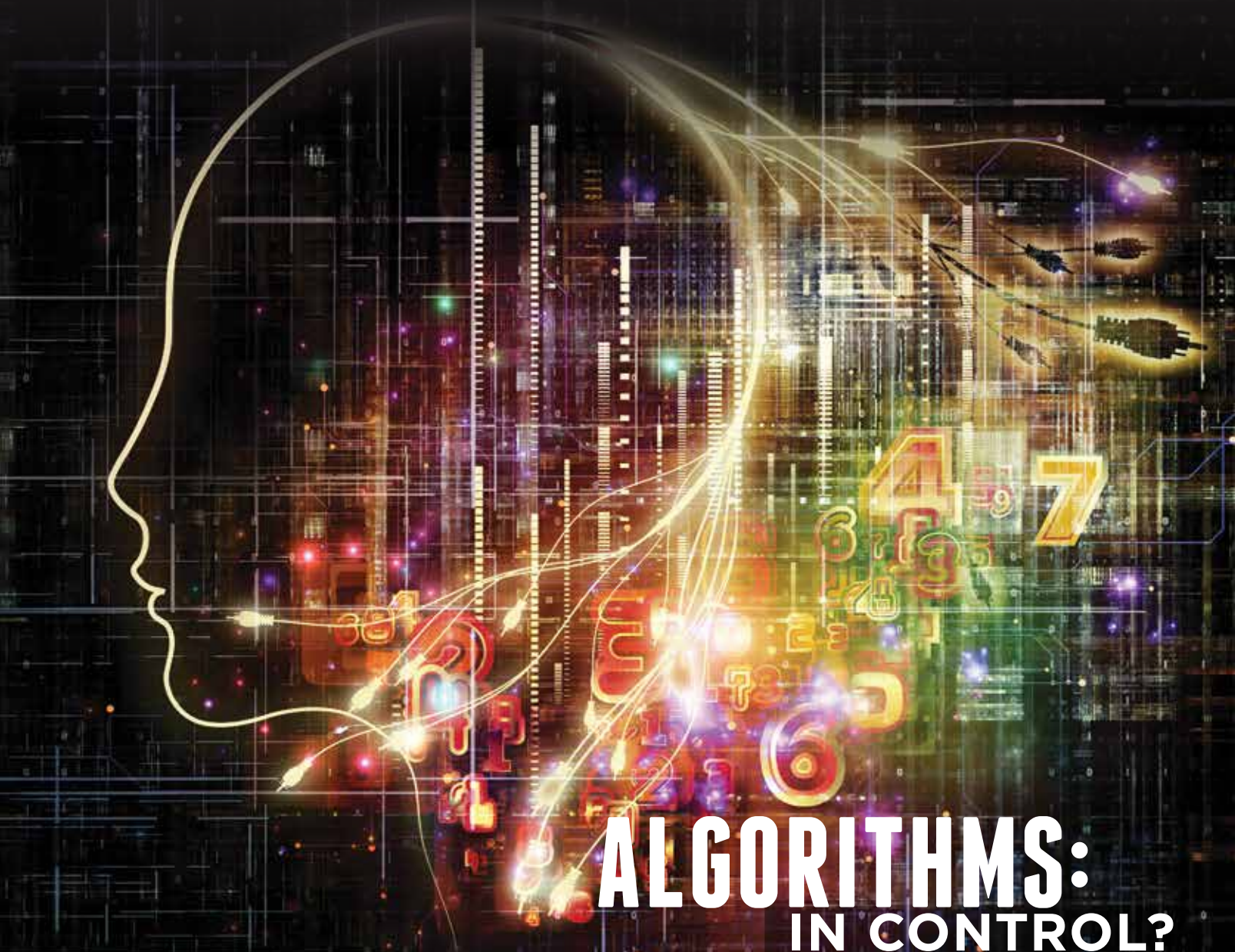
IN THIS ISSUE MACHINE LEARNING • MEDICINE • AUTONOMOUS VEHICLES

THE SciTech LAWYER

VOLUME 14 ISSUE 1 | FALL 2017 | SECTION OF SCIENCE & TECHNOLOGY LAW | AMERICAN BAR ASSOCIATION



ALGORITHMS: IN CONTROL?

LISA R. LIFSHITZ, LOIS D. MERMELSTEIN, AND LARRY W. THORPE, ISSUE EDITORS

THE SciTech LAWYER

The SciTech Lawyer (ISSN 1550-2090) is published quarterly as a service to its members by the Section of Science & Technology Law of the American Bar Association, 321 North Clark Street, Chicago, IL 60654-7598. It endeavors to provide information about current developments in law, science, medicine, and technology that is of professional interest to the members of the ABA Section of Science & Technology Law. Any member of the ABA may join the Section by paying its annual dues of $55. Subscriptions are available to nonmembers for $55 a year ($65 for foreign subscribers). Some back issues are available for $12 plus a $3.95 handling charge from the ABA Service Center, American Bar Association, 321 North Clark Street, Chicago, IL 60654-7598; 1-800-285-2221. Requests to reprint articles should be sent to ABA Copyrights & Contracts, www.americanbar.org/utility/reprint/Periodicals; all other correspondence and manuscripts should be sent to *The SciTech Lawyer* Contract Editor Melissa Vasich, melissa@vasich.com. For more information, visit www.americanbar.org/publications/scitech_lawyer_home.html. The material published in *The SciTech Lawyer* reflects the views of the authors and has not been approved by the Section of Science & Technology Law, the Editorial Board, the House of Delegates, or the Board of Governors of the ABA.

MESSAGE FROM THE CHAIR

Eileen Smith Ewing, Chair 2016–17

Artificial Intelligence: Revolution or Evolution?

Many fear advances in artificial intelligence (AI). No less a mind than that of Stephen Hawking said, "The development of full artificial intelligence could spell the end of the human race. . . . It would take off on its own, and re-design itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded."

Of course, other promising, revolutionary technological advances have raised similar, Armageddon-like fears: cloning and gene alteration, to name a few. As lawyers committed to the use of science for the betterment of humanity, we face a special challenge—it is we who must work alongside scientists to promote, monitor, and regulate the best uses of technological innovation—and to draft policies on potentially harmful uses. It's a significant job, but one the members of this Section (and indeed readers of and contributors to this magazine) are uniquely qualified to approach.

In their respective articles in this issue, Natasha Duarte and April Doss each take on broad ethical and policy issues affecting the development of AI across fields of knowledge. Duarte introduces possible ethical frameworks for the use of AI in analyzing big data and making automated decisions. Doss argues that our traditional, common-law approach to forming law and policy—namely, the gradual accumulation of judicial decisions—is simply not dynamic enough to meet the rapidity of change in areas like AI.

There are a number of excellent pieces in this issue on more granular AI topics as well. In medicine, Matt Henshon advises the use of AI in small ways, to enhance human decision making. Aubrey Haddach and Jeffrey Licitra demonstrate the high human cost of an AI-driven false positive in medical diagnostics. Privacy issues come to the fore in Kay Firth-Butterfield's article on data privacy and AI: the European Union, for one, seeks transparency in how automated decision-making systems may reach adverse decisions against consumers.

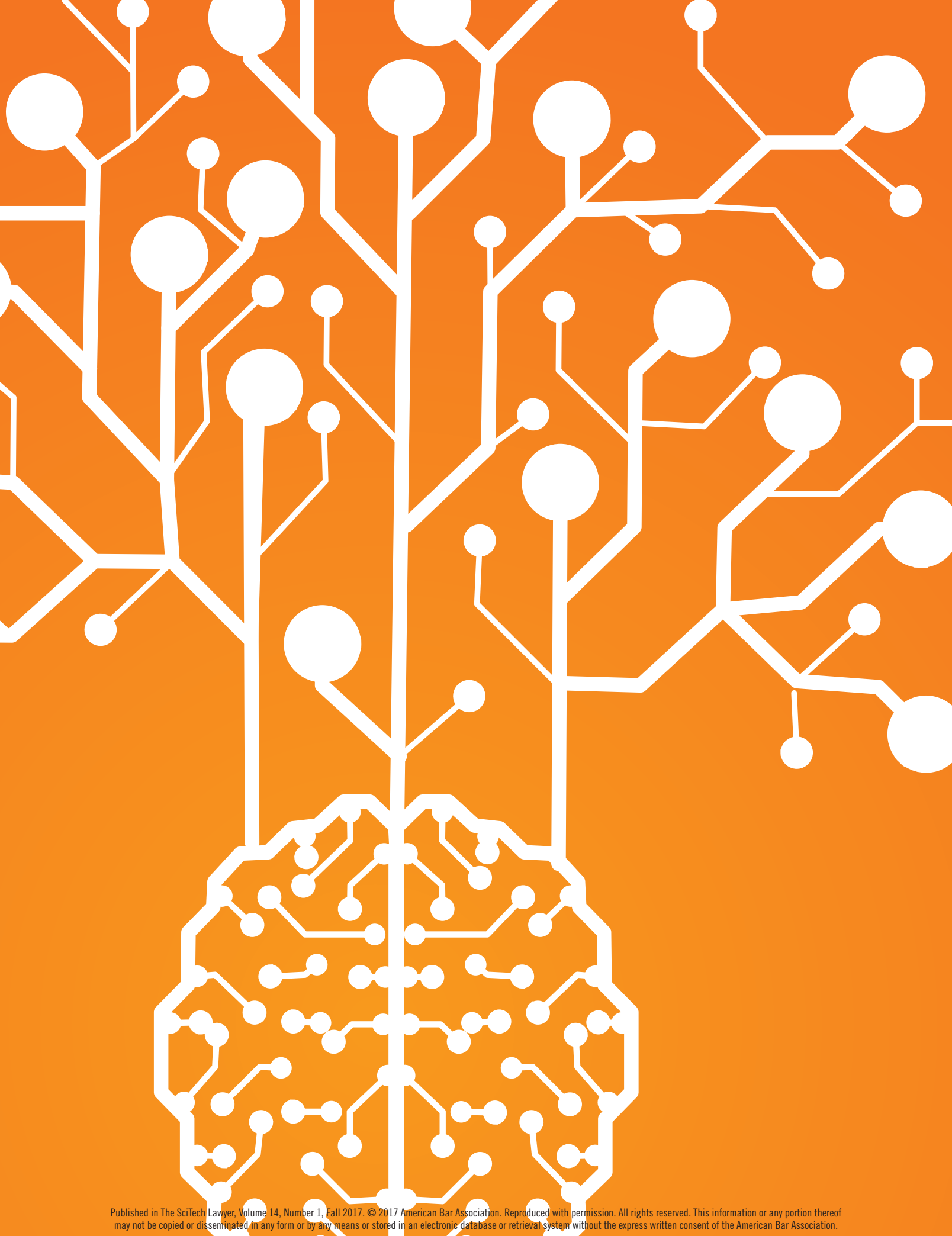
Professor Gary Marchant offers a hopeful note on how AI may affect the practice of law. He concludes we lawyers face an evolution, not a revolution, as many developing technologies will benefit our efforts but not replace the need for human oversight.

Speaking of change and evolution, this is my farewell column as Section Chair. In August, we had a very productive ABA Annual Meeting in New York City, during which we cosponsored a very successful ABA Showcase Program on Cybersecurity (the Section's Eric Hibbard was among the panelists). Our past Section Chair Heather Rafter offered a timely CLE panel on copyright issues affecting music and technology. Committee leaders Katherine Lewis (our incoming Secretary) and Barron Oda provided two fascinating arts-focused CLE programs—one on virtual and augmented reality; the other on technological advances in detecting art forgery.

