Sue Gluck, James Grimmelmann, Woody Hatzog, Margo Kaminsky, Gregory Keating, Mark Lemley, Christopher Millard, Helen Nissenbaum, Paul Ohm, Frank Pasquale, Laurel Riek, Andrew Selbst, Latanya Sweeney and Joel Zysman for advice and information..

† Canada Research Chair in Ethics, Law & Technology, University of Ottawa, Faculty of Law, with cross appointments to the Faculty of Medicine, Department of Philosophy and School of Information Studies. Thanks to the Social Sciences and Humanities Research Council and the Canada Research Chairs program for their generous support.

‡ William Dawson Scholar and Associate Professor, School of Computer Science, McGill University.

Someday, perhaps soon, diagnostics generated by machine learning (ML) will have demonstrably better success rates than those generated by human doctors. What will the dominance of ML diagnostics mean for medical malpractice law, for the future of medical service provision, for the demand for certain kinds of doctors, and—in the longer run—for the quality of medical diagnostics itself?

This article argues that once ML diagnosticians, such as those based on neural networks, are shown to be superior, existing medical malpractice law will require superior ML-generated medical diagnostics as the standard of care in clinical settings. Further, unless implemented carefully, a physician's duty to use ML systems in medical diagnostics could, paradoxically, undermine the very safety standard that malpractice law set out to achieve. In time, effective machine learning could create overwhelming legal and ethical pressure to delegate the diagnostic process to the machine. Ultimately, a similar dynamic might extend to treatment also. If we reach the point where the bulk of clinical outcomes collected in databases are ML-generated diagnoses, this may result in future decision scenarios that are not easily audited or understood by human doctors. Given the well-documented fact that treatment strategies are often not as effective when deployed in real clinical practice compared to preliminary evaluation, the lack of transparency introduced by the ML algorithms could lead to a decrease in quality of care. The article describes salient technical aspects of this scenario particularly as it relates to diagnosis and canvasses various possible technical and legal solutions that would allow us to avoid these unintended consequences of medical malpractice law. Ultimately, we suggest there is a strong case for altering existing medical liability rules in order to avoid a machine-only diagnostic regime. We argue that the appropriate revision to the standard of care requires the maintenance of meaningful participation by physicians in the loop.

Contents

Introduction

Someday, perhaps sooner,[1] perhaps later,[2] machines will have demonstrably better success rates at medical diagnosis than human physicians, at least in particular medical specialties.[3]

We can reasonably expect that machine-learning -based diagnostic competence, which we will sometimes call "AI" for short, will only increase. It is thus appropriate to consider what the dominance of machine-based diagnostics might mean for medical malpractice law, the future of medical service provision, the demand for certain kinds of physicians, and—in the longer run—for the quality of medical diagnostics itself.

In this article, we interrogate the legal implications of superior machine-generated diagnosticians, particularly those based on neural networks, currently a leading type of machine learning used in prediction.[4] We argue that existing medical malpractice law will eventually require superior ML-generated medical diagnosis as the standard of care in clinical settings. We further argue that—unless implemented carefully—a physician's duty to use ML in medical diagnostics could, paradoxically, undermine the very safety standard that malpractice law set out to achieve. Once mechanical diagnosticians demonstrate better success rates than their human trainers, effective machine learning will create legal (and ethical) pressure to delegate much if not all of the diagnostic process to the machine. If we reach the point where the bulk of clinical outcomes collected in databases are ML-generated diagnoses, this may result in future decision scenarios that are difficult to validate and verify. Many ML systems currently are not easily audited or understood by human physicians and, if this remains true, it will be harder to detect sub-par performance, jeopardizing the system's efficacy, accuracy, and

[1] See *infra* text at notes 10 to 17.
[2] See *infra* text at notes 22 to 27.
[3] See *infra* text at notes 19 to 21.
[4] Machine learning (ML) is the discipline of automated pattern recognition and making predictions based on patterns that are detected. Neural networks are one of several types of ML. "Deep Learning," another term of use, refers to neural networks with many layers. AI is a more general term applied to automated techniques that mimic human reason. Thus, deep learning systems are a subset of neural networks, which are a subset of ML, which is itself a subset of AI. IBM's Watson, which we also discuss, is perhaps the best-known example of a neural network-based medical diagnostic system. See *infra* text at notes 28 to 32.

reliability. We maintain that such unintended consequences of medical malpractice law must be avoided, and canvass various possible technical and legal solutions.

Our story has four acts.

1) We begin with the effect of existing law on the use of ML diagnostic technology, be it neural networks or some other form of AI. We argue that once a machine is demonstrably superior to human diagnosticians, malpractice law will require the use of the superior technology in certain sectors of medical diagnostics. Medical service providers who do not use machine-learning systems will be said to fall below the appropriate standard of care in cases where things go wrong, and hospitals that use human physicians rather than ML systems will be subject to claims in negligence–as will the treating physicians themselves.

2) Next, we consider the consequences that these novel legal requirements might have on the overall demand for physicians of certain types and the potentially diminished role that they might play in medical practice. We suggest that, in the same way that enhanced safety and efficacy will increase the demand for robot drivers and decrease the demand for truckers and occupations employing human drivers, so too will the advent of superior ML diagnosticians reduce the demand for human physicians.[5] These consequences, flowing from the requirements imposed by medical malpractice law, give rise to various narratives. To the extent that patient outcomes are now better and perhaps even cheaper—depending on automated system service provider pricing—these newly imposed legal requirements offer a desirable neoliberal result: better living through technology. Of course, the possible outcomes also comport just as well with the classic account of deskilling: overreliance on these machines could render obsolete the human cultivation of medical skills and knowhow developed over centuries.[6] Indeed, robotic surgery is already being accused of causing a loss of surgical skill among medical trainees.[7] That law has mandated the use of a new technology that produces improved health outcomes might also make this tale a happy outlier to more

[5] It will likely increase demand for certain types of medical technicians. A similar economic logic applies to robot surgeons and other medical specialties as they get robotized.

[6] See, e.g., Harry Braverman, Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century (1974); The Degradation of Work?: Skill, Deskilling, and the Labour Process (Stephen Wood, ed. 1982), *but see* Paul Attewell, *The Deskilling Controversy*, 14 Work & Occupations 323 (1987); Stanley Aronowitz & William DiFazio, *High Technology and Work Tomorrow*, 544 Annals Am. Acad. Pol. & Soc Sci. 52 (1996).

[7] *See* Matthew Beane, *Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail*, ADMINISTRATIVE SCIENCE QUARTERLY, https://doi.org/10.1177/0001839217751692 (Jan 9, 2018).

familiar stories of the law's interaction with technology, those in which law is disrupted by the technical change and perhaps even seeks to hold it back.[8]

3) Regardless of which narrative best describes our second act, we believe there is a third act that must also be considered: the development of a diagnostic monoculture and other dangers associated with an over-reliance on ML. By "diagnostic monoculture" we mean a scenario in which the medical and legal systems standardize on a particular mechanized approach to diagnosis in a given sub-specialty. Diagnostic monoculture exemplifies a more general problem that arises when we come to rely, to our detriment, on a dominant mode of thinking to the exclusion of other possible solutions. In this case, a diagnostic monoculture that leads to less input from human physicians could make quality control of diagnostic databases much more difficult. The problem becomes far more serious once reliance on ML goes beyond diagnosis to treatment. The reduction in new data from physicians—that is to say the creation of a loop in which outcomes added to the database are solely or overwhelmingly the result of ML-informed treatment decisions—creates scenarios in which more sub-optimal conclusions are reached. If a set of symptoms is consistently producing an erroneous ML diagnostic, and physicians act on that erroneous diagnostic, where will ML get the data to suggest a different diagnosis which lead to better treatment? If the answer is "nowhere" then we have a problem. Worse, it is not even clear that either the ML system or an outside observer necessarily would know that the results were sub-optimal. From a human perspective, the challenges associated with understanding and auditing an ML system's predictive diagnostic process will become significant. Those challenges become greater if the output of the ML diagnostic system is then fed into a second ML treatment system. In that case, absent personalized medicine, for any given set of symptoms one might get consistent treatment decisions leading to less variegated treatment-to-outcome data. The lack of variety in treatment could further mask any issues caused by sub-optimal diagnosis, and could lead to bad decision-making and, potentially, tragic medical outcomes. To guard against this possibility, we will need a mechanism. And until we know how to automate that too we may need a substantial corps of medical researchers on tap to help audit and monitor the machines in order to spot anomalies.

---

[8] For example, the DMCA is sometimes accused of propping up outdated or anticompetitive business models in the face of easy content-sharing. See, e.g., Ryan J. Shernaman, *The Digital Millennium Copyright Act: The Protector Of Anti-Competitive Business Models*, 80 UMKC L. Rev. 545 (2011). Likewise, DMCA-type legislation has also been shown to undermine privacy: Ian Kerr, Kerr, *If Left to Their Own Devices ... How DRM and Anti-Circumvention Laws Can be Used to Hack Privacy*, in IN THE PUBLIC INTEREST: THE FUTURE OF CANADIAN COPYRIGHT LAW (Michael Geist, ed., 2005), available at SSRN: https://ssrn.com/abstract=902448.

4) The approach taken in our fourth act is speculative and involves exploring different possible future scenarios and potential solutions. Our starting point imagines a future in which the reliability of the diagnostic ML is high enough that the human physician seems unnecessary or even—to the extent she may overrule valid diagnoses—unhelpful insofar as her inputs tend to reduce the probability of a successful outcome. We consider technological fixes in response to an ML monoculture, and also whether better liability rules might avoid or at least postpone the problem. One complicating factor which we must consider is that law is not the only driver here: Even without the malpractice push, if the price is right, economics could incentivize a very similar evolution. In either case, it is essential to examine several potential means of avoiding the risks associated with an ML diagnostic monoculture and an over-reliance on ML.

A possible legal strategy would be to change existing medical malpractice rules and thus reduce the incentives that drive medicine to reduce its reliance on people. We propose meaningful human participation in diagnostics as an essential requirement of the standard of care. This will blunt the legal aspect of the push towards replacing physicians with ML. But, alone, even this rule is not enough, as malpractice law will still tend to stay the human's hand in individual cases: if a physician overrides the machine, the physician (and her employer if any) will be taking a terrible malpractice risk if it remains the case that the machine has a significantly better probability of success on its own then does the physician. We thus will also need to formulate new rules that balance the social interest of having human judgement in the loop with the individual patient's interest in getting the best outcome. This, however, requires that we consider thorny ethical and legal issues.

Unless we are very confident in our technical solutions, we argue, there is a strong case for altering existing medical liability rules in order to maintain focus—when it comes to determining the appropriate role of humans and machines in medical diagnostics—on both ethics[9] and cost rather than defensive medicine. A revision of the standard of care to avoid allowing a machine-only diagnostic regime, would require meaningful participation by people in the loop. As such it risks being expensive since the machine will cost money and the rule we propose will negate potential cost savings from reducing the number of physicians in reliance on the new technology. We suggest, however, that our proposal could be a first step in preventing law from overriding these other important considerations, preserving

[9] See Jason Millar and Ian Kerr, *Delegation, Relinquishment, and Responsibility: The Prospect of Expert Robots* in ROBOT LAW (Ryan Calo, A. Michael Fromkin & Ian Kerr eds 2016).

many long-term beneficial outcomes that would otherwise be at risk due to pressure from the legal system and from cost-cutting.

I. Once a ML System is Demonstrably Superior, Malpractice Law Will Require That Medical Service Providers Use It

It seems inevitable that – at least for some medical specialties – ML diagnosticians someday will have demonstrably better success rates than human physicians. A number of ongoing initiatives suggest that ML will have, or perhaps already has,[10] great diagnostic power for a variety of diseases and conditions ranging from oncology to drug discovery. Google's neural net diagnoses skin cancer as effectively as do experienced dermatologists.[11] IBM's Watson uses oncological data to diagnose cancers that humans have difficulty identifying.[12] Google has tested an AI-based system that successfully identified eye diseases in retinal fundus photographs.[13] Other programs already beat humans: AIs beat humans at predicting heart attacks – without even considering the effects of diabetes or lifestyle.[14] "Machine-learning significantly improves accuracy of cardiovascular risk prediction, increasing the number of patients identified who could benefit from

---

[10] *See* Ian Steadman, *IBM's Watson Is Better at Diagnosing Cancer than Human Doctors*, WIRED (February 11, 2013), http://www.wired.co.uk/article/ibm-watson-medical-doctor.

[11] *See* Andre Esteva et al. *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115 (2017), https://www.nature.com/nature/journal/v542/n7639/full/nature21056.html.

[12] See Steve Lohr, *IBM Is Counting on Its Bet on Watson, and Paying Big Money for It*, NY TIMES B1 (Oct 17, 2016), http://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html.

[13] Varun Gulshan et al, Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, 316 J. AM. MED. ASSOC. 2402 (2016); see also Ariel Bleicher, Teenage Whiz Kid Invents an AI System to Diagnose Her Grandfather's Eye Disease, IEEE SEPCTRUM (Aug. 3 2017), https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/teenage-whiz-kid-invents-an-ai-system-to-diagnose-her-grandfathers-eye-disease (describing creation of "Eyeagnosis, a smartphone app plus 3D-printed lens that seeks to change the diagnostic procedure from a 2-hour exam requiring a multi-thousand-dollar retinal imager to a quick photo snap with a phone").

[14] Lulu Chang, *Machine Learning Algorithms Surpass Doctors At Predicting Heart Attacks*, DIGITAL TRENDS, http://www.digitaltrends.com/health-fitness/ai-algorithm-heart-attack/ (last visited Apr 19, 2017).

preventive treatment, while avoiding unnecessary treatment of others."[15] Researchers at MIT and Harvard are using ML for Alzheimer detection.[16] Similarly, "Watson for Drug Discovery rank ordered all of the nearly 1,500 genes within the human genome and proposed predictions regarding which genes might be associated with ALS. … eight of the top 10 ranked genes proved to be linked to the disease. More significantly, the study found five never before linked genes associated with ALS."[17] Diagnostic medicine seems a particularly good fit for what today's AIs can do best – pattern recognition – as well as being an area with real room for improvement. Five percent of U.S. adults who seek outpatient care each year experience a diagnostic error, leading to six to seventeen percent of adverse events in hospitals.[18]

Radiology seems to be a specialty particularly suited to replacement by ML.[19] One study reports that an AI correctly detected 92.4% of breast cancer tumors compared to the 73.2% detected correctly by human doctors.[20] Indeed University of Toronto Professor Geoffrey Hinton argues that radiologists are about to be obsolete: "I think that if you work as a radiologist you are like Wile E. Coyote in the cartoon … You're already over the edge of the cliff, but you haven't yet looked down. There's no ground underneath. … It's just completely obvious that in five years deep

---

[15] Stephen F. Weng et al., Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data?, 12 PLOS ONE e0174944 (2017).

[16] *See Predicting Change in the Alzheimer's Brain*, MIT CSAIL, http://www.csail.mit.edu/predicting_change_in_the_alzheimers_brain (last visited Sep 22, 2017); Adrian V. Dalca et al, *Predictive Modeling of Anatomy with Genetic and Clinical Data* (2015), http://www.mit.edu/~adalca/files/papers/miccai2015_predictiveModelling_precr.pdf.