At the conclusion of the Annual Meeting, it was my privilege to pass the gavel to our new Section Chair: David Z. Bodenheimer of Washington, D.C. Nationally ranked by Chambers USA in the area of government contracts, David's expertise in that field and in related areas, such as privacy, cybersecurity, and homeland security, will be a great boon to our Section. I commend him to you all. ◆

TABLE OF CONTENTS

Published in The SciTech Lawyer, Volume 14, Number 1, Fall 2017.



BY WARREN E. AGIN

A SIMPLE GUIDE TO MACHINE LEARNING

Lawyers know a lot about a wide range of subjects—the result of constantly dealing with a broad variety of factual situations. Nevertheless, most lawyers might not know much about machine learning and how it impacts lawyers in particular. This article provides a short and simple guide to machine learning geared to attorneys.

"Artificial intelligence" (AI) usually refers to machine learning in one form or another. It might appear as the stuff of science fiction, or perhaps academia, but in reality machine learning techniques are in wide use today. Such techniques recommend books for you on Amazon, help sort your mail, find information for you on Google, and allow Siri to answer your questions.

In the legal field, products built on machine learning are already starting to appear. Lexis and Westlaw now incorporate machine learning in their natural language search and other features. ROSS Intelligence is an AI research tool that finds relevant "phrases" from within cases and other sources in response to a plain language search. Through the use of natural language processing, you can ask ROSS questions in fully formed sentences. Kira Systems uses machine learning to quickly analyze large numbers of contracts.

These are just two of dozens of new, machine learning–based products. On the surface, these tools might seem similar to those currently available—but they actually do something fundamentally different, making them not only potentially far more efficient and powerful, but also disruptive. For example, machine learning is the "secret sauce" that enables ridesharing services like Uber to efficiently adjust pricing to maximize both the demand for rides and the availability of drivers, predict how long it will take a driver to pick you up, and calculate how long your ride will take. With machine learning, Uber and similar companies are rapidly displacing the traditional taxicab service. Understanding what machine learning is and what it can do is key to understanding its future effects on the legal industry.

What Is Machine Learning?

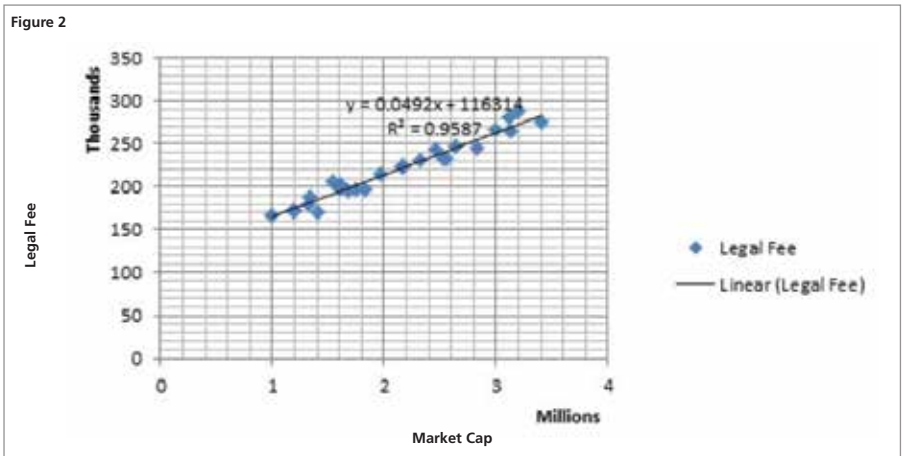
Humans are good at *deductive reasoning*. For example, if I told you that a bankruptcy claim for rent was limited to one year's rent, you would easily figure out the amount of the allowed claim. If the total rent claim was $100,000, but one year's rent was $70,000, you would apply the rule and deduce that the allowable claim is $70,000. No problem. You can determine the result easily, and you can also easily program a computer to consistently apply that rule to other situations. Now reverse the process. Assume I told you that your client was owed $100,000 and that the annual rent was $70,000, and then told you that the allowable claim was $70,000. Could you figure out how I got that answer? You might guess that the rule is that the claim is limited to one year's rent, but could you be sure? Perhaps the rule was something entirely different. This is *inductive reasoning*, and it is much more difficult to do.

Machine learning techniques are computational methods for figuring out "the rules," or at least

Warren E. Agin *(wea@swiggartagin.com) is a principal of Analytic Law LLC in Boston, which helps law firms and legal departments find quantitative solutions to legal problems. He also chairs the ABA Business Law Section's Legal Analytics Committee and teaches legal analytics as an adjunct professor at Boston College Law School. You can follow him on Twitter at @AnalyticLaw.* *Mr. Agin thanks Michael Bommarito of LexPredict and Thomas Hamilton of ROSS Intelligence for kindly reviewing and commenting on an earlier version of this article, published in the February 2017 issue of* Business Law Today, *but emphasizes that any errors are his, not theirs.*

Figure 1

Figure 2



approximations of the rules, given the factual inputs and the results. Those rules can then be applied to new sets of factual inputs to deduce results in new cases.

For instance, consider number series games. For example:

2 4 6 8 10 ?

The next number is 12, right? Here, the inputs are the series of numbers 2 through 10, and from this we induce the rule for getting the next number—add 2 to the last number in the series. Here is another one:

1 1 2 3 5 ?

The next number is 8. This is a Fibonacci sequence, and the rule is that you add together the last two numbers in the series.

With these games, what you are doing in your head is looking at a series of inputs and answers, and using inductive reasoning to figure out the rule. You then apply that rule to get the next number. Broken down a little, the prior game looks like this:

Input	Result
1 1	2
1 1 2	3
1 1 2 3	5
1 1 2 3 5	?

We look at the group of inputs and induce a rule that gives us the displayed results. Once we have derived a workable rule, we can apply it to the last row to get the result 8, but more importantly we can apply it to any group of numbers in the Fibonacci sequence. This is a simple (very simple) example of what machine learning does.

Let's take a more complex example. Assume we wanted to predict the amount of a debtor's counsel's fees in a Chapter 11 case. We could take a look at cases in the past and get information about each: for example, the number of creditors, the debtor's market capitalization, where the case was filed, and, of course, the eventual fee awarded to the debtor's counsel. We might compare these numbers and discover that if we graphed the fee awards against the debtor's market capitalization, it looks something like figure 1 (purely hypothetically).

There seems to be a trend. The larger the market capitalization (the *x* axis), the higher the legal fee seems to be. In fact, the data points look sort of like a line. We can calculate the line that best fits the data points using a technique called linear regression (see fig. 2).

We can even see the equation that the line represents. You take the market capitalization for the debtor, multiply it by 4.92 percent, and add $116,314 (these two variables are the "weighting mechanisms," explained in detail below). This is called a "prediction model." The prediction model might not perfectly fit the data used to create it—after all, not all the data points fall exactly on the line—but it provides a useful approximation. That approximation will provide a pretty good estimate for legal fees in future cases (that's what the R^2 number on the graph tells us). For the record, the data here is imaginary, hand-tailored to demonstrate the methodology.

Naturally, real-world problems are more complex. Instead of a short series of numbers as inputs, a real-world problem might use dozens, perhaps thousands, of possible inputs that might be applied to an undiscovered rule to obtain a known answer. We also do not necessarily know which of the inputs the unknown rule uses!

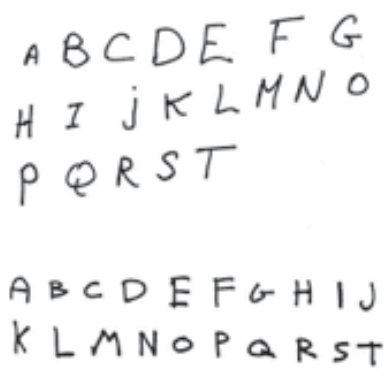
To solve a more complex problem, we might begin by building a database with the relevant points of information about a large number of cases, in each instance collecting the data points that we think might affect the answer. To build our prediction model, we would select cases at random to use as a "training set," putting the remainder aside to use as a "test set." Then we would begin to analyze the various relationships among the data points in our training set using statistical methods. Statistical analytics can help us identify the factors that seem to correlate

with the known results and the factors that clearly do not matter.

Advanced statistical methods might help us sort through the various relationships and find an equation that takes some of the inputs and provides an estimated result that is pretty close to the actual results. Assuming we find such an equation, we then try it out on the test set to see if it does a good job there as well—predicting results that are close to the real results. If our predictive model works on our test set, then we consider ourselves lucky. We can now predict a debtor's counsel's legal fees ahead of time; at least until changing circumstances—perhaps rules changes, a policy change at the U.S. Trustee's Office, or the effect our very own model has on which counsel get hired for cases—render our model inaccurate. If our model does not work on the test set data, than we consider it flawed and go back to the drawing board.

For real-world problems, this kind of analysis is difficult. The job of collecting the data, cleaning it, and analyzing it for relationships takes a lot of time. Given the large number of potential variables that affect real-world relationships, identifying those that matter is somewhat a process of trial and error. We might get lucky and generate results quickly, we might invest substantial resources without finding an answer at all, or the relationships might simply prove to be too complex for the methods I described to work adequately. Inductive reasoning is difficult to do manually. This brings us to machine learning. Machine learning can efficiently find relationships using inductive reasoning.

As an example of what machine learning can do, consider these images:



Assume we want to set up a computer system to identify these handwritten images and tell us what letter each image represents. Defining a rule set is too difficult for us to do by hand and come up with anything that is remotely usable, but we know there *is* a rule set. The letter *A* is clearly different from the letter *P*, and *C* is different from *G*, but how do you describe those differences in a way a computer can use to consistently determine which image represents which letter?

The answer is that you don't. Instead, you reduce each image to a set of data points, tell the computer what the image is of, and let the computer induce the rule set that reliably matches all the sets of data points to the correct answers. For the image recognition problem, you might begin by defining each letter as a 20 pixel by 20 pixel image, with each pixel having a different grayscale score. That gives you 400 data points, each with a different value depending on how dark that pixel is. Each of these sets of 400 data points is associated with the answer—the letter they represent. These sets become the training set, and another database of data points and answers is the test set. We then feed that training set into our machine learning algorithm—called a "learner"—and let it go to work.

What does the learner actually do? This is a little more difficult to explain, partially because there are a lot of different types of learners using a variety of methods. Computer scientists have developed a number of different kinds of techniques that allow a computer program to infer rule sets from defined sets of inputs and known answers. Some are conceptually easier to understand than others. In this article, I describe, in simple terms, how one of these techniques works. Machine learning programs will use a variation of one or more of these techniques. The most advanced systems include several techniques, using the one that fits the specific problem best or seems to generate the most accurate answers.

In general, think of a learner as including four components. First, you have the input information from the training set. This might be data from a structured, or highly defined, database, or unstructured data like you might find in a set of discovery documents. Second, you have the answers. With a structured database, a particular answer will be closely identified with the input information. With unstructured information, the answer might be a category, such as which letter an image represents or whether a particular email is spam; or the answer might be part of a relationship, such as text in a court decision that relates to a legal question asked by a researcher. Third, you have the learning algorithm itself—the software code that explores the relationships between the input information and the answers. Finally, you have weighting mechanisms—basically parts of the algorithm that help define the relationships between the input information and the answers, within the confines of the algorithm. Once you have these four components, the learner simply adjusts the weighting mechanisms in a controlled manner until it finds values for the weighting mechanisms that allow the algorithm to accurately match the input information with the known correct answers.

Let's see how this might work with my hypothetical system for estimating a debtor's counsel's fees. In the example (see fig. 2), the market capitalizations are the input information (X). The known legal fees for each case are the answers (Y). For purposes of illustration, let's assume the algorithm is $Y = aX + b$ (a vast simplification, but I'm going to use it to demonstrate a point). The weighting mechanisms are the two variables a and b. Instead of manually calculating the values of a and b using linear regression, a machine learning program might instead try different values of a and b, each time checking to see how well the line fits the actual data points mathematically. If a change in a or b improves the fit of the line, the learner might continue to change a and b in the same direction, until the changes no longer improve the line's fit.

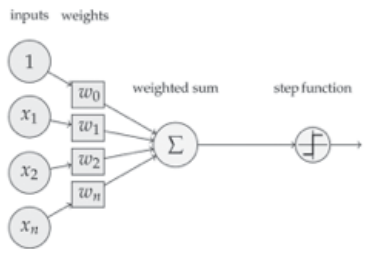
Of course, in my example it is easier just to calculate a and b using linear regression techniques. I don't even

need to have math skills to do it—the functionality is built right into Microsoft Excel and other common software products. Given a spreadsheet with the data, I can perform the calculation with a few mouse clicks. Machine learning programs, however, can figure out the relationships when there are millions of data points and billions of relationships—when modeling the systems is impossible to do by hand because of the complexity. Machine learning systems are limited only by the quality of the data and the power of the computers running them.

Now, let's look at an example of a machine learning system.

Neural Networks

The term "neural network" conveys the impression of something obscure and mysterious, but it is probably the easiest form of a machine learning system to explain to the uninitiated. This is because it is made up of layers of a relatively simple construct called a "perceptron."



Credit: https://blog.dbrgn.ch/2013/3/26/perceptrons-in-python/

This perceptron contains four components, the first being one or more inputs represented by the circles on the left. The input is simply a number, perhaps between 0 and 1. It might represent part of our input information, or it might be the output from another perceptron.

Second, each input number is given a weight—a percentage by which the input is multiplied. For example, if the perceptron has four inputs of equal importance, each input is multiplied by 25 percent. Alternatively, one input might be multiplied by 70 percent while the other three are each multiplied by 10 percent, reflecting that one input is far more important than the others.

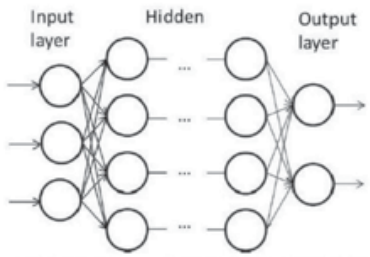
Third, these weighted input numbers are added to generate a weighted sum—a single number that reflects the weights given the various inputs.

Fourth, the weighted sum is fed into a step function. This is a function that outputs a single number based on the weighted sum. A simple step function might output a 0 if the weighted sum is between 0 and 0.5, and a 1 if the weighted sum is between 0.5 and 1. Usually a perceptron will use a logarithmic step function designed to generate a number between, say, 0 and 1 along a logarithmic scale so that most weighted values will generate a result at or near 0, or at or near 1, but some will generate a result in the middle.

Some systems will include a fifth element: a "bias." The bias is a variable that is added or subtracted from the weighted sum to bias the perceptron toward outputting a higher or lower result.

In summary, the perceptron is a simple mathematical construct that takes in a bunch of numbers and outputs a single number. By computing the weighted sum of the inputs, running that number through the step function, and adjusting the result using a final bias, the perceptron tells you whether the collection of inputs produces a result above or below a threshold level. This mechanism works much like a switch. The result of that switch might be fed to another perceptron, or it might relate to a particular "answer." For example, if your learner is doing handwriting recognition, you might have a perceptron that tells you the image is the letter *A* based on whether the output number is closer to a 1 than a 0.

In a neural network, the perceptrons typically are stacked in layers. The first layers receive the input information for the learner, and the last layer outputs the results.



Credit: http://www.intechopen.com/books/cerebral-palsy-challenges-for-the-future/brain-computer-interfaces-for-cerebral-palsy

In between are what are called "hidden layers" of perceptrons, each taking in one or more input numbers from a prior layer and outputting a single number to one or more perceptrons in the next layer. By stacking the layers of perceptrons, the "deep learner" acts a little bit like a computer circuit, one whose operations are programmed by the changes in the weights.

The computer scientist building the neural network determines its design—how many perceptrons the system uses, where the input data comes from, how the perceptrons connect, what step function gets used, and how the system interprets the output numbers. However, the learner itself decides what weights are given to each input as the numbers move through the network, and what biases are applied to each perceptron. As the weights and biases change, the outputs will change. The learner's goal is to keep adjusting the weights and biases used by the system until the system produces answers using the input information that most closely approximate the actual, known answers.

Returning to the handwriting recognition example, remember that we broke down each letter image into 400 pixels, each with a grayscale value. Each of those 400 data points would become a input number into our system and be fed into one or more of the perceptrons in the first input layer. Those outputs would pass through some hidden layers in the middle. Finally, we would have an output layer of 26 perceptrons, one for each letter. The output perceptron with the highest output value will tell us what letter the system thinks the image represents.

Then, we pick some initial values for the weights and biases, run all the samples in our training set through the system, and see what happens. Do the output answers match the real answers? Probably not even close the first time through. So, the system begins adjusting weights and biases, with small, incremental changes, testing against the training set and continuously looking for improvements in the results until it becomes as accurate as it is going to get. Then, the test set is fed into the system to

see if the set of weights and biases we just determined produces accurate results. If it does, we now have an algorithm that we can use to interpret handwriting.

It might seem a little like magic, but even a relatively simple neural network, properly constructed, can be used to read handwriting with a high degree of accuracy. Neural networks are particularly good at sorting things into categories, especially when using a discrete set of input data points. What letter is it? Is it a picture of a face or something else? Is a proof of claim filed in a bankruptcy case objectionable or not?

Machine Learning in Action

These examples are basic, designed to provide some understanding of what are fairly abstract systems. Machine learners come in many flavors—some suitable for performing basic sorting mechanisms, and others capable of identifying and indexing complex relationships among information in unstructured databases. Some systems work using fairly simple programs and can run on a typical office computer, and others are highly complex and require supercomputers or large server farms to accomplish their tasks.

To understand the power of machine learning systems compared with non-learning analytic tools, let's revisit an earlier example: ROSS Intelligence. ROSS is built on the IBM Watson system, although it also includes its own machine learning systems to perform many of its tasks. Watson's search tools employ a number of machine learning algorithms working together to categorize semantic relationships in unstructured textual databases. In other words, if you give Watson a large database of textual material dealing with a particular subject, Watson begins by indexing the material, noting the vocabulary and which words tend to associate with other words. Even though Watson does not actually understand the text's meaning, it develops, through this analysis, the ability to mimic understanding by finding the patterns in the text.

For example, when you conduct a Boolean search in a traditional service for "definition /s 'adequate protection,'" the service searches its database for an exact match where the word "definition" occurs in the same sentence as the term "adequate protection." ROSS does something different. Using the Watson AI systems and its own algorithms, it looks within the search query for word groups it recognizes and then finds the results it has learned to associate with those word groups. If you search for "what is the definition of adequate protection," the system will associate the query "what is the definition" with similar queries, such as "what is the meaning of" or just "what is." It will also recognize the term "adequate protection" as a single concept instead of two separate words, and likely, given the context, understand it as a word found in bankruptcy materials. Finally, it will have associated a successful response as being one that gives you certain types of clauses including the term "adequate protection." It will not understand specifically that you are looking for a definition, but because others who used the system and made similar inquiries preferred responses providing definitions, you will get clauses containing similar language patterns and, viola, you will get your definition.

You should not even have to use the term "adequate protection" to get an answer back discussing the concept when that is the appropriate answer to your question. So long as your question triggers the right associations, the system will, over time, learn to return the correct responses.

The key is that a machine learning system learns. In a way, we do the same thing ROSS does. The first time we research a topic, we might look at a lot of cases and go down a lot of dead ends. The next time, we are more efficient. After dealing with a concept several times, we no longer need to do the research. We remember what the key case is, and at most we check to see if there is anything new. We know how the cases link together, so the new materials are easy to find.

A machine learning–based research tool can do this on a much broader scale. It learns not just from our particular research efforts, but also from those of everyone who uses the system. As the system receives more use, it employs user feedback to assess how its model performs and to allow for periodic retraining. As a result, it will become extremely adept at providing immediate responses to the most common queries by users. It might also be able to eventually give you a confidence level in its answer, comparing the information it provides against the entire scope of reported decisions and its users' reactions to similar, prior responses, to let you know how reliable the results provided might be. Even though the system doesn't understand the material in the same manner as a human, its ability to track relationship building over a large scope of content and a large number of interactions allows it to behave as you might, if you had researched a particular point or issue thoroughly many times previously. This provides a research tool far more powerful than existing methodologies.

Legal tools based on machine learning have enormous application.

Legal tools based on machine learning have enormous application. Lawyers are already using learners to help with legal research, categorize document sets for discovery, evaluate pleadings and transactional documents for structural errors or ambiguity, perform large-scale document review in mergers and acquisitions, and identify contracts affected by systemic changes like Brexit. General Motors' legal department, and likely other large companies, are exploring using machine learning techniques to evaluate and predict litigation outcomes and even help choose which law firms they employ. Machine learning is not the solution for every question, but it can help answer a large number of questions that simply were not answerable in the past, and that is why the advent of machine learning in the legal profession will prove truly transformational. ◆

ARTIFICIAL INTELLIGENCE IN HEALTH CARE
APPLICATIONS AND LEGAL ISSUES

BY W. NICHOLSON PRICE II

Artificial intelligence (AI) is rapidly moving to change the healthcare system. Driven by the juxtaposition of big data and powerful machine learning techniques—terms I will explain momentarily—innovators have begun to develop tools to improve the process of clinical care, to advance medical research, and to improve efficiency. These tools rely on algorithms, programs created from healthcare data that can make predictions or recommendations. However, the algorithms themselves are often too complex for their reasoning to be understood or even stated explicitly. Such algorithms may be best described as "black-box."[1] This article briefly describes the concept of AI in medicine, including several possible applications, then considers its legal implications in four areas of law: regulation, tort, intellectual property, and privacy.

AI in Medicine

Medicine, like many other fields, is experiencing a confluence of two recent developments: the rise of big data, and the growth of sophisticated machine learning/AI techniques that can be used to find complex patterns in those data. Big data as a phenomenon is characterized by the "three Vs" of *volume* (large quantities of data), *variety* (heterogeneity in the data), and *velocity* (fast access to the data). In medicine, the data come from many sources: electronic health records, medical literature, clinical trials, insurance claims data, pharmacy records, and even information entered by patients into their smartphones or recorded on fitness trackers. Machine learning techniques, a subset of AI, use simple learning rules and iterative techniques to find and use patterns in these vast amounts of data. The resulting algorithms can make predictions and group sets—how long is a patient expected to live given his collection of symptoms, and does that picture of a patch of skin look like a benign or a cancerous lesion?—but typically, these techniques cannot explain *why* or *how* they reach the conclusion they do. Either they cannot explain it at all, or they can give explanations that are accurate but meaningless in terms of medical understanding.[2] Because of this inherent opacity (which might or might not be augmented with deliberate secrecy about how the algorithms were developed and validated), I describe this field as to "black-box medicine," though it has also been referred to as AI in medicine or "predictive analytics."[3] To add to the complexity, when more data are available for the machine learning algorithms, those data can be incorporated to refine future predictions, as well as to change the algorithms themselves. The algorithms at the heart of black-box medicine, then, are not only opaque but also likely to change over time.