[17] *Barrow Identifies New Genes Responsible for ALS using IBM Watson Health*, http://www.prnewswire.com/news-releases/barrow-identifies-new-genes-responsible-for-als-using-ibm-watson-health-300378211.html (last visited Feb 10, 2017)

[18] *See* Nicholas P. Terry, *Appification, AI, and Healthcare's New Iron Triangle*, at 45 (citing Institute of Medicine, Improving Diagnosis in Health Care (Sept. 2015)), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3020784.

[19] *See* Katie Chockley & Ezekiel Emanuel, *The End of Radiology? Three Threats to the Future Practice of Radiology*, 13 J. AMCOLL RADIOL. 1415. (Sept. 18, 2016), doi:10.1016/j.jacr.2016.07.010.

[20] Yun Liu et al., *Detecting Cancer Metastases on Gigapixel Pathology Images* ARXIV:1703.02442 [CS] (2017), http://arxiv.org/abs/1703.02442 (last visited Oct 27, 2017) (stating "[a]t 8 false positives per image, we detect 92.4% of the tumors, relative to 82.7% by the previous best automated approach. For comparison, a human pathologist attempting exhaustive search achieved 73.2% sensitivity."). Currently, however, the ML system's false positive rate remains greater than that of humans, *see* Dayong Wang et al, *Deep Learning for Identifying Metastatic Breast Cancer*, arXiv preprint arXiv:1606.05718 (2016)

learning is going to do better than radiologists.[21] Hyperbole notwithstanding, many ML experts share Hinton's vision regarding the 'inevitable' demise of human medical diagnosis for conditions where we have large amounts of high-quality data.

Skeptics point to issues with current trials and suggest that ML, not to mention AI, superiority remains purely speculative,[22] and that IBM's advertising over-promises what Watson can do.[23] Oren Etzioni, CEO of the Allen Institute for AI, went as far as to say that "IBM Watson is the Donald Trump of the AI industry—outlandish claims that aren't backed by credible data."[24] Indeed, IBM Watson's "Oncology Expert Advisor" suffered a high-profile setback when the University of Texas's cancer center cancelled a flagship collaboration when the project foundered on incompatibilities with hospital records system and alleged violations of hospital procurement regulations.[25] In the end, the "project appeared to fall apart because of cost overruns related to incompatible IT platforms and the extraordinarily complex work involved in structuring and preparing massive amounts of data to be ingested by Watson's machine learning systems." [26] Even a state-of-the-art AI was no match for "the idiosyncrasies of medical records: the acronyms, human errors, shorthand phrases, and different styles of writing."[27]

---

[21] Siddhartha Mukherjee, *A.I. Versus M.D.*, THE NEW YORKER, Apr. 3, 2017, http://www.newyorker.com/magazine/2017/04/03/ai-versus-md (last visited Apr 12, 2017).

[22] See, e.g., Casey Ross & Ike Swetlitz, *IBM Pitched Watson As a Revolution in Cancer Care. It's Nowhere Close*, STAT, Sept. 5, 2017, https://www.statnews.com/2017/09/05/watson-ibm-cancer/ (last visited Sep 7, 2017)

[23] A particularly egregious example is IBM, *Watson at Work*, https://www.youtube.com/watch?v=2cMQZ2l5XlU in which "Watson" has a dialog with basketball scouts on the court--although reportedly the Toronto Raptors are in fact using a version of Watson to help them rank scouted players based on various numerical metrics. *See* IBM, *Seeing Things the Other Teams Can't is the Key to Victory*, https://www.ibm.com/watson/stories/basketball-with-watson.html.

[24] Jennings Brown, *Why Everyone Is Hating on IBM Watson—Including the People Who Helped Make It*, GIZMODO, Aug. 10, 2017, http://gizmodo.com/why-everyone-is-hating-on-watson-including-the-people-w-1797510888 (last visited Sep 7, 2017) (quoting Oren Etzioni).

[25] *See* Matthew Herper, *MD Anderson Benches IBM Watson In Setback For Artificial Intelligence In Medicine*, FORBES (Feb 19, 2017), https://www.forbes.com/sites/matthewherper/2017/02/19/md-anderson-benches-ibm-watson-in-setback-for-artificial-intelligence-in-medicine.

[26] John Battelle, *A Trio of Tech Takedowns*, NEWCO SHIFT, Jul. 17, 2017, https://shift.newco.co/a-trio-of-tech-takedowns-b931c0df5ef6 (last visited Jul 22, 2017). *See also* Herper, *supra* note 25.

[27] Ross & Swetlitz, *supra* note 22.

There is no question that Watson has enjoyed a friendly press and significant hype.[28] It is also the case that not everything IBM currently markets as "Watson" is really true ML. For example, "Watson for Oncology" has been touted as giving "the same recommendations as professional oncologists in 99 percent of the cases"[29] in a test at the University of North Carolina. But the program is really a decision-support tool enhanced with preprogrammed suggestions based on what a committee of doctors at Sloan Kettering said they would do when presented with various symptoms and scenarios.[30]

We should not, however, allow the real ML wheat to be obscured by the marketing chaff. ML systems are being used for everything from dress designing to cooking, roadside assistance, business messaging, education, and movie direction.[31] In particular, researchers are using ML systems, including Watson, to find tumors in radiological data[32] making these the paradigmatic examples of the genre.

A. Machine Learning

1. ML Algorithms Today

At their core, ML systems are simply algorithms designed to draw on data to answer questions. Depending on the design of the algorithm, and the type and amount of data available, an ML system can answer very simple questions, such as predicting the expected weight gain for a patient receiving a given medication, or more complex questions, such as analyzing brain scans and delineating the location of a tumor.

The basic components of an ML system include:

---

[28] Mary Chris Jaklevic, *MD Anderson Cancer Center's IBM Watson Project Fails, and So Did the Journalism Related To It*, HEALTHNEWSREVIEW.ORG, Feb 23, 2017, https://www.healthnewsreview.org/2017/02/md-anderson-cancer-centers-ibm-watson-project-fails-journalism-related/ (last visited Jul 22, 2017)

[29] Ben Dickson, *How Artificial Intelligence Is Revolutionizing Healthcare*, THE NEXT WEB, https://thenextweb.com/artificial-intelligence/2017/04/13/artificial-intelligence-revolutionizing-healthcare/ (last visited Sep 30, 2017).

[30] "That training does not teach Watson to base its recommendations on the outcomes of these patients, whether they lived, or died or survived longer than similar patients. Rather, Watson makes its recommendations based on the treatment preferences of Memorial Sloan Kettering physicians." Ross & Swetlitz, *supra* note 22.

[31] Will Knight, *IBM's Watson is Everywhere—But What Is It?*, MIT TECH. REV. (Oct. 27, 2016), https://www.technologyreview.com/s/602744/ibms-watson-is-everywhere-but-what-is-it/.

[32] See Chockley & Emanuel, *supra* note 19.

- **Input**: The training examples fed into the algorithm. The examples are described by a set features (e.g. doctors' notes, clinical results, time-series recordings, images, etc.) that the machine will observe.
- **ML Algorithm**: The computer program that will digest the data and make a prediction (e.g. linear regression, neural networks, decision trees). We include in this component both the computer's representation of the knowledge extracted and the optimization routine used to train the representation.
- **Evaluation**: The criteria by which we measure the algorithm's performance (e.g. classification accuracy, prediction error, false positive rate)
- **Output**: The information that is produced by the algorithm for a given examples (e.g. predicted weight gain, tumor location, primary health outcome, recommended treatment strategy, prescribed medication dosage.)

In this article, we distinguish between ML systems that make **predictions** and ML systems that make **interventions**. Most of the components may be very similar in both cases, the distinction is primarily in terms of the output. Prediction-type ML systems produce outputs designed to inform medical personnel, enhance their knowledge, situational awareness, and understanding, which they can incorporate in their own decision-making about treatment strategy. Intervention-type ML systems produce outputs that are actionable and can be applied directly, such as a clinical test request, a prescription, or in some cases a direct intervention. Examples of interventions include the case of a neuro-stimulation device using ML to decide the timing and intensity of electrical stimulation applied to a patient with epilepsy in hope of reducing the incidence of seizures, or an artificial pancreas using ML to adapt the dosage of an implanted insulin pump on a diabetic patient.

While from a technical perspective Prediction-type ML and Intervention-type ML can be built using analogous technology and data, the distinction between them is potentially important in the context of discussing medical malpractice law because of the different degrees of human intervention that occur before the ML output is applied to a patient. It might seem obvious that a human's liability for relying on ML will be greater in the Intervention-ML scenario than in the mere Prediction-ML scenario. After all, if ML is only being used for prediction, there clearly is a human in the loop making the treatment decision rather than dare we say mechanically following the dictates of the Intervention-ML. In our view, however, the liability distinction between the two is less sharp than it may seem: If the downstream human's reliance on the Prediction-ML was the source of the patient's bad outcome, but this reliance was reasonable given the Prediction-ML's track record and/or its being part of the standard of care, then the liability of the human under the Prediction system may be no greater than under the Intervention system.

Neural networks are but one type of ML algorithms designed to answer questions using data. Earlier methods, including linear regression, decision trees, and simple probabilistic models, have been used for years to make predictions. Currently, researchers are making particularly rapid progress in training neural networks, especially those with many layers ("deep learning"), to recognize increasingly complex patterns in data. Neural networks are now the method of choice to analyze high-dimensional data, including images of all types, sound, and natural language text. Their power resides in their ability to extract patterns from large datasets with relatively little prior knowledge about useful features or variables.

A critical element of deep learning is that it trains synthetic 'neurons' in multiple layers, which extract information at different levels of abstraction.[33] One can think of each neuron as a simple unit of computation (typically performing a linear equation, followed by a non-linear transform). Groups of neurons are assembled into layers. Each neuron in a layer is in communication with the ones in the layer above it; and each successive layer tends to learn to recognize more general features of the network's input. The neurons in the very first layer observe the Input (raw) data. The neurons in the final layer are responsible for producing the Output.

A common denominator of all ML algorithms, including neural networks, is that they require training. Training methods vary, but they all depend on access to a sufficient – and usually quite large[34] – body of accurate training data. For tumor detection, the data set might be a set of input images, along with the annotations from expert radiologists about the target output (e.g. simple tumor / no tumor classification, or a detailed tumor contour segmentation).[35] The fact that the images come with a human-annotated label is crucial.[36] The ML algorithm relies on having

---

[33] A more formal description appears in David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams, *Learning Representations by Back-Propagating Errors*, 323 NATURE 533 (1986).

[34] *See* Prakash Jay,*Transfer Learning Using Keras Towards Data Science*, https://medium.com/towards-data-science/transfer-learning-using-keras-d804b2e04ef8 (last visited Nov 4, 2017) (noting that with "small" datasets of under 40,000 examples "it is difficult to achieve decent accuracy" for computer vision problems).

[35] *See* Permaln School of Medicine, Section for Biomedical Image Analysis, *Multimodal Brain Tumor Segmentation Challenge 2017*, http://braintumorsegmentation.org/ (last visited Jan 29, 2018).

[36] Some ML algorithms are trained by Unsupervised learning to recognize patterns without human labels. Current state-of-the-art for these techniques still lags behind Supervised learning, so we do not dwell on these approaches here.

that pairing between Input and Output in the data, and the process of "training" the ML system corresponds to the computer learning how to set its own representation such as to reliably select a good output for any new input it might observe. A key component of the training procedure is to assess the expertise level of the ML algorithm throughout training. This is typically done by keeping a portion of the data (e.g. 10%) aside as a "validation set", against which the results of the training will be evaluated using the specified Evaluation criteria.

Another significant feature for our purposes is that neural network systems are rarely static. Even after the successful processing of the initial training data, there are many reasons why one would want to give a deep learning AI additional data to digest. The most obvious is that additional data offers the possibility of better predictions. This is true when the new data is simply a greater quantity of the same type of data (e.g. more x-rays graded by experts), and assuming the data comes from the same distribution (i.e. collected in the same way, annotated in the same way, from the same type of patients). It is not inevitably the case, however, that more data is always better, in particular data collected from a different hospital, potentially with slight variations in procedure, may confuse the ML system. It is important to be vigilant about the quality of the data used to train the system, and in particular to ensure that the data used for training is collected under the same conditions as the ML system will be used in practice. If the inputs from which the ML is to make its decision change in some way over time, the deep learning system will need to be re-trained with new representative data. Changes in data distribution are not uncommon, and might be due to quality degradation caused by aging equipment,[37] or due to quality improvements resulting from the invention of better and more accurate data acquisition equipment (e.g. the invention of better quality imaging machines). Without representative examples of the new information, the AI will not be able to make the best predictions from them,[38] and indeed could in theory go badly wrong.[39]

Due to the very large number of variables, large neural networks are often thought to have a 'black-box' quality. In reality, it is possible to track very precisely the computation at each neuron and each layer. However, it is often difficult to extract a simple explanation for the decision at the end layer (output), since it depends on the combination of many small decisions by each neuron.[40] This

---

[37] *See* ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING 275 (2014)
[38] *See* ALPAYDIN, *supra* note 37, at 275.
[39] *See* MASASHI SUGIYAMA & MOTOAKI KAWANABE. MACHINE LEARNING IN NON-STATIONARY ENVIRONMENTS (2012).
[40] *See* IAN GOODFELLOW, YOSHUA BENGIO & AARON COURVILLE, DEEP LEARNING § 6.2 (2016).

highlights an important distinction: most machine learning algorithms have high traceability (they run on a computer, and can be re-run several times to generate the same results) but poor explainability (cannot extract a compact narrative explaining the logic behind their reasoning). In contrast, humans tend to have poor traceability (difficult to track, at the neural level, reasons for our decisions), but high interpretability (we can easily construct narratives to explain our behaviors). Neural networks in particular do not typically extract causal relationships between inputs and outputs, therefore it is important to interpret any relationship between input and output as a predictive one, no matter how intuitive such relationships might look on the surface."[41]

2. Our Assumptions about Tomorrow

For the purposes of this article, we make two predictive assumptions, one about AI's capabilities and one about its limits. As regards AI's abilities, we assume that at some future date—which may come soon—an ML has been shown to be measurably superior to humans in some specialized aspect of diagnostic medicine. We make this assumption because current trends point strongly in that direction given ML's advances in tumor-detection[42] as well as other areas.[43] For our purposes – and that of the legal system – a new diagnostic technique such as a ML system is superior if its diagnostic accuracy is greater to a statistically significant degree. For simplicity we assume here that the ML either makes fewer false positives (Type I errors) and no more false negatives[44] (Type II errors), or that it makes fewer false negatives and no more false positives, or that the ML system improves on humans to a statistically significant extent in both types of error.[45]

---

[41] *See* Cary Conglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147 (2017); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION 87 (2016); FRANK PASQUALE, THE BLACK BOX SOCIETY (2015).

[42] Chockley & Emanuel, *supra* note 19.

[43] For example, ML has made significant progress advancing computer vision, speech recognition, and machine translation.
See supra text at notes 10 to 21.

[44] Unsurprisingly, false negatives are the errors most likely to create malpractice claims in radiology. *See* Antonio Pinto and Luca Brunses, *Spectrum of diagnostic errors in radiology*, 2 WLRLD J. RADIOL. 377 (20101), doi: 10.4329/wjr.v2.i10.377.