Black-box medicine has tremendous potential for use throughout the healthcare system, including in prognostics, diagnostics, image analysis, resource allocation, and treatment recommendations. Machine learning is most familiar in the context of image recognition, and an algorithm has already been developed that can identify skin cancer by analyzing images of skin lesions; the algorithm performs as well as board-certified dermatologists.[4] A recent *New England Journal of Medicine* article suggests that such algorithms could soon enter widespread use in image analysis, aiding or displacing much of the work of anatomical pathologists or radiologists within the span of years.[5] Another current algorithm can predict which trauma victims are likely to hemorrhage by constantly analyzing vital signs and can in turn call for intervention to forestall catastrophe; such prognostic algorithms could come into use in a similarly short time frame.[6] A bit farther off, black-box algorithms could be used for diagnosis more generally, to recommend off-label uses for existing drugs, to allocate scarce resources to patients most likely to benefit from them, to detect fraud or problematic medical behavior, or to guide research into new diseases or conditions. In fact, black-box algorithms are already in use today in smartphone apps that aim to identify developmental disorders in infants based on facial features[7] or autism in young children based on eye movement tracking.[8] The potential for benefit from such black-box medicine is substantial, but it comes with its own challenges: scientific and medical, certainly, but also legal. How do we ensure that black-box medicine is safe and effective, how do we ensure its efficient development and deployment, and how do we protect patients and patient privacy throughout the process?

Regulation

The first question to ask is perhaps the most fundamental: How do we ensure that black-box algorithms are high quality—that is, that they do what they say, and that they do it well and safely? New and emerging medical technologies and devices are typically regulated for safety and efficacy by the Food and Drug Administration (FDA). Whether the FDA actually has statutory authority over free-standing algorithms used to make medical decisions (or to help make them) depends on the relatively complex question of what is a "medical device." The FDA's regulation

***W. Nicholson Price II**, PhD (wnp@umich.edu) is an assistant professor of law at the University of Michigan Law School. His work focuses on innovation in the life sciences, with a significant emphasis on the use of big data and artificial intelligence in health care.*

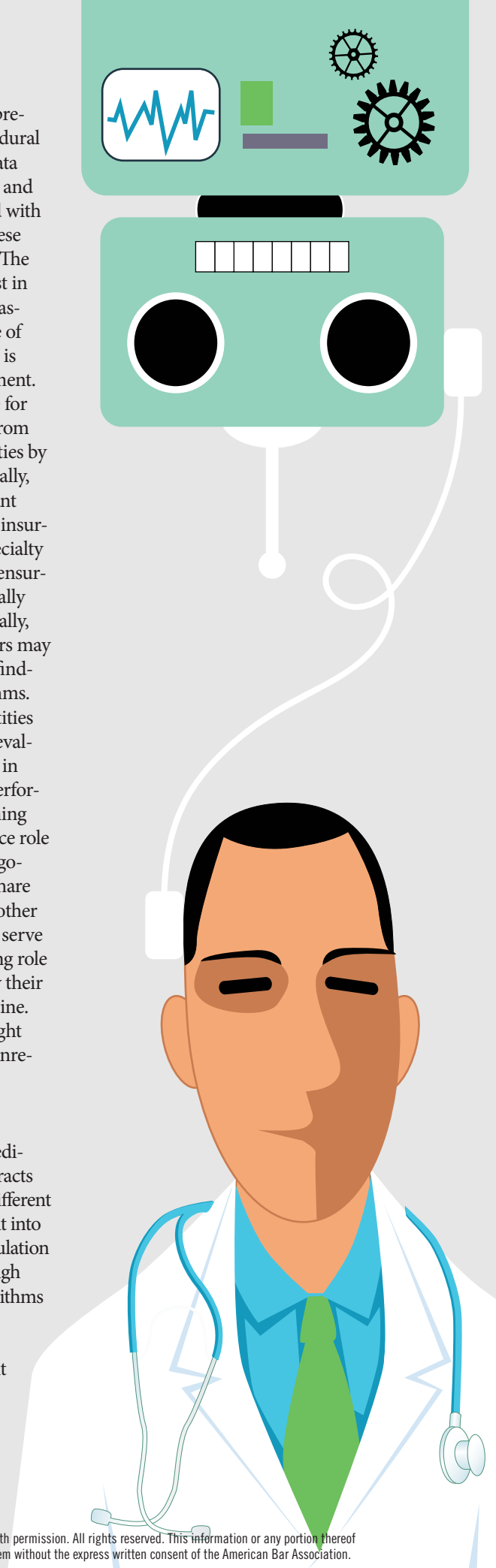
of black-box medical algorithms may also conflict with its long-standing statement that it does not regulate the practice of medicine.[9] Elsewhere, I argue that the FDA has this authority, probably over algorithms standing alone and almost certainly in the context of linked technology that may more readily be called a "medical device," but disputes may arise over this point.[10] Industry dynamics may also play a role here: Silicon Valley, the hub of much of the innovation in AI generally, traditionally has not worked closely with regulators like the FDA.

Assuming that the FDA can and will regulate AI in the healthcare system (and the agency has asserted this ability and intent),[11] typically two tools help ensure safety and efficacy of new medical technology: scientific understanding and clinical trials. Unfortunately, these two tools do not work well in the context of black-box medicine. Understanding does not work for obvious reasons—we do not understand how a black-box algorithm makes decisions, because the machine learning techniques generally cannot tell us their reasoning, and even when they can, the results are often too complex to understand. Using clinical trials for testing safety, efficacy, and validity might work for some algorithms, but will not work for many others. For algorithms that divide patients into groups and suggest a particular treatment, clinical trials could be used to test their efficacy. But some algorithms will make highly personalized treatment predictions or recommendations, so that the use of clinical trials would be infeasible. And even for algorithms that are amenable to trials, the benefits of black-box medicine—quick, cheap shortcuts to otherwise inaccessible medical knowledge—would be seriously delayed or even curtailed due to the slow, ponderous, expensive enterprise of clinical trials. For algorithms that change as they incorporate more data, the challenges are even more pronounced. In short, in black-box medicine, traditional methods of testing new medical technologies and devices are likely not to work at all in some instances, and to slow or stifle innovation in others.

So how should the FDA tackle this challenge? The most fruitful path, I argue, will likely be more flexible than rigid, involving somewhat lighter premarket scrutiny (focused on procedural safeguards like the quality of the data used, the development techniques, and the validation procedures) coupled with robust post-market oversight as these algorithms enter into clinical care. The FDA has recently expressed interest in this approach.[12] Of course, this is easier said than done; the parallel case of post-market surveillance for drugs is notoriously troublesome to implement. One attractive possibility would be for the FDA to enable oversight help from other sophisticated healthcare entities by collaborating with them and, crucially, enabling ways to get them important and useful information. Hospitals, insurance companies, and physician specialty associations all have an interest in ensuring that black-box algorithms actually work to help patients (and, potentially, their bottom lines). Rival developers may also have an interest, especially in finding problems with existing algorithms. In addition, these sophisticated entities may have the capacity to perform evaluations, especially as they are used in clinical practice, and to generate performance data. Nevertheless, performing this type of collaborative governance role requires information, and many algorithm developers are reluctant to share that kind of information with any other entities. Potentially the FDA could serve as a centralized information-sharing role to allow those other entities to play their part in regulating black-box medicine. However, exactly how this idea might become a reality is very much an unresolved question.

Tort

What do we do when black-box medicine goes awry? The law of tort interacts with black-box medicine in a few different contexts. First, if there are flaws built into the algorithms themselves, or if regulation fails to ensure that algorithms are high quality, then the developers of algorithms (or technologies that rely on them) might become liable under tort law. However, courts have been reluctant to extend or apply product liability theories to software developers, and even more reluctant in the context of healthcare software.[13]



Part of that reluctance has come from the fact that healthcare software to date has been characterized primarily as technology that helps healthcare providers make decisions by providing them with information or analysis, with the final decision always resting in the hands of the provider. Black-box medicine turns that notion on its head, or at least it can. Can and should healthcare providers be fully responsible for decisions suggested or made by black-box algorithms that they do not, or cannot, understand?

This raises a second set of questions. What must healthcare providers and healthcare institutions—doctors, nurses, hospitals, managed-care organizations, and the like—do to fulfill their duties of care to patients in a healthcare world with black-box algorithms? Must providers themselves evaluate the quality of black-box algorithms, based on procedural measures (validation undertaken, performance statistics, etc.) before relying on those algorithms in the course of providing care? And should healthcare institutions perform similar evaluations before implementing black-box software? I argue elsewhere that they should, but currently the information necessary for that type of evaluation is largely unavailable—just as in the parallel regulatory context mentioned above.[14] Similarly, if an algorithm suggests an intervention that seems mundane but unhelpful, useless and expensive, or dangerous, should the provider second-guess the recommendation? On the one hand, the answer seems an obvious "yes"—providers are trained to care for patients—but on the other hand, if providers only implement those decisions they would have reached on their own, they will leave on the table much of the benefit that black-box medicine promises to extract from otherwise inaccessible patterns in big data. This would not leave *everything* on the table—algorithms can still potentially perform the usual analyses more quickly and cheaply[15]—but excessive caution is not costless. Courts have not tackled these issues yet, but they will need to in the near future.

Intellectual Property

Intellectual property protection creates another set of challenges for the development of black-box medicine.[16] When firms invest in developing black-box algorithms, how can they protect that investment? Developing black-box algorithms can involve considerable expense. Developers must generate, assemble, or acquire the tremendous data sets needed to train their algorithms; they must assemble the expertise and resources to actually develop those algorithms; and they must validate them to make sure they work. Normally, we might expect intellectual property to provide some measure of protection for the information goods created by such expenditures, so that firms are willing to invest the necessary funds for their development without fear that resulting inventions will be appropriated by others.[17] However, intellectual property fits relatively poorly for black-box medicine.

Patents are a natural choice to protect technological innovation, but patents do not provide strong incentives for black-box medicine. A string of recent decisions by the U.S. Supreme Court interpreting section 101 of the Patent Act, which governs patentable subject matter, has made it very difficult to patent black-box algorithms.[18] In *Mayo Collaborative Services v. Prometheus Laboratories, Inc.*, the Supreme Court repeated its long-standing statement that laws of nature cannot be patented.[19] However, the Court applied that rule to a diagnostic test that used the measurement of a metabolite level in a patient's blood to adjust the dosage of a drug, which many, including the Federal Circuit below, had thought to be a patentable application of such a law. The Supreme Court used very broad language to invalidate the patent: "[W]ell-understood, routine, conventional activity previously engaged in by scientists who work in the field . . . is normally not sufficient to transform an unpatentable law of nature into a patent-eligible application of such a law."[20] Where underlying information about the biological world is the heart of the invention, merely using that information to guide medical treatment is unpatentable (as is the information itself). But this describes most black-box algorithms quite well and suggests that those algorithms are unlikely to be patentable subject matter. Further patent problems might arise under section 112, which requires a "written description" of the invention. Although this issue has not been tested in the courts, it is at least debatable how well one can describe an algorithm that is opaque, and how broad the resulting protection would be.[21]

Trade secrecy—or secrecy in general—seems an obvious solution but comes with its own problems. Trade secret law protects from appropriation information that is kept secret and gets commercial value from its secrecy. What better way than secrecy to protect an algorithm that is already opaque and cannot be understood? The data on which an algorithm is generated, the method by which the algorithm was developed, and the process of its validation can all be kept secret by firms looking to protect their investment in the algorithm's development. And indeed, firms that are developing black-box algorithms seem to be relying on just such secrecy. But while secrecy may be an effective intellectual property strategy, it runs headlong into the concerns raised above about safety, malpractice, and regulation. How willing will doctors, patients, and insurers be to accept medical algorithms where not only is the working of the algorithm a mystery, but also the way the algorithm was made and tested, along with the data underlying its development? And if third parties are indeed to be actively involved in ensuring algorithmic quality and validity, as I suggest above, how can they conduct such evaluations without the underlying information? The reliance of algorithm developers on trade secrecy echoes other past situations where information relevant to public health has been kept secret, and these experiences suggest that there may be similar fights over access to algorithmic information.[22]

However, if intellectual property incentives are unavailable to help protect investments in black-box medicine, will firms invest sufficiently? How can the government help drive this form of innovation while ensuring that it is safe and effective? These questions are and will remain pressing for the development of AI in health care.