[45] It is also possible that malpractice law might determine that an ML system which made substantially fewer false negative diagnoses but also a small number of increased false positives was legally superior either on its own or in conjunction with human diagnostician, but we need not consider that distracting case to make our argument.

It is also likely that even if a machine learning system has a better success rate than the average human doctor, ML and humans combined might be still better.[46] There are some reasons to suspect that today the combination might beat either one alone, as is the case in "centaur chess."[47] We also know that at present neural networks can make confident but erroneous identifications that no human would make.[48] Keeping a human around protects against those obvious errors, and might protect against other kinds of errors as well.

Indeed, if machine + human is demonstrably better than machine alone, then the combination will become the standard of care through the ordinary operation of the legal system without the need for external intervention unless the combination is seen as prohibitively expensive. At least until ML gets very good, there are scenarios in which the human doctor's role evolves more than evaporates. If ML makes prediction and correlation cheaper, that arguably increases the value of other inputs.

Even in this scenario, however, machine + human remains the standard of care only so long as AI technology does not improve to where the ML system alone is as good at some activity as machine + human. At that point, we posit, the ML system alone becomes, or suffices to meet, the standard of care for that activity (e.g. diagnosis), and the problems discussed below all reappear, making a policy intervention necessary. Perhaps at that point humans will need to switch to other activities such as "the application of ethics, and for emotional support" – and

---

[46] For context see *infra* text at notes 69 to 78.

[47] "The best chess players in the world are human-machine teams"—so long as teams are not time-limited for moves. Paul Scharre, *supra* note 129, at 39-40.

[48] *See* Anh Nguyen, Jason Yosinski & Jeff Clune, *Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images*, IEEE COMP. VISION & PATTERN RECOGNITION (2015), http://arxiv.org/pdf/1412.1897v4.pdf (discussing "a project that used neural networks to predict the probability of death for patients with pneumonia, so that low-risk patients could be treated as outpatients. The results were generally more accurate than those that came from handcrafted models that applied known rules to the data. But the neural network clearly indicated that asthmatic pneumonia patients are at low risk of dying and thus should be treated as outpatients. This contradicts what caregivers know, as well as common sense. It turns out that the finding was caused by the fact that asthmatic patients with pneumonia are immediately put into intensive care units, resulting in excellent survival rates."). See also David Weinberger, *Alien Knowledge*, BACKCHANNEL (Apr. 18, 2017), https://backchannel.com/our-machines-now-have-knowledge-well-never-understand-857a479dcc0e.

indeed, if ML allows us to diagnose and treat more diseases, the demand for those activities could increase.[49]

Conversely, for simplicity we assume that the diagnostic specialty which the AI excels in is one that ordinarily takes place away from the point of care, or if it is at the point of care forms only a part of the care-provider's diagnostic responsibilities. This second assumption allows us to assume that there will still be a physician present at the point of care, e.g. an oncologist who ordinarily would be informed by consulting with a radiologist but instead turns to a ML system.[50] In so doing we can avoid engaging, at least for now, with long-standing medical ethics debates about the appropriateness of fully robotic care.[51]

As set out in the next section, once ML diagnostics are statistically superior to humans, it will only be a short while before the various legal systems, including the US, treat machine-diagnosis as the "standard of care". That designation will mean that any physician or hospital failing to use machine diagnosis without a good excuse will be running a substantial risk of malpractice liability if the patient is incorrectly diagnosed.[52] In a fairly short time, every insurance company and every hospital will require the use of ML, at least as an assistant to physicians, because failure to do so will be actionable in the event of a bad outcome. There are some variables that might alter how quickly this will happen, notably cost and whether courts continue to make distinctions between types of practices and types of practice situations, e.g. teaching hospitals versus rural hospitals versus sole practitioners.

---

[49] Ajay Agawal, Joshua Gans & Avi Goldrarb, *The Simple Economics of Machine Intelligence*, HARV. BUS. REV. (Nov. 17, 2016), https://hbr.org/2016/11/the-simple-economics-of-machine-intelligence.

[50] One very substantial difference between consulting with a human oncologist and "consulting" with a computerized system is that there is no opportunity for any discussion or give and take. An AI gives a report, but can neither explain it nor alter it in light of reasoned argument. This could be real loss to the quality of care, although it is possible that increasing reliance on electronic health records as means of communication between specialists has already eroded those conversations and relationships.

[51] The so-called 'Standard View' of biomedical ethics holds "that the *practice of medicine and nursing* are ineluctably human." KENNETH W. GOODMAN, ETHICS, MEDICINE, AND INFORMATION TECHNOLOGY 26 (2015) (citing R.A. Miller, *Why the Standard View is Standard: People, Not Machines, Understand Patient's' Problems*, 15 J. MED. & PHIL. 581 (1990)).

[52] See Patricia Kuszler, *Telemedicine and Integrated Health Care Delivery: Compounding Malpractice Liability*, 25 AM. J. L. & MED. 297, 316-17 (1999); Kori M. Klustaitis, *Dr Watson Will See You Now: How the Use of IBM's Newest Supercomputer Is Changing The Field of Medical Diagnostics and Potential Implications For Medical Malpractice*, 5 BIOTECHNOLOGY & PHARMACEUTICAL L. REV. 88 (2011-2012).

But these are primarily questions of speed and detail rather than of trend. In fairly short order, it seems highly plausible that ML systems will be prescribed not by doctors but by tort law for certain forms of diagnosis and that medical service providers will comply. And, if a ML system proves statistically superior for treatment, then a similar argument will also apply. In which case, hospitals and other medical service providers will carry out AI-recommended treatment plans unless there is a very clear reason to do otherwise.

B. How Tort Law Incorporates Technical Change

Medical malpractice law is a species of negligence law, which itself is a type of tort, a civil wrong. A physician can commit malpractice by failing to get informed consent (an issue not especially relevant here), or by breaching her duty to provide the appropriate standard of care in a manner that causes injury to the patient. Defining the relevant standard of care is thus a central issue in many malpractice cases.

The standard of care for a medic is, at the most general level, that of a reasonably competent physician. i.e. one who uses a reasonable degree of care and skill. While there can of course be evidentiary issues as to what a physician actually did, in cases that involve whether a physician should have used a particular, relatively new, technology there can also be complicated questions as to whether the use of the new technology–or the failure to use the new technology—is itself negligence. Using new technology also may invite claims that perhaps the people who used it were not yet sufficiently familiar with it and thus used it improperly.

U.S. tort law recognizes that technology changes and thus the general standard of care for professions and trades may change too. Indeed, where once "custom"–what most people in the trade or profession do and have generally done—was the starting point for measuring the appropriate standard of care, US courts today are somewhat suspicious of custom-based arguments on the theory that these arguments provide too little incentive to modernize, and may favor entrenched modes of service provision at the expense of the victim.

This modernizing tendency traces back at least as far as the oft-cited *T.J. Hooper* case, where Judge Learned Hand ruled that it was negligent for a tugboat sailing the Atlantic in 1928 to fail to have a working radio on board in order to hear storm weather warnings. The trial court had found that if the *T.J. Hooper* had carried a radio, it likely would not have foundered. On appeal, Judge Hand first noted that there was no general and established custom of carrying a radio among coastwise carriers, and he admitted that courts sometimes treated the absence of

such a custom as a full defense. But he also noted that a suitable radio was not expensive, and that custom should not be definitive:

> [A] whole calling may have unduly lagged in the adoption of new and available devices. It never may set its own tests, however persuasive be its usages. Courts must in the end say what is required; there are precautions so imperative that even their universal disregard will not excuse their omission. But here there was no custom at all as to receiving sets; some had them, some did not; the most that can be urged is that they had not yet become general. Certainly in such a case we need not pause; when some have thought a device necessary, at least we may say that they were right, and the others too slack.[53]

Since *The T.J. Hooper*, U.S. courts have not been shy about demanding additional precautions even when an industry resisted them[54]—except in the case of medicine, where until recently the courts have been more cautious.

C. Medical Variations: Custom and Localities

To succeed in a medical malpractice case, the plaintiff must show that her injury, more likely than not, resulted from the treating physician's departure from "the generally recognized and accepted practices and procedures that would be followed by the average, competent physician in the defendant's field of medicine under the same or similar circumstances."[55] What constitutes average competence in a given field of medicine is a question of fact, for which parties commonly offer expert testimony.

In contrast, who makes up the set of comparable physicians is primarily an issue of law. For many years physicians, almost alone among professionals and tradespeople, enjoyed two special protections from professional negligence liability, both relating to who counted as comparable: a heightened ability to plead custom as

---

[53] The T.J. Hooper, 60 F.2d 737, 740 (2d Cir. 1932) (L. Hand, J.) (citations omitted).
[54] See, e.g., Bimberg v. Northern Pacific Ry, 14 N.W.2d 410, 413 (Minn. 1944) ("Local usage and general custom, either singly or in combination, will not justify or excuse negligence.").
[55] Hoard v. Roper Hosp., Inc, 694 S.E.2d 1, 4 (S.C. 2010); see also Pike v. Honsinger, 49 N.E. 760 (N.Y. 1898). The basic elements of the tort of negligence are duty, breach, causation and injury.

a defense,[56] and the so-called "locality rule"[57]. The effect of these two rules was to insulate a physician from liability so long as she provided treatment no worse than was common in her community. Since physicians were reluctant to testify against their colleagues until fairly late in the 20th century, these rules worked to greatly limit malpractice claims.

1. The Waning of the Locality Rule

The locality rule reflected a judicial belief that it would be unfair to apply a single standard of care to all physicians. Physicians vary as to their training and specialization, and also in their practice settings. A general practitioner should not be expected to have the same skill as a specialist, at least in matters touching on that specialty.[58] A small rural practice does not have access to the same equipment as a large urban teaching hospital; many courts also seemed influenced by the idea that it would be unfair to expect the prototypical rural practitioner to be as up-to-date as someone affiliated with a major hospital. Precisely what the comparatives were varied slightly: other physicians with similar training in the same or a similar community, or perhaps other physicians with similar training in similar communities in the state.[59]

Today the standard of care for physicians is largely national, reflecting the relative standardization of medical training. Physicians continue to be held to a varying standard depending on their training and type of practice, but the standard applied to members of a given specialty is more or less uniform nationally. The standard of care is that established by the "relevant community," which is now understood to be the national group of practitioners in that specialty. To whatever the extent the locality rule lives on, it applies primarily to general practitioners.

---

[56] Tim Cramm, Arthur J. Hartz & Michael D. Green, *Ascertaining Customary Care in Malpractice Cases: Asking Those Who Know*, 37 WAKE FOREST L. REV. 699, 699-700 (2002) ("Medical malpractice law has long modified the ordinary tort duty of reasonable care. Health care professionals must exercise the same care that other professionals customarily exercise. Thus, the duty applied to medical professionals is a purely factual one, unlike the normative 'reasonable care' standard invoked for non-professionals."). But see Steven Hetcher, *Creating Safe Social Norms in A Dangerous World*, 73 S. CAL. L. REV. 1 (1999) (critiquing reliance on custom as a measure of negligence).

[57] See *infra* § I.C.1.

[58] *See* 99 A.L.R.3d 1133 (1980).

[59] *See* Note, *Call In Houdini: The Time Has Come To Be Released From The Geographic Straitjacket Known As The Locality Rule*, 56 DRAKE L. REV. 753, 754-64 (2008) (tracing origins and evolution of the locality rule), 99 A.L.R.3d 1133 (1980).

2. Custom in Medical Malpractice Meets Technological Change

U.S. courts have, at least until recently, tended to accept evidence of customary practices as persuasive defenses against claims of medical negligence. The rule has been strongly criticized for deterring medical innovation. If the standard of care is defined by custom, then any physician who innovates takes on the risk of deviating from custom. If the innovative practice or device causes harm, that creates an exposure to malpractice liability for "unreasonable" behavior even if, on average, the innovation is beneficial.[60]

In part due to such criticism, and perhaps also due to the erosion of the view that physicians should be above criticism,[61] the privileged position of physicians that allowed them to plead custom in malpractice cases has greatly diminished:

> Gradually, quietly and relentlessly, state courts are withdrawing this legal privilege. Already, a dozen states have expressly rejected deference to medical customs and another nine, although not directly addressing the role of custom, have rephrased their standard of care in terms of the reasonable physician, rather than compliance with medical custom.
>
> Even more important than the raw numbers is the trend revealed by the decisions. The slow but steady judicial abandonment of deference to medical custom began in earnest in the 1970s, continued in the 1980s, and retained its vitality through the 1990s. Showing no signs of exhaustion, this movement could eventually become the majority position.
>
> Furthermore, many of the states that theoretically continue to defer to custom actually apply the custom-based standard of care in a way that operates very much like a reasonable physician standard.[62]

---

[60] *See* Gideon Parchomovsky & Alex Stein, *Torts and Innovation*, 107 MICH. L. REV. 285 (2008).

[61] Public deference to the judgment of medical professionals has gradually declined since World War II. See Philip G., Jr., *The Quiet Demise of Deference to Custom: Malpractice Law at the Millennium*, 57 WASH. & LEE L. REV. 163 (2000).

[62] Peters, *supra* note 61, at 164; *see also* Note, *supra* note 59, at 770-72 (concluding "The movement of nearly all jurisdictions has been to incorporate a national standard of care, and those that have not had the right case arise have continued to loosely apply the similar locality rule.").

In other words, in more and more states, the physician's duty under malpractice is being normalized and brought into alignment with the ordinary tort duty of care, permitting courts to hold that even a widespread medical practice can be negligent[63]--particularly if the innovation they have not adopted is a "precaution[] so imperative that even … universal disregard will not excuse their omission.[64]"

Indeed, as a general matter, the standard of care is not only national, but is subject to reasonably rapid change.[65]

D. Nature of Machine Learning Removes Common Obstacles to the Adoption of New Medical Technology

Much of the writing and thinking about the interaction between medical negligence rules and technical change concerns clinical techniques or devices that are not unambiguously good for the patients to whom the new technology may be applied. Most of these technologies create new risks as well as benefits; frequently they require new training without which physicians may fear they could fail to reap the benefits of the new technology or even misuse it in a harmful way; and

---

[63] Peters, *supra* note 61, at section II.B. (citing cases). Interestingly, studies show that as states switch from a custom-based measure of the standard of care to a national standard based on reasonableness, the rate of adoption of innovations converged to the national mean. This suggests that "this change in behavior was motivated by the change in tort law's test of reasonable care, not by any independent medical evaluation of whether compliance with the local or national custom was in the best interests of the patient." Mark Geistfeld, *Does Tort Law Stifle Innovative Medical Treatments?*, JOTWELL (June 2, 2015) (reviewing Anna B. Laakmann, *When Should Physicians Be Liable for Innovation?*, 36 CARDOZO L. REV. 913 (2015)), http://torts.jotwell.com/does-tort-law-stifle-innovative-medical-treatments/.

[64] The T.J. Hooper, *supra* note 53.