Privacy

Finally, privacy concerns run through the development and deployment of black-box medicine.[23] Privacy is important in at least two areas: gathering immense amounts of healthcare data to develop

algorithms, and sharing such data to oversee them. Algorithm developers need to assemble data from multiple sources to train machine learning algorithms. Those data—as well as data about how the algorithms perform in practice—may then be shared with other entities in the healthcare system for the purpose of evaluation and validation, as described above. In each case, patient-oriented data privacy is a concern, most notably as mandated under the Health Insurance Portability and Accountability Act's (HIPAA's) Privacy Rule. The Privacy Rule governs and restricts both disclosure and use of "protected health information" (that is, most individually identifiable health information) by "covered entities" (mostly, healthcare providers, health insurers, health information clearinghouses, and business associates of the same).[24] HIPAA creates a relatively complex set of permitted and restricted uses of protected health information. Notably, de-identified information is not governed by the Privacy Rule (though it raises its own concerns about data aggregation and the possibility of re-identification), and neither is information collected by noncovered entities like Google, Apple, or other aggregators of big data.[25] Navigating the HIPAA Privacy Rule—and otherwise managing and addressing the privacy concerns of those whose data is used throughout black-box medicine—creates yet another ongoing set of potential legal concerns.

Conclusion

Black-box medicine has tremendous potential to reshape health care, and it is moving rapidly to do so. Some healthcare black-box algorithms are already at work in consumer-directed smartphone apps, and others are likely to enter medical practice in the span of years. But the legal issues involved with the development and implementation of AI algorithms, which we do not and cannot understand, are substantial. As described here, regulation, legal causes of action such as medical malpractice and product liability, intellectual property, and patient privacy all have real implications for the way black-box medicine is developed and deployed. In turn, black-box medicine may change the way we approach some of these issues in the context of contemporary health care. Does entity-centered privacy regulation make sense in a world where giant data agglomerations are necessary and useful? Should intellectual property law find new ways to recognize the primacy of health data and the fast-moving nature of algorithms? Must the legal doctrine of the "learned intermediary" bow to the recognition that doctors cannot fully understand all the technologies they use or the choices such technologies help them make when they are not provided the needed and/or necessary information? Should the FDA change how it regulates new medical technology as AI software gains prominence? As black-box medicine develops and evolves, the need to consider these legal issues—and the need for scientifically literate lawyers who can understand them in context—will continue to grow. ◆

Endnotes

1. W. Nicholson Price II, *Black-Box Medicine*, 28 Harv. J.L. & Tech. 419 (2015).

2. Jenna Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, 3 Big Data & Soc'y 1, 5 (2016).

3. I. Glenn Cohen et al., *The Legal and Ethical Concerns That Arise from Using Complex Predictive Analytics in Health Care*, 33 Health Aff. 1139 (2014).

4. Andre Esteva et al., *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*, 542 Nature 115 (2017).

5. Ziad Obermeyer & Ezekiel J. Emanuel, *Predicting the Future—Big Data, Machine Learning, and Clinical Medicine*, 375 New Eng. J. Med. 1216 (2016).

6. Nehemiah T. Liu et al., *Development and Validation of a Machine Learning Algorithm and Hybrid System to Predict the Need for Life-Saving Interventions in Trauma Patients*, 52 Med. & Biological Engineering & Computing 193 (2014).

7. Megan Molteni, *Thanks to AI, Computers Can Now See Your Health Problems*, Wired (Jan. 9, 2017), https://www.wired.com/2017/01/computers-can-tell-glance-youve-got-genetic-disorders/.

8. *Autism*, Right Eye, https://www.righteye.com/tests-therapies/autism (last visited Oct. 17, 2017).

9. Patricia J. Zettler, *Toward Coherent Federal Oversight of Medicine*, 52 San Diego L. Rev. 427 (2015).

10. W. Nicholson Price II, *Regulating Black-Box Medicine*, 91 Mich. L. Rev. (forthcoming 2017), https://papers.ssrn.com/abstract_id=2938391.

11. *See* U.S. Food & Drug Admin., Medical Device Accessories—Describing Accessories and Classification Pathway for New Accessory Types: Guidance for Industry and Food and Drug Administration Staff (Jan. 30, 2017).

12. Press Release, FDA, FDA Selects Participants for New Digital Health Software Precertification Pilot Program (Sept. 26, 2017), https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm577480.htm.

13. Randolph A. Miller & Sarah M. Miller, *Legal and Regulatory Issues Related to the Use of Clinical Software in Health Care Delivery*, *in* Clinical Decision Support: The Road Ahead 423, 426 (Robert A. Greenes ed., 2007).

14. W. Nicholson Price II, *Medical Malpractice and Black-Box Medicine*, *in* Big Data, Health Law, and Bioethics (I. Glenn Cohen et al. eds., forthcoming 2018).

15. Megan Molteni, *If You Look at X-Rays or Moles for a Living, AI Is Coming for Your Job*, Wired (Jan. 25, 2017), https://www.wired.com/2017/01/look-x-rays-moles-living-ai-coming-job/.

16. W. Nicholson Price II, *Big Data, Patents, and the Future of Medicine*, 37 Cardozo L. Rev. 1401 (2016).

17. Mark A. Lemley, *Ex Ante versus Ex Post Justifications for Intellectual Property*, 71 U. Chi. L. Rev. 129 (2004).

18. Rebecca S. Eisenberg, *Diagnostics Need Not Apply*, 21 B.U. J. Sci. & Tech. L. 256 (2015).

19. 132 S. Ct. 1289 (2012).

20. *Id.* at 1298.

21. W. Nicholson Price II, *Describing Black-Box Medicine*, 21 B.U. J. Sci. & Tech. L. 347 (2015).

22. David S. Levine, *The People's Trade Secrets?*, 18 Mich. Telecomm. & Tech. L. Rev. 61 (2011).

23. Roger Allan Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 Mich. Telecomm. & Tech. L. Rev. 1 (2016).

24. 45 C.F.R. pts. 160, 164.

25. Nicolas Terry, *Big Data and Regulatory Arbitrage in Health Care*, *in* Big Data, Health Law, and Bioethics, *supra* note 14.

AI AND MEDICINE

HOW FAST WILL ADAPTATION OCCUR?

BY MATTHEW HENSHON

Artificial intelligence (AI) burst onto the popular scene in 2011, when IBM's Watson defeated two human champions (including all-time leader Ken Jennings) in a nationally televised two-part exhibition of *Jeopardy!*, the TV game show.[1] A previous Watson iteration (Deep Blue) had defeated then world champion Garry Kasparov in chess in 1997, but the game of chess perhaps seemed a simpler task for machines: a defined board, and 16 pieces on each side.[2] In contrast, the range of *Jeopardy!* clues (remember, as Alex Trebek reminds viewers regularly, to "phrase your response in the form of a question") is seemingly limitless, and the clues are often in the form of puns or slang, so a chess move like "rook to D1" in comparison seems simple.

To its credit, IBM has built an entire marketing campaign around Watson's victory on *Jeopardy!* But the methodology that enabled Watson to excel in the quiz show was perhaps unique: the machine first tries to identify a keyword in the clue, then compares that word against its (then) database of 15-terabytes of information. In Ken Jennings's words:

> It rigorously checks the top hits against all the contextual information it can muster: the category name; the kind of answer being sought; the time, place, and gender hinted at in the clue; and so on. And when it feels "sure" enough, it decides to buzz. This is all an instant, intuitive process for a human *Jeopardy!* player, but I felt convinced that under the hood my brain was doing more or less the same thing.[3]

Applying Watson to fields like medicine has been a bit rougher. IBM signed a high-profile partnership with the University of Texas MD Anderson Cancer Center in Houston in 2012, declaring it a "moon shot" to cure cancer in a press release.[4] But five years later, with progress significantly slower than initially anticipated, MD Anderson and IBM have parted ways. The university, which paid IBM a total of $39 million on a contract originally negotiated for less than 10 percent of that amount, had nothing to show for its money, except a cancer-screening tool that was still in the "pilot" stage.[5]

The problem for Watson and medicine may be related to its success in *Jeopardy!* In the game show, the correct answers are "known," so Watson sifts through data and tries to find the right one. And if the machine does not pick a winner, it can adjust its algorithm (so-called "machine learning"). But in medicine, it is perhaps harder to find the single correct answer. Computers excel at working with "structured data," such as billing codes or lab test results; but sometimes human medical judgment and doctor's notes are just as important in making a diagnosis, and those are much harder for a computer to analyze.[6]

But while IBM's Watson healthcare efforts appear (for the moment) to be retrenching,[7] other players are aggressively entering into the medical AI market. In 2014, Google acquired London-based DeepMind, for $400 million.[8] Among other research projects, DeepMind has developed a program that plays Go (the Asian board game that is more complicated than chess) and has begun to regularly beat the best players in the world, even when five Go champions combined their efforts to try and defeat the program!

Like Watson, Google's DeepMind is attempting to apply its technology to health care: last November, it announced a partnership with a London hospital system.[9] But DeepMind's Streams app appears to be built around

Matthew Henshon *(mhenshon@henshon.com) is a partner at the Boston boutique law firm of Henshon Klein LLP. He is chair of the Artificial Intelligence and Robotics Committee. Follow him on Twitter at @mhenshon.*

much more rudimentary AI, and its primary benefit at this point appears to be streamlining the process of notification for blood tests indicating acute kidney injuries (AKIs).[10] While AKI is one of the leading causes of death in the National Health Service (NHS), the "special sauce" at this point appears to be simply routing abnormal blood test results to the appropriate doctor's mobile device.