[65] Patricia Kuszler, *Telemedicine and Integrated Health Care Delivery: Compounding Malpractice Liability*, 25 AM. J. L. & MED. 297, 316-17 (1999). On the physician's duty to keep informed of new treatment methods, see Jolene S. Fernandes, *Perfecting Pregnancy via Preimplantation Genetic Screening: The Quest for an Elusive Standard of Care*, 4 U.C IRVINE L. REV. 1308-12 (2014); Alan Weintraub, *Physician's Duty to Stay Abreast of Current Medical Developments*, 31 MED. TRIAL TECH. Q. 329 (1985); Carter L. Williams, Note, *Evidence-Based Medicine in the Law Beyond Clinical Practice Guidelines: What Effect Will EBM Have on the Standard of Care?*, 61 Wash. & Lee L. Rev. 479, 508–12 (2004). Consider also Harbeson v. Parke-Davis, Inc., 656 P.2d 483 (Wash. 1983) (holding that physician's failure to conduct literature search on side effects of Dilantin justified liability for wrongful birth).

frequently there is concern that not all the long-term risks of the new techniques or devices will necessarily be evident at the time that the physician must decide whether to use the familiar procedure or the new one.[66] Each of these properties creates the specter of tort liability if something goes wrong, creating disincentives that may balance out or even overcome the purported advantages: A bad outcome following a new surgical procedure creates the risk that the patient may claim improper training. A new implantable device creates risks of unforeseen long-term complications or even failure. A new invasive diagnostic procedure may have side-effects. Some advanced diagnostic equipment may be too expensive to have in every hospital, much less in every physician's office.

Machine Learning systems are different from these common examples in many important respects. From the point of view of malpractice risk management, AI diagnostics should be much easier to implement than other recent medical advances that have required expensive equipment be present on-site. ML can be trained to work with any diagnostic materials that can be reduced to standardized data, including notably radiographic images; as the ML is fundamentally a computer program, the analysis need not be done on-site but can instead live anywhere else or even in the cloud.[67] Any medical facility capable of capturing clinical information, digitizing it, and transmitting it, could presumably, if affordable, access a machine-learning-based computer located anywhere else.

In short, the data collection needs to be done at the point of care, where the patient is; the data input and the processing can be anywhere. Rather than being equipment or a technique, machine learning systems present as a service. Unless the pricing is extortionate, this will not only increase the rate at which medical service providers adopt ML systems, but also will increase the speed with which hospitals and even local physicians feel legal pressure to use ML.[68]

---

[66] See Michael D. Greenberg, *Medical Malpractice and New Devices: Defining an Elusive Standard of Care*, 19 HEALTH MATRIX 423 (2009).

[67] Any remote location and especially cloud-based services raise issues of security and privacy outside the scope of this article. *See*, *e.g.*, Sebastian Zimmeck, *The Information Privacy Law of Web Applications and Cloud Computing*, 29 SANTA CLARA COMPUTER & HIGH TECH. L.J. 451 (2013); William Jeremy Robison, Note, *Free At What Cost?: Cloud Computing Privacy Under The Stored Communications Act*, 98 GEO. L.J. 1195 (2010).

[68] Watson-as-a service also raises some complex issues of what standards of liability would apply to Watson's errors, see Jessica S. Allain, Comment, *From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems*, 73 LA. L. REV. 1049 (2013). It also raises potentially difficult problems of proof, as one would need a perfect snapshot of the entire medical data base on which the ML could have relied at the moment of treatment in order to prove that had the ML been consulted it would have
(Continued)

There is one way, however, in which ML may not be different from other medical innovations: it will not be immune to all malpractice claims. Even if we can prove that an ML, on average, is a better diagnostician than the average physician, a patient misdiagnosed by an ML might seek to claim that even if the ML's overall average is better than most or all humans, a significant part of the ML's success occur in cases where humans would have failed, and that a significant part of the ML's errors fall on a group of patients who might have fared better with a human doctor,[69] and that the misdiagnosed patient fell into the group who would have fared better with an average—or a particular—human physician. Simply put, humans and ML systems might make very different kinds of mistakes. And these differences might affect the manner in which liability is assessed.

Currently we tend to train ML systems from databases that reflect the best judgments of panels of practicing physicians. One could in theory train on actual real-world outcomes, if the medical system commonly annotated diagnostic data files with outcome data at regular intervals. At present, however, it is not common to find, say, a database containing radiological images linked with data about whether and which tumors actually manifested in the patients over a set period of time. Given the hypothesis on which this article is based, that a ML system has managed to do substantially better on average than do human physicians, we would not expect in the short term[70] that the ML system's errors would tend to be in cases that humans would, on average, have diagnosed correctly. Nevertheless, since that tendency is only a matter of *probability*, the *possibility* cannot be excluded as a provable or mathematical certainty in general or indeed in any given case. Worse, as set on in Part I. A.1, the current state of the art for neural networks, with its lack of interpretability, creates some circumstances in which there is no practical way for humans to examine the reasoning for any given decision.[71] Furthermore the lack of causal connections, of the sort humans typically use to understand reasoning, makes it difficult to pinpoint a specific source of error in the ML-based prediction system. Any given diagnosis is the result of correlations based on the entire medical data base available at the moment of diagnosis. As a result, given current technology,[72] a physician or hospital relying on a neural network cannot back up any particular decision with evidence of a reasoned decision-making process beyond pointing to the program's overall batting average and perhaps (if the system is

---

made a better decision than the human. Unfortunately, these issues are beyond the scope of this project.

[69] See Millar and Kerr, *supra* note 9.

[70] We return to the issue of relative long-term accuracy in Part III.

[71] *See supra* text at note 40.

[72] For a discussion of ongoing efforts to provide explanation see *infra* text at note 154.

programed to provide it) to an 'evidence profile' that shows how it weighed different classes of information[73] or perhaps some number indicating the neural network's degree of confidence in its diagnosis.[74] Thus, for example, if a hospital is relying on an ML for its diagnosis, it will be open to both parties to provide *ex post* rationalizations based on expert testimony by humans, but while the defendants relying on the ML will have a chance to argue that the ML made the right call on the merits, they may have the disadvantage of not being able to explain how the actual decision came to be.

A neural network can learn from its successes and its mistakes—that is the key to how it is trained initially. So long as its decisions are being reviewed by human physicians on an ongoing basis[75] we would hope that its success rate continues to improve as its training data incorporates new information based on their input.[76] Likewise, such systems will improve as the quality and quantity of data increases. Furthermore, we would expect that, prior to the adoption of ML diagnosticians, researchers would have studied its outcomes carefully in order to see if any patterns of error emerge. Perhaps doctors using ML diagnoses could be warned not to rely on it for any identifiable sub-classes of cases where humans were still superior. It is worth noting, however, that the search for such patterns of error likely would require a careful review process external to the ML system because the ML itself is unlikely to be able to make these distinctions unless the subclasses to consider can be defined for it in advance. Worse, while doctors should be able to identify some false positives (Type I errors) fairly quickly – e.g. if they operate but

---

[73] For an example of this in the Jeopardy game show context see David Ferrucci et al, *Building Watson: An Overview of the DeepQA Project*, AI MAGAZINE (Fall, 2010), reprinted at Association for the Advancement of Artificial Intelligence, *The AI Behind Watson - The Technical Article*, http://www.aaai.org/Magazine/Watson/watson.php, in which weights are given to "location", "passage support", "popularity", "source reliability", and "taxonomic" categories for the answer to the question "Chile shares its longest land border with this country."

[74] *See* G. Papadopoulos, P.J. Edwards, A.F. Murray, *Confidence Estimation Methods for Neural Networks: A Practical Comparison*, 12 IEEE TRANSACTIONS ON NEURAL NETWORKS 1278 (2001). .

[75] Most commonly this would happen in batch mode, not real time: scientists train models first and deploy them into the wild in a static form. They might the release updated versions later that take into account new data, Working in batch mode allows for testing between releases and makes it easier to avoid error that can occur if the neural network is learning in real time. For an example of the dangers of continual real-time learning, see James Vincent, *Twitter Taught Microsoft's AI Chatbot to be a Racist Asshole in Less than a Day*, THE Verge (Mar 24, 2016), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

[76] But see *supra* text above note 37.

find no tumor[77] –false negatives (Type II errors) may take longer to manifest, and may pose real risk to patients if they are misdiagnosed as a result of reliance on the ML system. Ideally, rigorous external review would keep the number of meritorious malpractice claims based on a robust ML system's diagnoses low, and should keep the number of successful claims low as well, but the technical obstacles to achieving this ideal may be substantial.

E. Malpractice Law Will Require Machine Learning Systems When They Are Demonstrably Better

It is important to recall two basic rules of malpractice law: bad outcomes do not necessarily mean there was malpractice, and physicians are not expected to be perfect. Sadly, there are some cases that could not be cured with even the best medical care in the world. A physician (or hospital, or insurer) relying on a machine learning system will be held to no different a standard than if it relied on a human; indeed, from a legal point of view, the decision to rely on ML will be a human medical judgement like any other. As noted above, the law requires only that physicians exhibit the ordinary skill and judgement of a reasonably competent similarly situated physician. Thus a physician, hospital or insurer relying on an ML diagnosis will, at least initially, be held to no higher standard than that of the ordinary human. Once ML itself becomes the standard of care, ML will raise the bar. But even though a higher level of accuracy will now be the standard, the malpractice exposure of ML-users will actually shrink as henceforth by relying on ML they will be complying with the professional standard;[78] at that point reliance on human diagnosticians will become the risky legal strategy both for failing to use the an increasingly common technology of which they should have been aware and because (by hypothesis) the risk of error is in fact greater.

In states that have changed the standard of care to align with general tort principles, one would expect the legal pressure to adopt ML to be very strong once

---

[77] It should be noted, however, that some oncological treatment regimens do not involve initial surgery, for example those relying instead on chemotherapy. Error may be harder to detect in such cases since the absence of a subsequent cancer might falsely be attributed to successful treatment.

[78] One small caveat ought to be noted here: Were an ML system to provide a clearly ludicrous diagnosis, one that any reasonable physician ought to have noticed was wrong or inapposite, then--even after it becomes the standard of care--relying on ML in those circumstances could easily be characterized as negligence, and plausibly as gross negligence. This entails a need for continued comprehensive human training, even if the role of human physicians, like pilots, becomes secondary to the role played by machines.

the evidence was clear that an ML system was better than human physicians, for it would be unreasonable to fail to adopt ML unless the cost was very high, an issue we address below in Part II. In the decreasing number of states that still allow custom to act as a defense, medical malpractice law's definition of the standard of care can act as a brake on innovation. In those states, the legal push to use ML will not be as great until ML is in common use nationally in the relevant specialty; at that point ML usage itself becomes 'customary' and we would expect the law to provide a strong push towards compliance with the relevant general norm for any late adopters.[79]

From the point of view of the tort law theorist, at least of the law-and-economics persuasion, this is a happy story as tort law seems poised to do exactly what theorists would want it to do: tort law incentivizes a profession to adopt a new technology that likely will save lives. Indeed, even if tort law were neutral or a possible brake, as in the case of custom-dependent states before the national trend develops, once ML's success rate is demonstrably superior to human physicians we would expect that both medical ethics and cost considerations would drive medical care providers to choose to consult an ML system, and to rely on its judgments unless they could articulate good reasons not to. Thus, if the ML's track record is significantly better than most humans', then arguably ethics would counsel (most[80]) humans to rely on the ML even if they believed they had a superior diagnosis.[81] In time, perhaps even in a short time, a provably superior ML becomes the standard of care for diagnosis in a specialty in many jurisdiction, and certainly throughout the United States.

We turn now to the economic drivers towards ML—and to some speculation about ML's economic consequences. Our happiness may prove temporary.

---

[79] There is one persistent exception to this trend, the so-called "two schools of thought" doctrine. Under this doctrine doctors have a powerful defense against malpractice if they can show that the treatment they provided is supported by a minority of professionals in the field due to disagreement in the field as to which is the optimal treatment. See generally Douglas Brown, *Panacea or Pandora' Box: The Two Schools of Medical Thought Doctrine after* Jones v. Chidester, 44 J. URBAN & CONTEMP. LAW 223 (1993). Note that this defense would not generally apply if the minority consisted of doctors unwilling to modernize in the face of a demonstrably better new technique or technology, being limited to situations where evidence as to which 'school' is better is inconclusive.

[80] Presumably Dr. House would demur.

[81] See Millar & Kerr, *supra* note 9. The argument in the text presupposes that the human physician at least accepts that Watson's diagnosis is plausible. If the human physician believes Watson's diagnosis is erroneous, then she will have a duty to step in. See *supra* note 78. See also *infra* text at note 93 (discussing how errors can happen).

II. Machine Learning and the Demand for Specialist Physicians

A. ML and the Market for Diagnostic Physicians

Physicians are expensive to train, and expensive to keep on staff.[82] Given the necessity of acquiring training data, formatting it, and establishing compatible data exchange regimes with hospitals and other medical care providers,[83] we presume that ML diagnostics will follow the path of many other digital technologies and exhibit high fixed costs but relatively low marginal costs. The fixed costs will be the presumably high cost of first priming the system with training data, then arranging for compatible data input from the treating physician's office. The costs of processing individual requests we presume to be low by comparison, although this is at best only informed speculation on our part. Magnetic resonance imaging (MRI) may, however, be instructive: Early MRI machines cost around $2 million plus $1million for installation.[84] Modern state-of-the-art devices can cost up to $3 million.[85] Yet failure to use one would in many cases be malpractice. As the high capital cost of an MRI machine can be shared by the many patients who will use it during the machine's lifetime, the per-patient cost is low enough to make an MRI the standard of care, and therefore the standard diagnostic tool, for many different diseases and sets of symptoms.[86]

At present, the smart bet seems to be that ML systems will not be as expensive as a human physician. "Once a model has been 'trained,' it can be deployed on a relatively modest budget."[87] In any plausible cost scenario, however, the medical services provider's financial problem is that unless ML replaces all or part of some other cost—the human doctor being the natural target—ML is just one

---

[82] Doctors' salaries vary by specialty in the US, starting at $189,000 for pediatricians and going up to $376,000 or more for cardiologists and top surgeons. *Average Salary for People with Jobs as Physicians / Doctors*, Payscale.com (Feb. 11, 2017). Doctors also impose substantial overheads, plus require offices and support staff.

[83] For more on the importance of acquiring training data, see *infra* text at notes 139-150.

[84] Ben L. Holmes, *Current Strategies for the Development of Medical Devices* 219, 220 in INSTITUTE OF MEDICINE STAFF, TECHNOLOGY AND HEALTH CARE IN AN ERA OF LIMITS (1992).

[85] Lacie Glover, *Why Your MRI or CT Scan Costs An Arm and a Leg*, THE FISCAL TIMES (July 21, 2014), http://www.thefiscaltimes.com/Articles/2014/07/21/Why-Your-MRI-or-CT-Scan-Costs-Arm-and-Leg.

[86] Klustaitis, *supra* note 52, at § III.2.

[87] Andrew Bean & Isaac S. Kohane, Editorial, *Translating Artificial Intelligence Into Clinical Care*, 316 J. AM. MED. ASSOC. 2368, 2369 (2016).

more cost, whether small, medium, or large. And as is well known, the medical sector is under pressure to cut costs.