The lesson learned may be that applying AI to real-world problems requires small steps that can supplement and enhance—rather than replace—human decision making. Streams is not attempting to replace doctors and specialists at this point—merely get them key information faster. Another factor is that the move to full electronic health records began only about 10 years ago, and is still in process; AI will get better as it has more data to evaluate. One site that is currently analyzing healthcare data is Modernizing Medicine, which uses a tablet and data provided from 3,700 doctors on over 14 million patient visits to recommend treatments or drugs based on symptoms, much like Netflix suggesting a new movie.[11]

We also may have to revise our view of what AI will do: the apparent early promise of Watson was in finding a single "cure for cancer." But a more promising side of AI may be in simply helping patients manage their own conditions. For instance, type 2 diabetes can often be managed—and in some cases reversed—by controlling the patient's diet and lifestyle. The problem is that such control requires extensive oversight, from a seemingly full-time doctor in the home. But with smartphones and home monitoring devices like Fitbit, the patient can provide such information in real-time, to be integrated into a larger database. The doctor can then quickly assess the changing conditions of the patient. Preliminary testing of an app-based system by one company (Virta Health) has shown that 87 percent of the type 2 diabetic patients in the study reduced their insulin dose or eliminated it outright.[12]

Medical care is not the only arena that AI hopes to move into: games like chess and Go are supposed to be a "test bed" (to use the industry term) for legal work, crime prevention, and business negotiations, among others. But as one IBM AI researcher said, "There are precious few zero-sum, perfect-information, two-player games that we compete in in the real world."[13]

The pace of adaptation in medicine relates to the nature of the test-bed AI itself: namely, the gaming world. There's little real-world consequence (other than to Garry Kasparov himself) in a chess game. There are no lives at risk, and a mistake might lead to the early loss of a rook. But in medicine, and other real-world events, mistakes have consequences. A missed AKI marker by DeepMind means the life of a real patient is potentially at risk. Thus, there is a natural tendency to be conservative with AI algorithms: the cost of a false positive (the equivalent of a false alarm) is low; the cost of a false negative can be catastrophic.

AI will continue to progress with each advance in semiconductors; note the computing power in your smartphone is more than that of Deep Blue 20 years ago. But getting to the next stage, where we rely on AI to make judgments on life-and-death decisions, may take longer than we currently anticipate. The incremental steps shown by Streams, Virta Health, Modernizing Medicine, and others may be more promising—and more successful in the short to medium term—than a "moon shot." ◆

Endnotes

1. Ken Jennings, *My Puny Human Brain*, Slate (Feb. 16, 2011), http://www.slate.com/articles/arts/culturebox/2011/02/my_puny_human_brain.html.

2. *Kasparov vs. Deep Blue*, NPR (Aug. 8, 2014), http://www.npr.org/2014/08/08/338850323/kasparov-vs-deep-blue.

3. Jennings, *supra* note 1.

4. Press Release, IBM, MD Anderson Taps IBM Watson to Power "Moon Shots" Mission Aimed at Ending Cancer, Starting with Leukemia (Oct. 18, 2013), https://www-03.ibm.com/press/us/en/pressrelease/42214.wss.

5. David H. Freedman, *A Reality Check for IBM's AI Ambitions*, MIT Tech. Rev. (June 27, 2017), https://www.technologyreview.com/s/607965/a-reality-check-for-ibms-ai-ambitions/.

6. Daniela Hernandez, *Artificial Intelligence Is Now Telling Doctors How to Treat You*, Wired (June 2, 2014), https://www.wired.com/2014/06/ai-healthcare/.

7. Indeed, even IBM's more recent press releases seem more modest: "[IBM's Watson will be] collaborating with more than a dozen leading cancer institutes to accelerate the ability of clinicians to identify and personalize treatment options for their patients." Press Release, IBM, Clinicians Tap Watson to Accelerate DNA Analysis and Inform Personalized Treatment Options for Patients (May 5, 2015), https://www-03.ibm.com/press/us/en/pressrelease/46748.wss. The institutes include Ann & Robert H. Lurie Children's Hospital of Chicago; BC Cancer Agency; City of Hope; Cleveland Clinic; Duke Cancer Institute; Fred & Pamela Buffett Cancer Center in Omaha, Nebraska; McDonnell Genome Institute at Washington University in St. Louis; New York Genome Center; Sanford Health; University of Kansas Cancer Center; University of North Carolina Lineberger Comprehensive Cancer Center; University of Southern California Center for Applied Molecular Medicine; University of Washington Medical Center; and Yale Cancer Center.

8. Oliver Roeder, *The Bots Beat Us. Now What?*, FiveThirtyEight (July 10, 2017), https://fivethirtyeight.com/features/the-bots-beat-us-now-what/.

9. Mustafa Suleyman, *A Milestone for DeepMind Health and Streams*, DeepMind (Feb. 27, 2017), https://deepmind.com/blog/milestone-deepmind-health-and-streams/.

10. *Streams in NHS Hospitals*, DeepMind, https://deepmind.com/applied/deepmind-health/working-nhs/how-were-helping-today/ (last visited Oct. 17, 2017).

11. Hernandez, *supra* note 6.

12. Kevin Maney, *How Artificial Intelligence Will Cure America's Sick Health Care System*, Newsweek (May 24, 2017), http://www.newsweek.com/2017/06/02/ai-cure-america-sick-health-care-system-614583.html.

13. Roeder, *supra* note 8.

PRENATAL GENETIC TESTING: WHERE ALGORITHMS MAY FAIL

On the forefront of the much anticipated arrival of genomic medicine, the field of noninvasive prenatal genetic testing (NIPT) has lived up to its expectations as a game changer. NIPT is estimated to be worth $500 million in 2013, with potential to grow to 2.38 billion by 2022.[1] NIPT not only has revolutionized access to prenatal testing for genetic disorders through a mere blood test, but it also is unique enough to defy classification by both intellectual property law and the existing Food and Drug Administration (FDA) regulatory framework.

Much attention has been given to the Federal Circuit's 2015 decision in *Ariosa Diagnostics, Inc. v. Sequenom, Inc.*,[2] invalidating the patent that protected Sequenom's commercially offered Maternit21® test as non-patent-eligible subject matter under the Supreme Court's *Mayo v. Prometheus* decision.[3] Less discussed is the technology itself, a screening test for a genetic condition, which relies significantly on "next-gen" DNA sequencing and proprietary algorithms to translate massive amounts of data into clinical results. The NIPT technology is described below before turning to a discussion of the law.

Algorithms and NIPT

Advancements in computational genomics, which focuses on developing probability models to interpret the data generated by DNA sequencing, have allowed genetic testing to scale upwards. The next-gen sequencing technology that powers these tests is capable of generating billions of DNA base pair reads from a single sample. This vast quantity of data must be reduced to an interpretable form.

Using this sequencing technology and the accompanying algorithms, genetic testing can be done by isolating fragments of fetal and maternal DNA present in a pregnant woman's blood and then amplifying those fragments. Through "amplification," scientists produce millions of copies of the relatively small amount of DNA found in the maternal blood serum to create a sample size large enough for the sequencing algorithms. Of the total DNA found in the maternal blood serum, approximately 3.4–6.2 percent is fetal, with the rest belonging to the mother.[4] This DNA is often referred to as "cellular free" because it is not found in the cell nuclei and originates from cells that die naturally, leaving DNA fragments in the expectant mother's bloodstream.

The most common prenatal genetic tests are for trisomy disorders, such as Down syndrome (trisomy 21), Edwards syndrome (trisomy 18), and Patau syndrome (trisomy 13). The latter two conditions can lead to early miscarriage and significant rates of clinical morbidity or mortality following birth.[5]

Typically, humans have 23 pairs of chromosomes, with each parent contributing a single chromosome to the pair. Trisomy occurs when an extra chromosome or fragment thereof is associated with the typical pair, resulting in three chromosomes where there should only be two. A trisomy on the 21st chromosome is referred to as "trisomy 21." Individual chromosomes themselves are made up of DNA "base pairs," molecules that occur in tandem on opposing strands of DNA, and are abbreviated by the letters A (adenine), T (thymine), C (cytosine), and G (guanine), so as to form what is referred to as the "genetic code." DNA is "sequenced" to determine the number of base pairs present and if those base pairs correlate to a specific gene, and thus a genetic condition.

Five companies (Sequenom, Illumina, LabCorp, Ariosa, and Natera) currently offer prenatal tests for trisomy conditions, each relying on different variations of the same sequencing methods. Though these methods all involve the same basic steps of isolating, amplifying, and sequencing DNA, each company applies those steps in different and proprietary ways. This matters because each test can offer an incorrect screening result due to errors inherent to individual testing methods and their underlying statistical assumptions.

Sequenom, Illumina, and LabCorp use a method called "massively parallel shotgun sequencing," while Ariosa and Natera use "targeted sequencing." Shotgun sequencing amplifies the entirety of the genetic material collected, regardless of chromosome number, and then sorts the sequenced genetic content according to

Aubrey Haddach *(ahaddach@dinsmore.com) is co-chair of the Section's Biotechnology Law Committee and a member of the Intellectual Property Department at Dinsmore & Shohl LLP.* ***Jeff Licitra*** *(jlicitra@licitralegal.com) is a recent graduate of the American University Washington College of Law and a member of the District of Columbia Bar.*

previously known and expected results for the human genome. Targeted sequencing, as the name implies, amplifies only a region or set of genes on the desired chromosome.

Each company then uses its own proprietary algorithms of "quantitative read counting" to count the total number of base pairs sequenced for the chromosome region being examined. One sample alone undergoing shotgun sequencing may generate, on average, 10,800,000 base pair reads, as opposed to 32,000 for targeting sequencing.[6] If the algorithm finds too many base pairs associated with a particular chromosome, then the statistical inference is that the extra DNA base pairs indicate a trisomy condition on that chromosome.

What distinguishes prenatal genetic screening from other types of genetic testing is its reliance on the tiny amount of fetal DNA fragments present in the maternal serum. That is to say, NIPT does not actually sequence the entire fetal genome because it only has fragments to work with. It is not so much reading the code as it is counting it. In this way, any amount of extra DNA present in the base pair count, even if that extra DNA is not associated with an extra chromosome, can lead to a false-positive result.

One such false-positive case involved a pregnant woman who had a copy number variation. A copy number variation occurs when there is more DNA in certain chromosome regions due to the variance in length of certain DNA strands. The nonstandard length of maternal DNA with a copy number variation, particularly when inherited in fetal DNA, and thus doubly represented in the sample, could yield extra DNA in the test result, and thus a false-positive reading.[7] Interestingly, both Ariosa and Sequenom replied to this case separately, not so much to refute its findings but to explain that they each compensate for copy variance differently in their respective algorithms.[8]

The Validity of Screening Results

While copy variance is certainly not the only factor that can affect false-positive results for NIPTs, it is instructive because it shows how indispensable underlying statistical assumptions are to the test results. Maternal or confined placental mosaicism,[9] vanishing twin pregnancies, and maternal malignancy may all yield false-positive results for trisomy because each of these conditions allows for extra DNA in the maternal blood serum.[10]

Various journalists have reported on the anguish experienced by parents who relied on false-positive results and unfortunately terminated the pregnancies as a result or who otherwise awaited (or planned) for the arrival of a child with a trisomy condition.[11] Yet peer-reviewed journals (and the companies themselves) continue to conclude that NIPTs are fundamentally sound in their ability to make true-positive predictions at a greater than 99 percent rate for trisomy 21.[12]

A review of results for trisomy 18 and 13 studies showed a lower true-positive rate (the rate at which tests are both positive and correct) of 97–99 percent and 92–95 percent, respectively.[13] A meta-analysis of all studies estimated the true-positive rates for trisomy 18 and trisomy 13 to be approximately 95 percent when performed in combination with the trisomy 21 test, but possibly lower when performed alone.[14] Moreover, the vast majority of studies have been done on women at a high risk for trisomy, increasing the likelihood of a correct positive test result.[15]

As the FDA considers NIPTs to be laboratory diagnostic tests (LDTs), it does not regulate prenatal genetic testing beyond a policy of "enforcement discretion"—meaning the FDA chooses whether or not to oversee pre-market testing or impose post-market safety requirements.[16] This is because the tests are developed and used entirely in one location, regardless of where the samples originate. LDTs were traditionally used by hospitals billing directly to Medicare. Their "home brew" labs have long been regulated for scientific accuracy and precision by the Clinical Laboratory Improvement Amendments (CLIA) statute.[17]

Notably, none of this oversight reaches the level of clinical validity, and, perhaps more importantly, adverse results are not required to be reported to the FDA. Contrastingly, clinical results for drugs and diagnostic tests otherwise falling under the FDA's purview are subject to extensive clinical data review for marketing approval and then continuous adverse event reporting after going to market. This allows an analysis of all adverse events occurring to a far more robust population than could ever be constructed in a clinical study.