Whatever the pricing scenario, the more that a ML system becomes the diagnostician of choice, the less there should be demand for similar human diagnosticians.[88] Instead, all that will be necessary is for someone to collect the patient's data and feed it to the system. (Recall our second simplifying assumption above, that Prediction-ML is replacing a consulting specialist not the point-of-care physician.[89] The legal issues created by purely automated medicine of the Treatment-ML variety are both more remote in time and more complex than those discussed here.[90]) If it becomes the case that all that ML requires is the input of data, in many cases those data could be collected by less-trained technicians, just as today nurses or trained medical technicians, not physicians, take blood samples, EKGs, MRIs and CT-scans. Or, in time, other specially trained AIs may do the intake interview as well.[91]

At first, medical service providers and insurers will treat ML diagnosis as another tool that is available to physicians. Thus, at first, hospitals will feel required to keep the same number of physicians around in order to double-check what the ML does. This will be costly since the hospitals and insurers will have to pay both the physicians and whoever provides the diagnostic service. In addition, as big-data-based diagnosis takes off, it seems likely that hospitals will be expected to collect increasing amounts of data to supply the AI with the information it needs to continue to learn in order to improve its diagnoses. Thus, hospitals will find themselves paying for more recording equipment, for more nurses and technicians to apply the recording equipment, for the same number of physicians, and for the AI. Again, in the short run, bills go up.

Once, however, confidence in the AI increases, insurers will inevitably seek cost savings by decreasing the use of physicians to do diagnosis. These savings are

---

[88] For a general argument that "the number of workers-intellectual as well as manual-is reduced by quantum measures in computer-mediated labor" see Aronowitza and DiFazioa, *supra* note 6.

[89] *See supra* text at note 42.

[90] For a taste of the issues see Ryan Abbott, *The Reasonable Computer: Disrupting the Paradigm of Tort Liability* (November 29, 2016), https://ssrn.com/abstract=2877380.

[91] For an account of early attempt to train an AI to do patient interviews in China see *Baidu Announces Melody, a New AI-Powered Conversational Bot for Doctors and Patients NASDAQ.com* (OCT. 11, 2016), http://www.nasdaq.com/press-release/baidu-announces-melody-a-new-aipowered-conversational-bot-for-doctors-and-patients-20161011-00083 (last visited Oct 18, 2016)

likely to be small in comparison to what might be achieved from having machines do treatment as well as diagnosis, but one could see these small savings as the vanguard of a possible future in which the push to replace doctors with machines is more widespread. The real action occurs once ML capably encroaches on areas of medical treatment—including not only the development of treatment plans but also their delivery.

Initially, rather than remove humans entirely from the diagnostic loop, hospitals and insurers likely will seek to have a physician 'review' ML diagnoses. Since the cost saving is predicated on reducing the number of physicians, the inevitable result of this 'human in the loop' policy is that each remaining physician will be tasked with reviewing a larger number of cases per day than they previously handled. At some point, perhaps quite soon, the load on the physicians will rise to the point where one might question their ability to do more than a basic reality check.[92] Even that check undoubtedly will have some value, because at present MLs can become confused—such as when the Jeopardy-playing Watson suggested Toronto is a US City.[93]

However, we question how often a physician presented with a large volume of cases would be able to detect relatively subtle errors. As the load increases, the carefulness of the review must inevitably decrease; meanwhile, it seems probable that the human's malpractice liability would remain the same, making the physician a moral and possibly financial 'crumple zone'.[94] Ultimately either the

---

[92] Cf. Juan Mateos-Garcia, *To Err Is Algorithm: Algorithmic Fallibility And Economic Organisation*, NESTA, May 10, 2017, http://www.nesta.org.uk/blog/err-algorithm-algorithmic-fallibility-and-economic-organisation (last visited Jun 1, 2017), which argues that "supervisors need to check each decision individually. This means that as the number of decisions increases, most of the organisation's labour bill will be spent on supervision, with potentially spiralling costs as the supervision process gets bigger and more complicated. …. When considered together, the decline in algorithmic accuracy and the increase in labour costs … are likely to limit the number of algorithmic decisions an organisation can make economically."

[93] Despite surface appearances, it is not. For an explanation of the error see Steve Hamm, *Watson on Jeopardy! Day Two: The Confusion over an Airport Clue*, Smarter Planet Blog (Feb 15, 2011), http://asmarterplanet.com/blog/2011/02/watson-on-jeopardy-day-two-the-confusion-over-an-airport-clue.html.

[94] See Madeline Elish, *Moral Crumple Zones: Cautionary Tales in Human Robot Interaction* (UPDATED: final draft), We Robot 2016, http://robots.law.miami.edu/2016/wp-content/uploads/2015/07/ELISH_WEROBOT_cautionary-tales_03212016.pdf. Machine Learning can be used to choose which cases are most uncertain, and present those only to reduce the volume. But there remains the risk that the ML system gets it wrong, i.e. misses

(Continued)

physicians will rebel, or the cost of their insurance will wipe out at least a chunk of the savings, or MLs will become so reliable that insurance companies and hospitals force physician out of the loop. In this scenario, bills go down unless ML providers react to the removal of the human doctors by charging even higher monopoly prices—something that presumably would be prohibited by the Sherman Act.[95]

Indeed, the removal of humans from the practice of radiology has already begun. Krista Jones wrote of her son's decision to become a radiology technician,

> After seeing what this radiation treatment was able to do for me, my son applied to a university program in radiology technology to explore a career path in medical radiation. He met countless radiology technicians throughout my years of treatment and was excited to start his training off in a specialized program. However, during his application process, the program was cancelled: He was told it was because there were no longer enough jobs in the radiology industry to warrant the program's continuation.[96]

Whatever the current demand for radiologists, future doctors and even radiology technicians are being exposed to strong signals that radiology is a field with no future. "They should stop training radiologists now," asserts University of Toronto Professor Geoffrey Hinton.[97]

That said, the future in which a patient in the US consults an AI directly without seeing even a primary care physician seems highly implausible if not far, far away—not only from a scientific point of view but from a legal perspective as well. In 2015 the Federal Trade Commission settled claims against marketers of "MelApp" and "Mole Detective" for "deceptively claiming their mobile apps could

---

some important cases that need to be reviewed, and we are back to the problem of humans having too many cases to review.

[95] See Sherman Antitrust Act (Sherman Act), 26 Stat. 209, codified at 15 U.S.C. §§ 1–7 (making it a felony to "monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations").

[96] Krista Jones, *I Was Worried About Artificial Intelligence-Until It Saved My Life*, QUARTZ (Aug. 20, 2017), https://qz.com/1056817/i-was-worried-about-artificial-intelligence-until-it-saved-my-life/ (last visited Sep 19, 2017).

[97] Quoted in Mukherjee, *supra* note 21.

detect symptoms of melanoma, even in its early stages."[98] Direct-to-patient services of this type would face legal and regulatory obstacles of their own, not least unauthorized practice of medicine claims in many states.[99]

B. Machine Learning and the Deskilling Debate

Medical observers have repeatedly warned that new technology causes the loss of old skills.[100] It remains unclear whether ML causes the loss of diagnostic skills or whether we should better "hypothesize that the use of [ML], especially their ability to identify and rank differential diagnoses, might actually improve diagnostic acumen."[101] We may never know; if ML actually eliminates all or most of the demand for the diagnostic services of a physicians in a given specialty, there will be some kind of loss of human know-how, however one characterizes it. The reduction in demand for physicians in a specialty inevitably will have knock-on effects in medical schools, as students, and especially interns and residents, steer away from the subject. Soon, hiring committees decide to use scarce resources elsewhere. The knowledge is not 'lost'–it lives on in the few remaining specialists and researchers and in a database—but it is no longer being added to in the same manner because humans contribute few if any new diagnoses paired with outcomes to the ML system's database. Instead new data about outcomes come primarily from situations where ML itself provided the diagnosis. One can only speculate about the extent to which the future of human medical knowledge will be compromised after a generation or two of diagnostic or treatment decisions generated exclusively by machines.

ML may also have other deskilling effects beyond the elimination of a specialty. We will still need physicians to act upon ML's conclusions, to do the surgery (at least until we have good robot surgeons, which seems to involve a much more complex set of challenges[102]). On the other hand, we may not need physicians

---

[98] FEDERAL TRADE COMMISSION, FTC CRACKS DOWN ON MARKETERS OF "MELANOMA DETECTION" APPS, https://www.ftc.gov/news-events/press-releases/2015/02/ftc-cracks-down-marketers-melanoma-detection-apps (last visited Sep 30, 2017)

[99] They might, however, have promise for countries with less-developed economies or large, dispersed, rural populations. *See*, *e.g.*, *Your Face could Reveal if You Have a Rare Disease,* WIRED UK, http://www.wired.co.uk/article/fdna-rare-disease-facial-recognition-algorithms (last visited Jun 11, 2017) (describing use of phones to detect rare diseases). .

[100] See GOODMAN, *supra* note 51, at 56 (noting that "[e]very generation enjoys the services of at least a few pessimists who despair of the current state of affairs" in medicine).

[101] GOODMAN, *supra* note 51, at 58.

[102] *See*, *e.g.,* Ian Kerr, Jason Millar and Noel Corriveau, *Robots and Artificial Intelligence in Healthcare*, in CANADIAN HEALTH LAW AND POLICY, 5TH ED. 257 (2017).

to interview the patient. An ML system could do the job, or perhaps – initially – a nurse practitioner (or even a nurse) guided by questionnaires, updated on the fly, provided by an expert system; tomorrow the questionnaire may be informed by a full AI interacting with information from real-time sensors.[103] The more that AI medicine provides occasions for substituting less expensive personnel for physicians and other highly paid medical service providers,[104] the more we can expect simple economic pressure to push towards the same ends we ascribed to malpractice liability above. A further push likely will come from the need to force the data collected to be as standardized as possible, in order to become quality fodder for future AI training and testing.

Anticipating some version of this future, an opinion column in the *Journal of the American Medical Association* recently suggested that in order to maintain their relevance perhaps radiologists and pathologists should rebrand themselves as Information Specialists "whose responsibility will not be so much to extract information from images and histology but to manage the information extracted by artificial intelligence in the clinical context of the patient." Even so, the article suggested there would be enormous economies of scale, allowing the specialists to export their work: "A single information specialist, with the help of artificial intelligence, could potentially manage screening for an entire town in Africa."[105] Indeed, this more or less is the business model of startup Alexapath.[106]

Extrapolating the future of AI-based diagnostic medicine is not easy. Current trials offer hope that ML systems will find cures for new diseases without human help, particularly at the molecular level.[107] In a world of partial successes,

---

[103] For a discussion of the difference between sensor data and electronic health record data, and the greater utility and ease of analysis of the sensor data, see Iyad Batal, *Temporal Data Mining for Healthcare Data*, in HEALTHCARE DATA ANALYTICS 379, 380 (Chandan K. Reddy & Charu C. Aggarwal, eds 2015). Many U.S. states have regulated limits on the role of so-called physician extenders which might block this scenario. *See* Amanda Swanson and Fazal Khan, *The Legal Challenge of Incorporating Artificial Intelligence into Medical Practice*, J. HEALTH & LIFE SCI. L. 90, 116 (October 2012).

[104] We can dream about replacing hospital administrators, but they likely will be the last to go.

[105] Saurabh Jha and Eric J. Topol, Viewpoint, *Adapting to Artificial Intelligence*, 316 J. AM. MED. ASSOC. 2353 (2016).

[106] *See* Jessica Leber, KILL TIME IN TRAFFIC BY DIAGNOSING CANCER NEO.LIFE (Sept. 28, 2017), https://medium.com/neodotlife/lou-auguste-and-alexapath-46f7b5f724ca (last visited Sep 30, 2017).

[107] For a suggestive example of AI being used to find a drug to cure a new disease see Jordana Divon, *Toronto startup has a faster way to discover effective medicines*, THE GLOBE AND MAIL, July 27, 2015, https://www.theglobeandmail.com/report-on-business/small-

(Continued)

we would expect ML to be able to identify which treatments work best.[108] Researchers are also working on using ML to customize treatments for patients based on their genetics or on the similarity of their symptoms to earlier success stories.[109]

III. Dangers of Over-Reliance on Machine Learning in Medicine

Our third Part is the most speculative in part because it imagines events farthest in the future. Machine learning works by using as inputs what is in effect big data of medicine: symptoms, test results, diagnosis, and outcomes from a substantial number of patients. In the case of ML and radiology the 'outcomes' are the opinions of a panel of physicians.[110] In other cases, and perhaps for future iterations of ML too, the inputs might be based on real-life outcomes. In either case, the training process is path dependent, and the quality of answers depends on how the system is trained.[111] Inevitably, the quality of an AI's outputs is subject to the quality of the data – GIGO (garbage in, garbage out) remains as true as ever.[112] Nonetheless, as we have seen in Part II, there may come a point where the reliability of the AI is so high that the human physician seems unnecessary or even – to the extent she may overrule valid diagnoses – unhelpful in that her inputs tend to reduce the probability of a successful outcome.

---

business/startups/toronto-startup-has-a-faster-way-to-discover-effective-medicines/article25660419/?arc404=true (last visited Sep 30, 2017) (describing use of AI to find potential treatment for Ebola).

[108] This includes noting correlations that have escaped humans. See Andrew H. Beck et al., *Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival*, 3 SCI. TRANSLATIONAL MED. 108 (Nov. 9, 2011) (describing use of "C-Path (Computational Pathologist)" to identify stromal morphologic structures, a "previously unrecognized prognostic determinant for breast cancer".); *see also* David L. Rimm, *C-Path: A Watson-Like Visit to the Pathology Lab*, 3 SCI. TRANSLATIONAL MED. 1 (Nov. 9, 2011) (noting importance and limits of study).

[109] Not that this prospect is itself without unique legal issues. See W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419 (2015) for a survey.

[110] *See supra* text at notes 35-36.

[111] Syed Shariyar Murtaza et al, *How to Effectively Train IBM Watson: Classroom Experience*, 49th Hawaii Int'l Conf. on System Sciences (2016), https://www.computer.org/csdl/proceedings/hicss/2016/5670/00/5670b663.pdf.

[112] See, e.g., James Vincent, *Twitter Taught Microsoft's AI Chatbot To Be A Racist Asshole In Less Than A Day*, THE VERGE (Mar. 24, 2016), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

But what happens once we take the human physicians out of the equation? Now the outcome data being input to the ML system are no longer produced by human decisions or AI plus human decisions, but only from outcomes based on ML-generated diagnoses.

This could happen in either of two ways, depending on whether we rely on ML solely for diagnosis, or use it also for identifying the course of treatment dictated by the diagnosis. First, and earlier in time, assume the ML takes over the diagnosis function from people but human doctors continue to choose the appropriate treatment. When the ML needs new training data, for example as new and improved sensors or imaging equipment come on line, if humans with the necessary diagnostic training are no longer available because they have been displaced by machines we face a problem. Rather than creating new training data by consulting expert physicians, we will need to create the new data by some other means. Relying on an ML trained on old training data has problems.[113] In this scenario there is a danger that the diagnostic decisions in a closed-universe of ML systems might take a wrong path, one not as good as the one that would have been taken if human physicians continued to provide training data. On the other hand, trying for an evidence-based approach in which we examine treatment outcomes based on human treatment decisions and then associate those outcomes with the diagnostic materials introduces substantial problems of its own. One is that it is a lot of work. Another is that it can take a long time, since all of the 'outcomes' we are interested in may take years to manifest.