This lack of regulatory oversight for NIPTs leaves patients and medical professionals in a position to interpret the results of tests that often influence a woman's decision to terminate pregnancy. Although NIPTs are screening tests, intended as an intermediate step before a more invasive diagnostic procedure, patients may not appreciate the nuance of a 10 mL blood sample yielding a statistical guess, even if an incredibly accurate one, at the presence of a genetic condition. Similarly, the nature of NIPTs as screening tests has left the courts to grapple with the importance of NIPT as a breakthrough technology and its place within the larger context of patent jurisprudence.

Ariosa, *Mayo*, and Section 101 Patent Eligibility

There are two central tenets of patent law that determine whether an invention such as NIPT is patent-eligible subject matter under 35 U.S.C. section 101. On one hand is the now infamous expression that Congress intended statutory subject matter to

"include anything under the sun that is made by man."[18] On the other is the "natural law exception," which holds that naturally occurring phenomena are not patentable.[19] The problem arises when a patent on a man-made invention—even a ground-breaking one such as NIPT—relies so heavily on a scientific discovery that a court finds the patent invalid for only claiming a natural law.

Sequenom obtained the first patent on NIPT in 2005, when it purchased from Isis Innovation Ltd., the commercial research arm of Oxford University, an exclusive license to a patent for "amplifying" and "detecting" cellular-free fetal DNA (cffDNA).[20] In 2011, Sequenom became the first company to offer prenatal genetic testing. Only months later, Sequenom entered into litigation in the Northern District of California with Ariosa, Natera, and Verinata over infringement of its patent. If the patent was valid, Sequenom would have exclusive control of the nascent NIPT market. If invalid, its competitor companies would be able to enter the market without risking infringement.

The task before the district court was to apply the Supreme Court's two-part *Mayo* test to determine whether the patent was invalid for lack of eligible subject matter: first, decide whether the method claimed is to a natural law or abstract idea; if yes, proceed to step two, and determine if there is an "inventive concept" sufficient to overcome the patent's reliance on an abstract idea.[21] In October 2013, the district court held the patent invalid as ineligible subject matter, concluding "the only inventive concept contained in the patent to be the discovery of cffDNA, which is not patentable."[22]

Sequenom appealed only to have the Federal Circuit affirm the patent's invalidation and deny rehearing en banc.[23] In June 2015, the Supreme Court declined Sequenom's petition for certiorari, putting an end to its five-year quest to maintain the patent in the face of noninfringement suits brought by its competitors. Along the way, members of the legal community and biotech industry filed numerous briefs imploring the Federal Circuit to avoid a ruling that would nullify patent protection for future genetic screening technology.

The *Mayo* decision itself was comparatively "low-tech" and involved a method for administering an intravenous drug to a patient, where the dosage of the drug was increased or decreased to obtain an efficacy level based on measuring metabolites in the bloodstream.[24] Calculations and dosage adjustments are decisively abstract compared to the multitude of tangible steps involved in noninvasive testing, from the separation of cffDNA in the blood sample to the next-gen sequencing and algorithmic determination of a clinical result. Do these steps constitute a sufficient "inventive concept" to transform the abstract science to patentable subject matter? The *Mayo* framework may ultimately be inadequate if it confuses patent-ineligible abstract methods with the several, complex, laboratory steps used in molecular diagnostics.

The Federal Circuit applied *Mayo* to conclude that (1) Sequenom was merely utilizing a natural law, i.e., the presence of fragmented fetal DNA in the maternal bloodstream; and (2) the patent lacked sufficient inventive concept because its claim involved the standard DNA sequencing steps of amplification and detection routinely practiced by scientists.[25] However, the claim may have been written so broadly as to render these methods abstract regardless of the patentability of NIPT itself. Indeed, Judge Lourie acknowledged as much in his concurrence, denying rehearing en banc, positing that the patent may have been invalid for lack of specificity regardless of *Mayo*, and that "the finer filter of § 112 might be better suited to treating these as questions of patentability."[26]

The judges, even in agreeing on the result, expressed misgivings about the breadth of the "natural laws" restriction imposed by *Mayo*. Judge Lourie acknowledged that the holding of *Mayo* had been correctly applied as binding Supreme Court precedent, but he found the rule "unsound" insofar as it "takes inventions of this nature out of the realm of patent-eligibility."[27] Similarly, Judge Dyk, in his own concurrence, observed that "the major defect is not that the claims lack inventive concept but rather that they are overbroad."[28] Judge Linn, concurring in the panel decision, explained he did so "bound by the sweeping language of the test" set out by the Supreme Court, noting that "the amplification and detection of cffDNA had never before been done."[29] Judge Newman, in her dissent to the denial of rehearing, viewed Sequenom's patent to be patentable subject matter. In her opinion, the patent at issue involved the "discovery and development of a new diagnostic method" for cffDNA, itself a discovery, in contrast to a situation where "both the medicinal product and its metabolites were previously known, leaving sparse room for innovative advance."[30] In other words, the subject matter of the process patent was neither an abstract process nor natural law, and in failing that, there was no need to look for an "inventive concept."

Much of the underlying next-gen sequencing technology is already patented or proprietary subject matter. If anything, it is the use of NIPT algorithms themselves to connect the next-gen sequencing technology to a meaningful screening result that may be the "inventive concept."

Even if the *Ariosa* decision stands, the patent at issue may not have been representative of the potential patentable subject matter in NIPT. Presently, litigation among these companies continues in federal

court and at the Patent Trial and Appeal Board to decide whether the patent is invalid on other bases.[31]

Moving Forward

Noninvasive prenatal genetic testing involves technology that at present resists classification under FDA regulations and confounds the intuitive notion that ground-breaking inventions deserve patent protection. This is in part because the technology itself is so much more complex than what the legal profession is accustomed to. NIPT is not just a traditional laboratory diagnostic test, but a screening exam for a genetic condition with clinical implications. Similarly, it is not just the application of a natural law, in the discovery of cffDNA, but a novel way of applying next-gen sequencing technology to that discovery.

The public has an interest both in seeing the biotech industry continue to innovate and in receiving the benefits of more robust testing and oversight. The two interests are not contradictory, and a better understanding of the technology that makes these tests possible should in turn lead to better laws and patient outcomes. ◆

Endnotes

1. Press Release, Transparency Mkt. Research, Non-Invasive Prenatal Testing (NIPT) Market (BambniTest, Harmony, informaSeq, MaterniT21 PLUS, NIFTY, Panorama, PrenaTest, verifi, VisibiliT and Others)—Global Industry Analysis, Size, Volume, Share, Growth, Trends and Forecast 2014–2022 (May 21, 2015), http://www.transparencymarketresearch.com/noninvasive-prenatal-diagnostics-market.html.

2. 788 F.3d 1371 (Fed. Cir.), *reh'g en banc denied*, 809 F.3d 1282 (Fed. Cir. 2015), *cert. denied*, 136 S. Ct. 2511 (2016).

3. *Mayo Collaborative Servs. v. Prometheus Labs., Inc.*, 132 S. Ct. 1289 (2012).

4. Y.M. Dennis Lo, *Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis*, 62 Am. J. Hum. Genetics 768 (1998); *see also* Yuk Ming Dennis Lo, *Non-Invasive Prenatal Diagnosis by Massively Parallel Sequencing of Maternal Plasma DNA*, 2 Open Biology, June 2012 (reporting that fetal DNA concentrations have been found as high as 10 percent using next-gen, digital polymerase chain reaction (PCR)).

5. Errol R. Norwitz et al., *Noninvasive Prenatal Testing: The Future Is Now*, 6 Revs. in Obstetrics & Gynecology 48, 49 (2013).

6. Andrew B. Sparks et al., *Selective Analysis of Cell-Free DNA in Maternal Blood for Evaluation of Fetal Trisomy*, 32 Prenatal Diagnosis 3, 4 (2012).

7. Matthew W. Snyder et al., *Copy-Number Variation and False Positive Prenatal Aneuploidy Screening Results*, 372 New Eng. J. Med. 1639 (2015).

8. *Correspondence: Copy-Number Variation and False Positive Prenatal Aneuploidy Screening Results*, 373 New Eng. J. Med. 2583 (2015).

9. Mosaicism occurs when cells carry more than one genotype, allowing for more DNA than normal.

10. Zandra C. Deans et al., *Recommended Practice for Laboratory Reporting of Non-Invasive Prenatal Testing of Trisomies 13, 18, and 21: A Consensus Opinion*, 37 Prenatal Diagnosis 699 (2017).

11. *See, e.g.*, Beth Daley, *Oversold and Misunderstood: Prenatal Screening Tests Prompt Abortions*, New Eng. Ctr. for Investigative Reporting (Dec. 13, 2014), *available at* https://eye.necir.org/2014/12/13/prenatal-testing/; *Prenatal Tests Have High Failure Rate, Triggering Abortions*, NBC News (Dec. 14, 2014), http://www.nbcnews.com/health/womens-health/prenatal-tests-have-high-failure-rate-triggering-abortions-n267301.

12. Megan Allyse et al., *Non-Invasive Prenatal Testing: A Review of International Implementation and Challenges*, 7 Int'l J. Women's Health 113, 114 (2015); Norwitz et al., *supra* note 5, at 59.

13. Allyse et al., *supra* note 12, at 114.

14. M.M. Gil et al., *Analysis of Cell-Free DNA in Maternal Blood Screening for Fetal Aneuploidies: Updated Meta-Analysis*, 45 Ultrasound Obstetrics & Gynecology 249, 261–62 (2015).

15. Norwitz et al., *supra* note 5, at 50.

16. U.S. Food & Drug Admin., Draft Guidance for Industry, Food and Drug Administration Staff, and Clinical Laboratories: Framework for Regulatory Oversight of Laboratory Developed Tests (LDTs) 21 (Oct. 3, 2014).

17. 42 U.S.C. § 263(a).

18. *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980).

19. *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 130 (1948).

20. U.S. Patent No. 6,258,540.

21. *Alice Corp. Pty. Ltd. v. CLS Bank Int'l*, 134 S. Ct. 2347, 2353 (2014) ("[A] court must first 'identif[y] the abstract idea represented in the claim,' and then determine 'whether the balance of the claim adds significantly more.'" (alteration in original)); *Mayo Collaborative Servs. v. Prometheus Labs., Inc.*, 132 S. Ct. at 1289, 1299 (2012) ("'These other steps apparently added to the formula something that in terms of patent law's objectives had significance—they transformed the process into an inventive application of the formula.").

22. *Ariosa Diagnostics, Inc. v. Sequenom, Inc.*, 19 F. Supp. 3d 938, 950 (N.D. Cal. 2013).

23. *Ariosa Diagnostics, Inc. v. Sequenom, Inc.*, 788 F.3d 1371 (Fed Cir.), *reh'g en banc denied*, 809 F.3d 1282 (Fed. Cir. 2015), *cert. denied*, 136 S. Ct. 2511 (2016).

24. *Mayo*, 132 S. Ct. at 1294.