The situation looks even more concerning if ML systems also take on the job of choosing and applying the course of treatment. Now, we face a closed-loop system, one in which the outcomes themselves owe their origins to ML generated choices. In such a scenario, the very distribution of observed cases and outcomes is a result of the ML system's decision strategy. If the ML system does not consider the right optimization function, things may derail. When a clinician is in the decision loop, she has the ability to adjust the optimization criteria (e.g. balance symptom reduction with side-effects), and incorporate additional variables into that criteria (e.g. multiple types of side-effects) to refine the decision strategy. An ML-system optimizes a fixed performance criteria, but it does not have the same normative ability to self-correct and gradually incorporate new dimensions to its value system.

Before going into further, it may be useful to emphasize the relative modesty of our claim. We are not claiming that closed-loop retraining *must* result in the degradation of an AI's predictive abilities. And we are certainly not echoing Juan

[113] *See supra* text at notes 37-39.

Mateos-Garcia's claim that "'entropic forces' that degrade algorithm accuracy will win out in the end: no matter how much more data you collect, it is just impossible to make perfect predictions about a complex, dynamic reality"[114] – not least because this claim is addressed primarily to systems where humans have an incentive to game against the AI, a condition that we trust does not apply to diagnostic medicine. Rather, our concern is whether in the closed-loop scenario we can be confident that over time the AI's diagnoses will remain of the high quality that originally led the medical and legal systems to prefer the AI to human diagnosticians. And even if we have some confidence that degradation is unlikely, as we explain below, there is the larger risk that improvement will not continue; indeed, especially if we rely on ML to plan and deliver treatments upon diagnosis, there is some real risk of the ML system reinforcing its original decisions when some other path might be better. If, as we believe, both law and medical ethics should require that we have this confidence before we rely solely on AI diagnosticians, then we may have a problem.

Statistical systems require feedback.[115] "The ideal technique for testing the obtained model is to use an external validation dataset that is collected independently of the training dataset on which the model was built."[116] Indeed, this testing and improvement is a continual process. Ideally one would check and retrain the AI on new data, making for a workflow of collect data, train a model, get new data, retrain, repeat. Retraining does not necessarily require a human in the loop. But for more complex real-life problems retraining may require human input, to check data quality and to generate labels for the new data.[117] And here is where the problem lies: If the AI always recommends a particular drug regime for a given type of cancer, we will never get any new data on the efficacy of radiation. As a result, we will never learn whether radiation could end up being better in some circumstances. In essence, the AI's initial diagnosis decisions will decide the training examples available to downstream. Of course, similar problems bedevil cancer treatments run by humans: ethics and humanity prevent the use of control groups of patients with deadly diseases.

---

[114] Juan Mateos-Garcia, *To Err Is Algorithm: Algorithmic Fallibility And Economic Organisation*, NESTA (May 10, 2017), https://www.nesta.org.uk/blog/err-algorithm-algorithmic-fallibility-and-economic-organisation.

[115] CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION 6 (2016).

[116] Sanjoy Dey et al., *Predictive Models for Integrating Clinical and Genomic Data*, in Reddy &. Aggarwal, *supra* note 103, at 433, 450.

[117] Martin Zinkevich, *Rules of Machine Learning: Best Practices for ML Engineering*, http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdf.

How much humans need to be involved in ML retraining varies with the type of problem being solved. Physical processes that can be observed and measured objectively like object grasping, or motor learning in robotics, lend themselves to automated retraining,[118] essentially via trial and error. We do not, however, wish to subject patients to random error as an ML system learns by doing. Automated retraining works best for problems where the preferred objective can be described precisely (mathematically), such as winning or losing in the game of Go.[119] Indeed, DeepMind's latest Go-playing AI, AlphaGo Zero, learned using no external training data at all: "With each iteration of self-play, the system learns to become a stronger player."[120] "It can do this efficiently because all the other uncertainties are known. … There is complete information. … There is a way to measure success. In short, the behavior of the game of Go is predictable, real world systems however are not."[121] In contrast to playing Go, retraining on diagnostic technique will require human input and supervision until such a time as we can sufficiently describe the conditions we are testing for in advance.[122]

One might reasonably ask why, once the AI is up and running and routinely outperforming human doctors, it cannot simply learn from its mistakes. One part of the answer is that in the case of tumor detection at least, we may only learn of its mistakes several years after the fact. Even assuming that medical systems are engineered to gather the feedback years later, that still leaves the possibility of an AI running on the wrong path for some significant period of time. Indeed, AI

---

[118] Retraining with no humans in the loop is sometimes called self-supervised learning. See, e.g. Dave Gershgorn, *Google's Robots Are Learning How To Pick Things Up*, POPULAR SCI (Mar. 8, 2016), http://www.popsci.com/googles-robots-are-learning-hand-eye-coordination-with-artificial-intelligence.

[119] Google's AlphaGo Zero is the perfect example of a system that can train itself in a closed loop. "The network learns by comparing itself not from external training data but from synthetic data that is generated from a previous version of the neural network." Carlos E. Perez, *Why AlphaGo Zero is a Quantum Leap Forward in Deep Learing*, MEDIUM (Oct. 22, 2017), https://medium.com/intuitionmachine/the-strange-loop-in-alphago-zeros-self-play-6e3274fcdd9f. AlphaGo Zero can do this, however, only because the rules of Go can be described mathematically. *Id*.

[120] Perez, *supra* note 119.

[121] Perez, *supra* note 119.

[122] For a description of the technique of "sparse representations" -- in which an AI is trained with general criteria that require less and more general training data, then left to train itself, then "fine-tuned" by humans (which includes checking to see if the results make any sense at all) – see Dinggang Shen et al., *Deep Learning in Medical Image Analysis*, 19 ANNUAL REVIEW OF BIOMEDICAL ENGINEERING 221–248 (2017), http://www.annualreviews.org/doi/10.1146/annurev-bioeng-071516-044442 (last visited Oct 11, 2017).

applications with long delays between prediction and real-world validation are among those at the greatest risk of 'concept drift', a known source of error.[123] Another risk is that learning from new training data can overwrite the learning from older data, which may not lead to an improvement in performance,[124] although this danger ought to be able to be mitigated by careful validation against the original training data.

Worse, in some cases, especially if the initial training data has systematic errors, that automated feedback, and even human-assisted feedback, can amplify the errors rather than correct them. Thus, for example, if a crime database is biased because officers have tended to stop minorities or to patrol disproportionately in minority neighborhoods, a predictive system based on that data will continue to steer police in those directions, and the arrests they make will be seen as confirmation of the initial bias.[125]

For these and other reasons, some computer experts, such as Cathy O'Neill, have suggested that AI-based-predictions should only be relied on if someone is continuously checking predictions against reality. O'Neill thinks AIs are too prone to error for us to rely on them when making important decisions unless a human remains in the loop.[126]

Some types of updating cause new difficulties. Typically, including new sensor data in a training set means we can no longer use the old data. And of

---

[123] Institute for the Cities, *Concept Change in Machine Learning*, https://www.wisc.warwick.ac.uk/files/6814/7922/2663/AdamG.pdf (citations omitted).
In the real world concepts and data distributions are often not stable but change with time. This problem, known as concept drift, complicates the task of learning a model from data and requires special approaches, different from commonly used techniques, which treat arriving instances as equally important contributors to the target concept. Among the most popular and effective approaches to handle concept drift is ensemble learning, where a set of models built over different time periods is maintained and the best model is selected or the predictions of models are combined.
A. Tsymbal, M. Pechenizkiy, P. Cunningham, S. Puuronen, Handling Local Concept Drift with Dynamic Integration of Classifiers: Domain of Antibiotic Resistance in Nosocomial Infections, 19TH IEEE SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS 679 (2006).
[124] *See* Carlos E. Perez, The Deep Learning AI Playbook 110 (2017).
[125] *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION 87 (2016); FRANK PASQUALE, THE BLACK BOX SOCIETY (2015).
[126] See CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016). Of course, humans are known to suffer from the same problems, which is what causes bias in the data to begin with. Having a human in the loop may help mitigate problems of bias, but it is not in itself any guarantee.

course, that new sensor data needs to be associated with 'correct' diagnoses for which at present we rely on human experts. Plus, a diagnostic ML with revised training data based on data derived from improved technology, will need to demonstrate anew that it is at least as good as its predecessor. That requires validation data, also at present created by humans. As noted above, however, producing that new data becomes even more difficult if treatment decisions as well as diagnosis have become the province of machines.

Conversely, imagine a period in which new types of data are not coming on stream, but the ML system is making poor diagnoses. What does it do then? If the same set of symptoms is producing the same diagnosis in all cases, where will the ML get the data to suggest which different diagnosis would be better? If the answer is "nowhere" then we have a problem. Again, the problem is likely even more serious if ML takes over treatment as well as diagnosis.

Or, even worse, imagine that the data on which the AI relies has been modified in some way, turning it into a "BadNet".[127] How long would it take before doctors first suspected, then were able to confirm, the existence of a problem? As a leading report on robotics and AI recently warned,

> The whole field of formal modelling, verification measurement and performance evaluation of [Robotics and AI (RAI)] systems is still very much in its infancy: it is critical that one should be able to prove, test, measure and validate the reliability, performance, safety and ethical compliance–both logically and statistically/ probabilistically–of such RAI systems before they are deployed. It should be noted that the verification of systems that adapt, plan and learn will involve the development of new modelling and verification approaches; moreover, such modelling and verification is a prerequisite for informed certification and regulation of RAI systems, which in turn is a factor in public acceptance of RAI.[128]

---

[127] For chilling scenarios see Tianyu Gu et al., *BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain*, ARXIV:1708.06733 [CS] (2017), http://arxiv.org/abs/1708.06733 (last visited Aug 29, 2017).

[128] Joint written evidence submitted by AAAI and UKCRC (ROB0021), http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/robotics-and-artificial-intelligence/written/32533.html (cited with approval in House of Commons, Science and Technology Committee, Robotics and artificial intelligence Fifth Report of Session 2016–17 at p. 16 (Oct. 12, 2016)).

Even with better validation protocols than currently exist, human observers may have real difficulty observing that a problem exists: As systems become more complex, "human operators may have greater uncertainty regarding the conditions under which the system will fail" due to an inability to confidently verify the behavior of the system under all possible operating conditions.[129]

Several of these problems likely apply to deep learning systems in general, and it might be unfair to expect that future proponents of AI-based health care solve them on their own. Either way, there are two extremely important problems that accompany the delegation of medical diagnostics and treatment to ML: the extent to which legal as well as economic pressure will drive actors to prefer the AI over humans, and the risk to life that might be caused by an over-dependence on AI-produced training data in the future.

In our next part we canvass possible solutions to the risk of over-reliance on AI diagnosticians.

[129] Paul Scharre, Center for a New American Century, *Autonomous Weapons and Operational Risk* 18 (February 2016).

IV. Sorting Potential Solutions

One of the simplest potential solutions, at least conceptually, is to impose legal rules or other governance mechanisms that ensure we have an adequate cadre of human physician-diagnosticians. Of course, the goal is not merely to impose a quota of warm bodies. It is to retain and retrain scientists and physicians who will continue experimentation with better solutions,[130] and who will maintain a meaningful and complimentary role, working with ML to create new training data, adjust the performance criteria, and certify the decisions of the ML-system. This aim is clearly in tension with the trends suggested in Parts I and II above, and would certainly be costly. Nevertheless, we return to this idea after first canvassing a variety of other potential technical, economic, and legal solutions.

A. Desiderata

The perfect, or at least good, solution to avoiding a scenario in which both legal rules and economic choices result in vastly reduced if not outright collapse of human participation in the improvement of various diagnostic and treatment specialties (thus eliminating the expertise needed to monitor the performance of ML systems and to create new training data when needed), would have the following properties:

- It would be consistent with *primum non nocere*, in that it would not involve any rule change with negative side-effects on other area of law, ethics, or technology.
- It would at best create incentives to give patients the best medical treatment affordable. At the very least it would impose no impediment to an evolving standard of care, and would never incentivize the definition of a legal standard of care worse than what could reasonably be provided given the overall state of the art.
- It would not create incentives that would tend to reduce the progress of medical research, or tend to leave us less well-able to react to medical emergencies such as new diseases and epidemics.
- It would be resistant to, or ideally invulnerable to, the dangers of monoculture and over-reliance on ML as identified above in Part III.
- It would at best allow capture of any cost-savings enabled by new technology. At the very least it would incentivize cost savings consistent with the ethical

[130] We would also need to ensure that there is a mechanism which allows ML systems to respond quickly to scientific and medical findings by overriding whatever the ML systems had previously been doing.

and legal obligations to give patients at least the standard of care, given the overall state of the art.
- It would have a bottom line that is consistent with the 'Standard View' of biomedical ethics—namely, "that the practice of medicine and nursing are ineluctably human."[131]

Spoiler alert: We do not have a perfect solution that meets all these criteria. In what follows we discuss various imperfect solutions and warn against particularly bad ones. Even our best solution has negative characteristics.

One challenge which seems to emerge from what follows results from the interaction between economic and legal incentives. It seems to be the case that a change to legal rules that fails to adequately deal with the effects of the economic incentives likely will not achieve much. To a great extent the reverse also seems to hold. So to be viable, it would appear that a solution must overcome both sets of incentives.

In spite of this one-two punch, it is important to state as a framing principle that we should not allow the entanglement of law and economics to become an impermeable barrier. If pressure from law and cost do indeed lead us down a path of over-reliance on mechanized medicine and this truly does create a risk of either bad outcomes or a reduction in the creation of better outcomes, then in accord with our bottom line desiderata stated above, we must be sure not to relinquish the human element in medicine. This especially includes access to and human control over the creation of medical knowledge. This point distinguishes our approach to economic considerations regarding ML from how one might approach other crucial diagnostics tools such as functional magnetic resonance imaging (fMRI). One could decide to bite the bullet on costs with either technology purely on the basis of the medical benefits that they provide, but the potential long-run consequences of ML—especially with regards to our ability to understand, control, and access future medical knowledge—remind us that in this case we need to look beyond short-run economic benefits: both Kantian- and utilitarian-based ethics may support the need for a human-centred approach to medicine.

B. Possible Technical and Economic Changes

We could attempt to engineer the national health system in order to enjoy as much of the benefit of ML's enhanced diagnostic abilities as possible without falling into the trap of monoculture or an over-reliance on ML. Depending on their nature,

[131] Supra note 51.

technical changes can be required by law, by the imposition of agreed standards, or self-imposed in response to ethical or market concerns.

1. Create a Control Group?

A potential technical solution would be to divide the population into two groups. One group would receive ML-informed care, while the other group, the control, would not. This is likely a nonstarter if one is convinced that ML is better than physicians, since the control group would then be getting substandard care. The ethical and legal difficulties are complex.

Beyond ethical questions are the practical concerns: running a very large control group would be highly impractical. Not only would it be difficult to decide how big the control group needed to be, but it would be equally challenging to decide how long the experiment needed to run before we reach conclusive results. There is at present no obvious point beyond which we can safely say that if the problems we have identified have yet to manifest we are likely in the clear forever. Conversely, there is no extant standard by which we can decide the ML is so good that the problems we highlighted above are no longer a concern.

Yet, without a control group, relying on human physicians to spot and correct a ML system's errors or especially failure to improve is perilous since the human doctors may not have anything to compare to in order to help them notice. If competing firms have equal access to the entire database, or have access to separate databases that are roughly equal in size and quality, competition might supply the needed monitoring. Unfortunately, for reasons discussed below, access to data may prove to be substantial barrier to entry unless the law changes in some way.[132]

2. Require a 'Red Team' and a 'Blue Team'?

A slightly less bad variant on the control group solution might be to divide the population into two or more groups each of which would be separate for database purposes and have the different groups' data be used by different ML systems. Thus, in effect, we have Dr. Abdul Watson, Dr. Betty Watson, and perhaps even Dr. Chia Watson and so on, each using a different population's data to shape their advice. Every so often – how often? and how? – they would have a virtual medical conference in which they exchange their "best ideas" (or would that be their most telling data?) and in effect upgraded each other's diagnostic suggestions. This seems a poor solution because in the usual case an ML system's accuracy is

[132] See i*nfra* text at notes 128-149.

positively correlated with the sized of the database. Splitting the database into shards creates a risk of sub-optimal care for everyone. Furthermore different systems may offer different trade-offs (e.g. more/less Type I vs Type II error; more explainability vs more accuracy) so cannot be compared directly.

3. Alternate AIs?

A third and perhaps better, if somewhat unlikely, technical solution might be to allow each ML to have the same full database,[133] but require that their programming or training differ in some meaningful way – if this difference can be defined, measured, and (most importantly) maintained, all without subjecting one group to inferior treatment. Using multiple models can add accuracy; were one model best, ethics and law might force us to use it uniquely. If this condition holds over time, the diagnostic problem becomes akin to the hurricane forecasting problem currently faced by meteorologists. There are several competing models, some with different algorithms, others with different coverage and "[t]he best forecasts are made by combining the forecasts from three or more models into a 'consensus' forecast."[134] One group of researchers recently demonstrated that a consensus of multiple models plays Atari video games better than any of the models alone.[135] Because Atari video games are like Go in that identifying the 'success' criteria is automatic and requires no human input,[136] the applications to medical diagnostics remains, at best, for the future. Nonetheless the use of ensemble learning has often been shown to surpass a single learner.[137]

Achieving this scenario would require us to overcome a number of legal and economic complexities. First, we would probably need to have multiple competing providers of AI diagnostic services for it is hard to see what would incentivize a single firm to provide multiple possibly conflicting diagnostic suggestions. Second,

---

[133] A valuable byproduct of a national ML system is that we would not only have more and thus better data for ML systems to chew on, but also we'd have valuable public health data. Identifying environmental issues, e.g. cancer clusters, will be much easier if all patients' diagnostic info is going into a national database in a standard format.

[134] Jeff Maters, *Hurricane and Tropical Cyclones*, WEATHER UNDERGROUND, https://www.wunderground.com/hurricane/models.asp (last viewed Feb. 20, 2017). We are indebted to Jonathan Frankle for pointing us to weather models as an analogy.

[135] Matteo Hessel et al., *Rainbow: Combining Improvements in Deep Reinforcement Learning*, ARXIV:1710.02298 [CS] (2017), http://arxiv.org/abs/1710.02298 (last visited Oct 16, 2017)

[136] *See supra* text at notes 120-121.

[137] Saso Džeroski & Bernard Ženko, *Is Combining Classifiers Better than Selecting the Best One?*, 54 MACHINE LEARNING 255 (2004).

we would need to evolve a standard of care that addressed whether it would suffice to consult (purchase) just one AI model or whether multiple AI opinions would be required. Third, we would need to evolve a method of combining, or sorting among, the competing diagnoses if AI's disagreed that would not expose the person making the decision to unreasonable liability.

Having multiple competing providers of AI diagnostic services that each used a different algorithm should prevent diagnostic monoculture. But any plan that intends to rely on multiple providers must address economic and legal obstacles to creating and sustain multiple providers.

The economic obstacle arises from the nature of the industry, a special case of the winner-take-all phenomenon often observed in markets relying on new technology.[138] We noted above that the economics of deep learning neural networks involved high fixed costs, including gathering and formatting the training data, designing and tuning the relevant algorithms, and perhaps (although here predications vary) the cost of the equipment hosting the AI.[139] Indeed, a widely quoted analysts' report recently cast doubt on the profit potential of IBM's Watson despite its being "one of the more mature and broad cognitive computing platforms today" precisely because users face a high cost of data gathering and curation.[140] In contrast, however, the marginal cost of diagnosing a patient is comparatively small. This account of high fixed costs and low marginal costs resembles the economic profile of a so-called natural monopoly in most respects, save one: other than the contingent question of whether there is sufficient demand to support the capital costs of running multiple competing AIs there is nothing that is an absolute barrier to entry.

---

[138] For discussions of the general phenomenon of winner-take-all in high technology industries see, e.g., Mark A. Lemley & David McGowan, *Legal Implications Of Network Economic Effects*, 86 CAL. L. REV. 479 (1998); Ronald Cass, *Antitrust And High-Tech: Regulatory Risks For Innovation And Competition*, Thomas A. Piraino, Jr., *A Proposed Antitrust Approach To High Technology Competition*, 44 WM. & MARY L. REV. 65 (2002); Cass R. Sunstein, Robert H. Frank, Sherwin Rosen & Kevin M. Murphy, *The Wages Of Stardom: Law And The Winner-Take-All Society: A Debate*, 6 U. CHI. L. SCH. ROUNDABLE 1 (1999).

[139] See *supra* text at note 87.

[140] James Kisner, Jefferies Franchise Note, IBM, *Creating Shareholder Value with AI? Not so Elementary, My Dear Watson* (Jul. 12, 2017), (rating IBM "underperform" due to doubts about Watson) https://javatar.bluematrix.com/pdf/fO5xWjc.

For the multiple-competing-provider scheme to work, all providers need access to sufficient training data[141] and, ideally, they all would have access to all of it since large data sets tend to increase accuracy.[142] Some firms may, however, be able to interpose a legal obstacle to their rivals' access to training data. Training data is not inherently rivalrous. Training an AI is not like siting a water turbine on a river, where there can be only one at any point.[143] But early indications are that would-be providers of AI health-related services see their access to data as a strategic asset to which they wish to have exclusive access. If our strategy for avoiding monoculture relies on having multiple equally competent providers, then, as Amanda Levendowski has argued in the context of avoiding training bias, the legal system may need to remove existing regulatory obstacles to data sharing. Levendowski suggests that using training data be per se fair use.[144] But if trade secret and proprietary first-mover advantages are among the main obstacles to access,[145] then even a copyright workaround may not be enough; in time we may need to impose some sort of compulsory licensing scheme on holders of the data. Compulsory license schemes require the owner of an intellectual property right to share it on reasonable terms. US law does not tend to give compulsory licenses, but they do exist as antitrust remedies[146] and in relatively unusual provisions of existing law relating to patents in essential foods[147] and atomic energy[148] and for

---

[141] "Deep learning requires very large quantities of data in order to build up a statistical picture." Alex Hern, *Why data is the new coal*, THE GUARDIAN, Sep. 27, 2016, https://www.theguardian.com/technology/2016/sep/27/data-efficiency-deep-learning (last visited Oct 1, 2016) (quoting Imperial College Professor Murray Shanahan).

[142] To this end, the U.S. Department of Energy and the National Cancer Institute are partnering in a "three-year pilot project called the Joint Design of Advanced Computing Solutions for Cancer," designed to assemble and integrate large amounts of data about how tumors respond to treatment. CANCER'S BIG DATA PROBLEM, http://cacm.acm.org/careers/208869-cancers-big-data-problem/fulltext (last visited Oct 21, 2016).

[143] Algorithms, however, are patentable, creating at least temporary monopolies.

[144] Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, -- WASH. L. REV – (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3024938.

[145] For a daunting list of obstacles, see Technology Legal Interoperability: Initial Steps Towards an Analytical Framework, PRIVACY & DATA SECURITY LAW RESOURCE CENTER (BLOOMBERG BNA), http://privacylaw.bna.com/pvrc/7057/split_display.adp?fedfid=121122251&vname=pvlrnotallissues&jd=0000015da36bd172abdfeb7fbdf90002&split=0 (last visited Sep 22, 2017).

[146] See United States v. Besser Mfg. Co., 343 U.S. 444, 447 (1952) (imposing compulsory licensing on a "fair" basis).

[147] 7 U.S.C. § 2404 (empowering Secretary of Agriculture to "declare a protected variety open to use on a basis of equitable remuneration to the owner, not less than a reasonable

(Continued)

copyrights in certain music.[149] Then again, foreign companies based in countries that have national policies designed to encourage access to training data as part of a pro-AI industrial policy may fill the gap without the need for radical changes in US law.[150]

4. Encourage Transparency?

A big part of what makes the monoculture story worrying is how difficult it could be to detect a problem if it occurred. As we noted above, decision making by deep-learning-based AI is notoriously opaque.[151] For example, IBM Watson, as currently engineered, does not clearly explain its decision-making processes in terms that are understandable to most humans. It is possible to formally trace (in the computer's memory) how Watson made its decisions, but it takes time and effort

---

royalty, when the Secretary determines that such declaration is necessary in order to insure an adequate supply of fiber, food, or feed in this country and that the owner is unwilling or unable to supply the public needs for the variety at a price which may reasonably be deemed fair").

[148] 42 U.S.C. § 2183.

[149] 7 U.S.C. § 115.

[150] As Chinese AI expert and investor Kai-Fu Lee says, "The U.S. and Canada have the best AI researchers in the world, but China has hundreds of people who are good, and way more data." Will Knight, *China's AI Awakening*, MIT TECH. REV. (Oct. 10, 2017) https://www.technologyreview.com/s/609038/chinas-ai-awakening/ (quoting Mr. Lee). See also Professor Dame Wendy Hall and Jérôme Pesenti, UK Department for Digital, Culture, Media & Sport and UK Department for Business, Energy & Industrial Strategy, *Growing the Artificial Intelligence Industry in the UK* (Oct. 15, 2017), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf (making multiple recommendations to facilitate UK-based AI access to training data).

[151] See *supra* note 40; *see also* Finale Doshi-Velez et al., *Accountability of AI Under the Law: The Role of Explanation*, ARXIV:1711.01134 [CS, STAT] (2017), http://arxiv.org/abs/1711.01134 (last visited Nov 15, 2017) (discussing technical requirements for AI systems that could provide kinds of explanations that are currently required of humans in light of EU GDPR); Aaron M. Bornstein, *Is Artificial Intelligence Permanently Inscrutable?*. 40 LEARNING NAUTILUS (Sept. 1, 2016), http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable (last visited Sep 7, 2016); *see also* Will Knight, *The Dark Secret at the Heart of AI*, MIT TECHNOLOGY REVIEW (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (last visited Apr 13, 2017) (describing "Deep Patient" and AI that can "anticipate the onset of psychiatric disorders like schizophrenia surprisingly well" using methods opaque to its designers).

to understand the result of that trace.[152] The same problem is present in other ML systems.

Although researchers are increasingly aware of the need for "explainable AI", we are still far from something the average doctor could use in real time to help decide what weight to put on a diagnosis. To the extent, for example, that the explanation consists of a set of weights of various bits of evidence without much in the way of context as to how the neural network chose those weights we are a long way from the user-friendly easy-to-use summary a doctor would need. Moving in that direction, we now have neural networks that can provide a confidence number with the decision. Humans can then use that information to prioritize checking the results with lower confidence. This, however, presumes that the confidence estimate is sufficiently well informed, i.e. that the machine "knows what it knows". So far ML only can guarantee this in some limited settings.[153]

Researchers today are actively working on the explainability problem,[154] and thus there is reason to hope that it will get better. The more that an ML system can provide explanation for its diagnoses, the more scope there will be for people to evaluate it meaningfully and, one presumes, spot mistakes or add value. It follows that the 'centaur' model is most likely to endure if AI becomes less opaque, since there will still be something meaningful for people to do. As noted above, however, should there come a point where the AI is so good that humans are not adding value, all the arguments we make here come rushing back into play.

---

[152] *See* Hamm, *supra* note 93 (describing how Watson erroneously concluded Toronto was in U.S.). Similar attempts have been made to reconstruct AlphaGo's move #37 in game #2 of the first match against Lee Sedol: Cade Metz, *In Two Moves, AlphaGo and Lee Sedol Redefined the Future* (Mar. 16, 2016), https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/

[153] *See* Zachary C. Lipton, *The Mythos of Model Interpretability*, ARXIV:1606.03490 [CS, STAT] (2016), http://arxiv.org/abs/1606.03490 (last visited Jan 15, 2018).

[154] Examples include Dong Huk Park et al., *Attentive Explanations: Justifying Decisions and Pointing to the Evidence* (July 25, 2017), arXiv:1612.04757v2 (using neural network based natural language processing and generation techniques to cooperatively explain the behavior of other neural networks); [*Note: cite to Leilani Gilpin, *Reasonableness Monitors: Steps Towards Explaining Complex Machines* (using deductive reasoning to create a "reasonableness monitor" that detects when cyberphysical systems violate rules encoded in formal logic) once it becomes available]; Tao Lei, Regina Barzilay, and Tommi Jaakkola, *Rationalizing Neural Predictions*, arXiv:1606.04155v2 (Nov. 2, 2016) (exploring how to determine the minimum fragment of the input to a neural network necessary for the decision it reached, thus offering some clarity about the network's rationale).

We are grateful to Jonathan Frankle for pointing us to these examples.

5. Tax ML to Change Incentives?

If the medical industry seeks to substitute ML for radiologists we would expect that in the short-term radiologist salaries might drop, blunting the economic pressure to eliminate them. But, as we have argued above, in the longer run demand could shrink to near zero; meanwhile those medical students whose choice of specialty is influenced by salary will avoid that specialty.

One way to discourage over-reliance on ML, therefore, is to change the economic calculus using tax law. If we can maintain a role for doctors in a manner that is more attractive financially, that will remove the economic incentive to undermine human participation in diagnostic decisions and the planning and delivery of treatment. The malpractice law incentive to choose ML would remain but, as we discuss below, there are some possible legal solutions that do not address the economics and thus a tax solution might be combined with a legal solution.

In theory, one could either tax the use of ML, subsidize the employment of human physicians, or both—perhaps even having the ML tax provide the funds for the subsidies. The idea of a robot tax is a popular one, having been endorsed by none less than science and tech celebrities such as Bill Gates,[155] Elon Musk,[156] and Stephen Hawkins.[157] The idea of a tax has also been criticized as impractical, given we do not have agreed definitions of what constitutes a robot,[158] a critique that applies with nearly equal force to AI and machine learning. The EU Parliament flirted with the idea of a robot tax but ultimately rejected it.[159] The biggest problem, not considered by any of the proposals mentioned here, is that in our view the ultimate aim of the tax is not to create *en masse* disincentives for the development of effective medical ML but, rather, to incentivize the successful development of

---

[155] *See* Kevin J. Delaney, *The Robot That Takes Your Job Should Pay Taxes, Says Bill Gates*, Quartz, (Feb. 17, 2017), https://qz.com/911968/bill-gates-the-robot-that-takes-your-job-should-pay-taxes/.

[156] *See* Catherine Clifford, *Elon Musk: Robots Will Take Your Jobs, Government Will Have To Pay Your Wage*, CNBC (Nov. 4, 2016), https://www.cnbc.com/2016/11/04/elon-musk-robots-will-take-your-jobs-government-will-have-to-pay-your-wage.html.

[157] *See* Doug Bolton, *Stephen Hawkings Says Robots Could Make Us All Rich and Free – But We're More Likely to End up Poor and Unemployed*, INDEPENDENT (Oct 9, 205), https://www.independent.co.uk/life-style/gadgets-and-tech/stephen-hawking-says-robots-could-make-us-all-rich-and-free-but-were-more-likely-to-end-up-poor-and-a6688431.html.

[158] *See*, *e.g.*, Robert J. Kovacev, *The Challenges of Administering a Robot Tax* (Sept. 25, 2017), https://www.steptoe.com/publications-12181.html

[159] Reuters, *European Parliament Calls For Robot Law, Rejects Robot Tax* (Feb. 16, 2017), https://www.reuters.com/article/us-europe-robots-lawmaking/european-parliament-calls-for-robot-law-rejects-robot-tax-idUSKBN15V2KM.

(centaur-type) ML that leaves a meaningful role for human doctors and, most importantly, avoids monoculture by ensuring human access to future medical knowledge and know-how.

How to devise a tax strategy that achieves these ends might prove an insurmountable challenge. In any event, a tax on ML would ultimately be a loss for patients, who would see costs rise; a subsidy from general revenues would not hurt patients as directly but would impose a similar deadweight loss on society. And to the extent that the tax discouraged medical service providers from using ML, patients would suffer from being deprived of a probably superior diagnosis.

6. Tax ML to Support an Expert Corps of Radiologists?

Rather than trying to change incentives, which risks a large deadweight loss, a more interesting scenario would be to set the ML tax at a level sufficient to support a corps of expert radiologists who would be charged with keeping tabs on the ML systems' accuracy, creating new training data as needed, conducting research to improve detection and analysis of scan data, and responding to medical emergencies.

Since there will be few if any relevant market signals, one should not underestimate the difficulty of fixing the right size of such a corps, determining its budget, recruiting and training highly competent persons to join it. Nevertheless, the idea of a reserve corps of specialists at the National Institute of Health, or perhaps spread out among teaching hospitals, does have some allure. Since it would be much smaller than the current number of radiologists, supporting a group of experts would presumably be less expensive than attempting to preserve the entire profession, even at reduced salaries.

An important challenge in setting up such a corps is in designing the appropriate training curriculum for these experts. The ideal profile would be people with both medical training and advanced machine learning training. This is a challenging program of study. The shift in curriculum, requiring medical students to incorporate training in probability, statistics and algorithms, may prove hard to sell for some of the more conservative medical faculties.

C. Possible Changes to Legal Rules

1. Revive the Locality Rule?

In Part I.C we showed how the demise of the locality rule eliminated the ability of physicians to assert a defense of custom, local or otherwise. This, we argued, makes malpractice an engine that will drive the progression towards AI monoculture or at least a potentially dangerous over-reliance on ML. Would a return to the locality rule stop this trend and thus prevent malpractice law from creating the incentives that would tend to make ML displace too many doctors?

The answer is that it would not. Even if the revival of the locality rule was able to delay or blunt malpractice law's impetus to switch to ML, it seems unlikely that a (politically improbable) revival of the locality rule would do much to prevent the problems we have identified above. So long as ML seems to offer significant accuracy increases and cost savings, the push to adopt them and in time reduce the use of human doctors will remain strong. As a result, the hospitals, insurers and private medical practices that choose not to use ML will in time find themselves painted as outliers and laggards even when compared to other hospitals and physicians who are similarly situated geographically or by type of practice.

Furthermore, unless the revival of the locality rule was narrowly cabined to AI based medical technology, it could have vast and unpredictable side-effects as it infected first malpractice claims generally, and then perhaps other areas of the law of professional negligence. As law and economics scholars haves shown, the locality rule imposes substantial costs on society, as it disincentivizes innovation, which means that patients will lose the advantages they would have gained from the adoption of new medical technology.[160] Intuitively, the long-term costs in lost advances would seem very likely to exceed the value of any temporary gains.

2. Create a Broad "ML Exception" to Malpractice Law?

Perhaps, therefore, instead of looking for broad-brush solution, we should just create a judicial or legislative "ML Exception" to malpractice law, by which we would agree that failing to use an ML system in diagnosis was not malpractice.

Unfortunately this broad ML Exception suffers from most of the same problems as the idea that we might revive the locality rule: it fails to take account of economic incentives to deploy ML, which exist independently from the push

[160] See *supra* note 60.

provided by malpractice law.[161] Also, like the locality rule revival, the broad ML exception also seems likely to impose greater social costs than benefits, for to the extent that it removes an incentive to use ML even carefully, it degrades the quality of patient care.

3. Create a Narrow 'ML Exception' to Malpractice Law?

If a broad ML exception is too much, how about a more narrowly tailored one, such as a rule that a human doctor's overruling of an ML system is not malpractice unless grossly negligent, but that failing to do so when needed would be actionable error. In other words, the standard of care would still require *consulting* the ML but it would not be *per se* error to deviate from its diagnostic conclusions. Indeed, we might go further and say the ML's diagnosis was not admissible evidence, although this is probably only a short-term fix at best: Over time one would expect that juries would come to understand that ML was the norm and expect to hear about its diagnosis.

This narrower exception would not relieve medical providers from liability for failing to use ML once it became the standard of care, but would provide a safe harbor from liability for overruling an ML system unless the human's decision was indefensible. We suggested above that under current liability rules, especially in the increasing number of states that have abandoned the locality rule, even a human doctor who believes with some justice that her diagnosis is better than the computer's will face moral risks and obstacles in displacing the AI's suggestion.[162] If nothing else, we suggested, the fact that ML has a better success rate will mean that the physician will run a very great malpractice risk in supplanting its judgment, and that insurers will be loath to permit such decisions as a result. The second form of the "ML Exception" removes, or at least greatly reduces, this risk. In so doing, it departs from the pattern in other contexts, such as piloting, where we believe machines outpace humans.[163]

The second part of the exception, in which human doctors are liable for failing to overrule a ML system when they should have, is not, on its face a change from current law. Under current law, an ML system, being a machine, has no

---

[161] See *supra* § II.A. The incentives could, however, be overcome by taxes. See *supra* § IV.B.5.

[162] Cf. Millar & Kerr, *supra* note 9.

[163] "A court may … infer negligence on the part of the pilot from evidence that suggests that the pilot switched from automatic pilot to manual in a crisis situation." James E. Cooling & Paul V. Herbers, *Considerations in Autopilot Litigation*, 48 J. AIR L. & COMM. 693, 710 (1983).

identity nor agency for legal purposes, and hence its decisions will in all cases be ascribed to the human(s) or corporation(s) responsible for acting on its diagnoses.[164] On the other hand, once ML has a better batting average than the average human, it will, as we've said repeatedly, be a courageous human who overrules it in any but the most obvious cases. Under current law, cases where the computer's decision was arguably plausible but courageously overruled anyway will invite litigation if the outcome goes badly, but cases where the doctor should have overridden the computer but did not will be much harder for plaintiffs to prove if and when ML alone becomes the standard of care.

Thus, the second part of the exception can be characterized as no more than a savings clause, a way to emphasize that while liability for overruling ML is changing, liability for not using ML and for not overruling it remains in place. Alternately, one can see the second clause as a means to emphasize the importance of keeping a human in the loop: liability will lie not only for failing to use ML when one should, but also for failing to overrule it when one should.

Although undoubtedly preferable to any of the rules canvassed so far, the social welfare consequences of this narrower ML Exception are hard to predict with any certainty. Even if we assume, somewhat heroically, that on average humans will overrule ML approximately as often as we would want them to, that leaves open the door for errors in both directions, i.e. overruling the ML system when it was right, and failing to overrule the ML system when it was wrong. The patients in the first group, who would have had the benefit of the ML system's correct diagnosis, will be made worse off compared to the treatment they would have received if the narrow ML Exception did not exist. In contrast, the patients in the second group, who would have suffered from the machine's error in any case, are no worse off than they would have been.

How we measure the cost of the errors to the first group is inevitably difficult; but without any defensible idea of how big that group would be – something we could only establish empirically – it is even more impossible to say. Unfortunately, we can say with some confidence that humans will feel freer to overrule ML systems under this rule than under the current, default, rule. Arguably, this means that the number of patients harmed by ignoring ML's correct diagnosis ought to grow above the baseline.

Furthermore, if this narrow exception suffices to incentivize medical service providers and malpractice insurers to keep a human doctor fully in the loop, then

---

[164] *See* Neil M. Richards & William D. Smart, *How Should the Law Think About Robots* in ROBOT LAW (Ryan Calo, A. Michael Froomkin & Ian Kerr, eds 2017).

we also will lose all or part of any cost savings from having ML replace humans, with the size of the loss depending on both the relative costs and the extent to which human doctors can 'work more efficiently' when paired with ML–i.e. diagnose more quickly and/or more accurately.

Against these costs one should put the speculative, but potentially large, gains caused by creating a data set of human decisions and resulting outcomes that can be used to provide ongoing training data for ML systems. If – and we stress that this may be a big 'if' – humans end up deciding enough cases differently from ML to provide enough examples for training purposes, this may suffice to head off what would otherwise be the monoculture of training data that we warned about in Part III.

One other caveat should be noted: for the human-generated training data to have real value, it needs to include a significant number of cases in which the human's decision was better than ML's, something which likely will turn on how great ML's success rate is. As this point may be obscure, a short elucidation is in order. We assume ML is on average more accurate than people. But neither is 100% accurate. The less accurate the humans are, the less accurate ML needs to be in order to be noticeably better than humans. The less accurate a better-than-humans ML is, the more scope will remain for potential cases in which, were a human to overrule the ML system, they might improve the patient outcome. (Of course, there is also the possibility that they might also both be wrong in different ways but we can collapse that scenario by defining "right" as "better than the other diagnosis".) Conversely, the more accurate ML is overall, the less frequently we would expect to see a human decision to override the ML diagnosis lead to a better outcome.

We return to this issue below.

4. Define the Standard of Care to Require a Human Doctor Plus ML

Rather than create a malpractice exception for human-ML interactions, we could instead fix the legal standard of care (either legislatively or judicially) to require ML plus meaningful review by a human doctor. At present -- while human diagnosticians remain on average superior to ML -- any doctor who uses ML as a decisional aid is in effect subject to this standard of care. We suggested above that once ML is provably superior to the average human, the standard of care would change, setting off a chain of events we fear could be deleterious in the long term. Freezing the standard of care to require meaningful human participation would head off those consequences. Indisputably, "meaningful" is a somewhat vague term, and it invites some fact-based debate as to what level of review by a human doctor would suffice. In the abstract, however, it is very hard to define the appropriate

level of review with any precision; litigation in courts may actually be a good way of developing the factual records needed to put more detail into this standard.

Both the broad and narrow "ML Exceptions" to malpractice take large swaths of human liability out of the equation; in so doing they leave the choice of using a person or an AI to other factors, namely ethics[165] and cost. In contrast, setting the standard of care to require both ML and humans invokes law to override those ethical and economic concerns, but does so at the possible price of forgoing a larger number of beneficial outcomes that will not happen because the AI plus physician is too expensive. The risk here is that some people may not be able to afford care that they otherwise might have had.

On the other hand, freezing the standard of care makes it more likely than does the narrow ML Exception that the rate of humans overrides of ML will tend towards the optimal level, where 'optimal' refers to individual patient outcomes without considering systemic effects on training data. Under the narrow exception, humans are protected from liability for overruling ML in the absence of gross negligence and this opens the door to excessive overrides. In contrast, setting the standard of care leaves current standards in place. Plaintiffs who wish to argue that a physician should have deferred to the ML will not be able to argue a *per se* violation of the standard of care, but doctors challenged for overriding ML will have to make the ordinary fact-based showing that their decisions were appropriate.

Even if the above is correct, and that this proposal comes closest to incentivizing an 'optimal' rate of human overrides of ML diagnoses, we cannot be confident that it will necessarily provide a sufficient supply of human-generated accurate training data. How much data people will create depends on a number of variables that can only be estimated once ML is up and running full speed. The two chief variables are ML's failure rate, and what fraction of those failures are detected and corrected by the human reviewers. (Recall that when humans wrongly override a correct diagnosis, this does not produce useful training data for ML; it might, however, provide useful training data for medical students.) We cannot know at this early stage whether the correct corrections will suffice, but this option probably gives as much hope as any, and more than most; the only one that comes close is the narrow ML Exception, and that because its incentive effects are likely to be similar.

[165] Compare Millar and Kerr, *supra* note 9, with sources cited *supra* note 51.

V. Conclusion: The Least Worst Solution Will be Expensive

We have argued that if and when AI can outperform human doctors both malpractice law and, if pricing warrants it, economic imperatives will push providers to substitute machines for human doctors. This is not as wonderful as it may sound to technophiles because it creates a subtle risk of a closed loop as well as the obvious (short-run) opportunities for better patient care.

The risk is a result of AI's great promise. If, as we assumed for the purposes of this article, some future machine learning system becomes significantly better at some types of diagnosis, such as reading X-rays and other radiological studies, then medical skills may suffer; if and when ML takes over treatment, some specialties may all but disappear. The problem we are concerned with is not directly the employment prospect of present or future radiologists. The problem is that the over-reliance on AI, and the resulting loss of medical knowledge, can create a closed loop in which future training and validation data sets are the result of decisions by the AI itself. At that point, we may lose the ability to discover new better treatments, in the case where the ML system settles for a sub-optimal solution or the ML chooses a solution that optimizes a narrow performance criterion.

We can head off this scenario in a number of ways. The simplest legal change would be to require that a human be fully and meaningfully in the loop in all cases. Preventing a ML alone from becoming the standard of care, and thus defining the standard as ML plus a physician meaningfully involved in reviewing the diagnostic decision, could alleviate the problem. We may also need to tinker with malpractice rules in order to prevent humans from being too unwilling to overrule an AI for fear of liability.

Admittedly, keeping physicians fully in the loop is likely to prove expensive compared to an AI-only world. Further, even if it may be a long-term fix we should not expect it to be permanent. We will need to continue to revisit the level at which machines and humans integrate and exchange information, and make decisions. Perhaps worst of all, our solution has more than enough of a whiff of the Luddite to make any robot or AI enthusiast uncomfortable.[166] Nevertheless, we see no better answer at present; the remaining challenges will focus on the proper alignment of humans and machines in order to integrate and exchange information, to make and

[166] That said, the public is at present showing a Luddite tendency: In a 2017 poll of 2200 American adults, 65% were "very uncomfortable" (44%) or "somewhat uncomfortable (21%) with the idea of "an AI making a medical diagnosis." Morning Consult National Tracking Poll #170401, Table BRD7_& at p. 62, https://morningconsult.com/wp-content/uplods/2017/04/170401_crosstabs_Brands_v3_AG.pdf.

carry out medial decisions. Figuring out how best to deal with the alignment questions will be a key consideration in the modernization of medical school curricula, so that next generation of medical professionals are adequately trained to work with ML.

Modern auto-pilots are capable of flying jets from takeoff to landing, yet we still require human pilots to be in the cockpit in case of emergency and despite the arguable duplication of expense. Meanwhile, whether reliance on automation has caused a dangerous deskilling of pilots is a live debate. Now it's medicine's turn.