25. *Ariosa*, 788 F.3d at 1378 ("Thus, in this case, appending routine, conventional steps to a natural phenomenon, specified at a high level of generality, is not enough to supply an inventive concept.").

26. *Ariosa,* 809 F.3d at 1286 (Lourie, J., concurring).

27. *Id.* at 1287.

28. *Id.* at 1293 (Dyk, J., concurring).

29. *Ariosa*, 788 F.3d at 1380–81 (Linn, J., dissenting).

30. *Ariosa*, 809 F.3d at 1293–94 (Newman, J., dissenting).

31. *See, e.g.*, *Verinata Health, Inc. v. Ariosa Diagnostics, Inc.*, No. 12-cv-05501-SI (N.D. Cal. Jan. 19, 2017).

ARTIFICIAL INTELLIGENCE AND THE FUTURE OF LEGAL PRACTICE

BY GARY E. MARCHANT

The alarming headlines and predictions of artificial intelligence (AI) replacing lawyers have no doubt created discomfort for many attorneys already anxious about the future of their profession: "Rise of the Robolawyers." "Here Come the Robot Lawyers." "Why Hire a Lawyer? Machines Are Cheaper." "Armies of Expensive Lawyers, Replaced by Cheaper Software." "Law Firm Bosses Envision Watson-Type Computers Replacing Young Lawyers." "Why Lawyers and Other Industries Will Become Obsolete. You Should Stop Practicing Law Now and Find Another Profession." And so on.

Despite these dire headlines, AI will fortunately not replace most lawyers' jobs, at least in the short term. One in-depth study of the legal field estimated that AI would reduce lawyers' billing hours by only 13 percent over the next five years.[1] Other estimates are a little less sanguine, but still not projecting a catastrophic impact on attorney employment. A database on the effect of automation on over 800 professions created by McKinsey & Company found that 23 percent of the average attorney's job could be replaced by robots.[2] A study by Deloitte estimated that 100,000 legal jobs will be eliminated by automation in the United Kingdom by 2025.[3] And last year JPMorgan used an AI computer program to replace 360,000 billable hours of attorney work, with one report of this development observing that "[t]he software reviews documents in seconds, is less error-prone and never asks for vacation."[4]

***Gary E. Marchant**, PhD (gary.marchant@asu.edu) is a Regents' Professor, Lincoln Professor of Emerging Technologies, Law & Ethics, and Faculty Director of the Center for Law, Science & Innovation at the Sandra Day O'Connor College of Law at Arizona State University. His teaching and research interests focus on the governance of emerging technologies. Prior to his joining the faculty of ASU in 1999, Professor Marchant was a partner at Kirkland & Ellis in Washington, D.C.*

As with many new technologies, there is a cycle of hype at the outset that creates inflated expectations, even though the long-term implications of that technology may be profound and enormous. As Bill Gates perceptively noted in his book *The Road Ahead*, "[w]e always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next ten."[5] Right now, AI in the practice of law is more of an opportunity than a threat, with early adopters providing more efficient and cost-effective legal services to an expanding portfolio of existing and potential clients.

The use of AI in law will thus be an evolution, not a revolution.[6] But make no mistake, AI is already transforming virtually every business and activity that attorneys deal with, some more quickly and dramatically than others, and the legal profession will not be spared from this disruptive change. Incorporation of AI into a law firm's systems and operations is a gradual, learning process, so early adopters will have a major advantage over firms that lag in adopting the technology. The lawyers, law firms, and businesses that do not get on the AI bandwagon will increasingly be left behind, and eventually displaced. As a recent *ABA Journal* cover story explained, "Artificial intelligence is changing the way lawyers think, the way they do business and the way they interact with clients. Artificial intelligence is more than legal technology. It is the next great hope that will revolutionize the legal profession."[7]

What Is Artificial Intelligence

At its simplest, AI is the development and use of computer programs that perform tasks that normally require human intelligence. At this time and for the foreseeable future, current AI capabilities only permit computers to approach, achieve, or exceed certain but not all human cognitive functions. While some researchers are working on developing computers that can match or eclipse the human mind, sometimes referred to as "general intelligence" or "superintelligence,"[8] such an achievement is likely decades away. That is why important legal skills based on human judgment, inference, common sense, interpersonal skills, and experience will remain valuable for the lifetime of any lawyer practicing today.

While AI has many attributes for its many different applications, two are currently most important for legal applications. First, *machine learning* is the capability of computers to teach themselves and learn from experience. This means that the AI can do more than blindly adhere to what it has been programmed to do, but can learn from experience and data to constantly improve its capabilities. This is how Google's Deep Mind system was able to defeat the world's best human Go players. Second, *natural language processing* is the capability of computers to understand the meaning of spoken or written human speech and to apply and integrate that understanding to perform human-like analysis.

AI is rapidly being applied to all major sectors of the economy and society, including medicine, finance, national defense, transportation, manufacturing, the media, arts and entertainment, and social relationships, to name just some. Many of these applications will create new legal issues for lawyers, such as the liability issues of autonomous cars, the legality of lethal autonomous weapons, financial bots that may run afoul of antitrust laws, and the safety of medical robots. But in addition to changing the subject matter that lawyers work on, it will also transform the way lawyers practice their craft.

AI Applications for Legal Practice

AI is rapidly infiltrating the practice of law. A recent survey of managing partners of U.S. law firms with 50 or more lawyers found that over 36 percent of law firms, and over 90 percent of large law firms (>1,000 attorneys), are either currently using or actively exploring use of AI systems in their legal practices.[9] The following summary describes some of the major categories and examples of such applications.

Technology-Assisted Review

Technology-assisted review (TAR) was the first major application of AI in legal practice, using technology solutions to organize, analyze, and search very large and diverse data sets for e-discovery or record-intensive investigations. Going far beyond keyword and Boolean searches, studies show that TAR provides a fifty-fold increase in efficiency in document review than human review.[10] For example, predictive coding is a TAR technique that can be used to train a computer to recognize relevant documents by starting with a "seed set" of documents and providing human feedback; the trained machine can then review large numbers of documents very quickly and accurately, going beyond individual words and focusing on the overall language and context of each document. Numerous vendors now offer TAR products.

Legal Analytics

Legal analytics use big data, algorithms, and AI to make predictions from or detect trends in large data sets. For example, Lex Machina, now owned by LexisNexis, uses legal analytics to predict trends and outcomes in intellectual property litigation, and is now expanding to other types of complex litigation. Wolters Kluwer leverages a massive database of law firm billing records to provide baselines, comparative analysis, and efficiency improvements for in-house counsel and outside law firms on staffing, billing, and timelines for various legal matters. Ravel Law, also recently purchased by LexisNexis, uses legal analytics of judicial opinions to predict how specific judges may decide cases, including providing recommendations on specific precedents and language that may appeal to a given judge. Law professor Daniel Katz and his colleagues have utilized legal analytics and machine learning to create a highly accurate predictive model for the outcome of Supreme Court decisions.[11]

Practice Management Assistants

Many technology companies and law firms are partnering to create programs that can assist with specific practice areas, including transactional and due diligence, bankruptcy, litigation research and preparation, real estate, and many others. Sometimes billed as the first robot lawyer, ROSS is an online research tool using natural language processing powered by IBM Watson that provides legal research and analysis for several different law firms today, and can reportedly read and process over a million legal pages per minute. It was first publicly adopted by the law firm BakerHostetler to assist with its bankruptcy practice, but is now being used by that firm and several others for other practice areas as well. A similar system is RAVN developed in the United Kingdom and first publicly adopted by the law firm Berwin Leighton Paisner in London in 2015 to assist with due diligence in real estate deals by verifying property details against the official public records. According to the law firm attorney in charge of implementation: "once the program has been trained to identify and work with specific variables, it can complete two weeks' work in around two seconds, making [it] over 12 million times quicker than an associate doing the same task manually."[12] Kira is another AI system that has already been adopted by several law firms to assist with automated contract analysis and data extraction and due diligence in mergers and acquisitions.

Legal Bots

Bots are interactive online programs designed to interact with an audience to assist with a specific function or to provide customized answers to the recipient's specific situation. Many law firms are developing bots to assist current or prospective clients in dealing with a legal issue based on their own circumstances and facts. Other groups are developing pro bono legal bots to assist people who may not otherwise have access to the legal system. For example, a Stanford law graduate developed an online chat bot called DoNotPay that has helped over 160,000 people resolve parking tickets, and is now being expanded to help refugees with their legal problems.

Legal Decision Making

AI is enabling judicial decision making in a number of ways. For example, the Wisconsin Supreme Court recently upheld the use of algorithms in criminal sentencing decisions.[13] While such algorithms represent an early use of primitive AI (some may not consider such algorithms AI at all), they open the door to use more sophisticated AI systems in the sentencing process in the future. A number of online dispute resolution tools have or are being developed to completely circumvent the judicial process. For example, the Modria online dispute resolution tool, developed from the eBay dispute resolution system, has been used to settle many thousands of disputes online using an AI system. The U.K. government is developing an Internet-based dispute resolution system that will be used to resolve minor (<£25,000) civil legal claims without any court involvement. Microsoft and the U.S. Legal Services Corporation have teamed up to provide machine learning legal portals to provide free legal advice on civil law matters to people who cannot afford to hire lawyers.

The Future of AI and the Law

These initial applications of AI to legal practice are just the early beginnings of what will be a radical technology-based disruption to the practice of law. AI "represents both the biggest opportunity and potentially the greatest threat to the legal profession since its formation."[14] The transformative impacts of AI on legal practice will continue to accelerate going forward. AI will take over a steadily increasing share of law firm billable hours, be applied to an ever-expanding set of legal tasks, and require knowledge and abilities outside the existing skill set of most current practicing attorneys. Today AI represents an opportunity for a law firm or an attorney to be a leader in efficiency, cost-effectiveness, and productivity, but soon incorporation of AI into practice will be a matter of keeping up rather than being a leader.

AI in the practice of law raises many broader issues that can only be briefly listed here. How will AI change law

firm billing, where a smart AI system can conduct searches and analyses in a few seconds that formerly would have taken several weeks of an associate's billable time? If AI eliminates many of the more routine tasks in legal practice that are traditionally performed by young associates, how will this affect hiring and advancement of young attorneys? How will legal training and law schools need to change to address the new realities of AI-driven legal practice? How will AI affect the competitive advantage of large law firms versus small and medium-sized firms? Will companies start obtaining legal services directly from legal technology vendors, skipping law firms altogether? Will AI systems be vulnerable to charges of unauthorized practice of law? Given that AI systems increasingly use their own self-learning rather than preprogrammed instructions to make decisions, how can we ensure the accuracy, legality, and fairness of AI decisions? Will lawyers be responsible for negligence for relying on AI systems that make mistakes? Will lawyers be liable for malpractice for not using AI that exceeds human capabilities in certain tasks? Will self-learning AI systems need to be deposed and take the stand as witnesses to explain their own independent decision making?

One thing is certain—there will be winners and losers among lawyers who do and do not uptake AI, respectively. As one senior lawyer recently remarked, "Unless private practice lawyers start to engage with new technology, they are not going to be relevant even to their clients."[15] The AI train is leaving the station—it is time to jump on board. ◆

Endnotes

1. Dana Remus & Frank S. Levy, Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law 46 (Nov. 27, 2016) (unpublished manuscript), https://ssrn.com/abstract=2701092.

2. David Johnson, *Find Out If a Robot Will Take Your Job*, Time (Apr. 19, 2017), http://time.com/4742543/robots-jobs-machines-work/.

3. *Deloitte Insight: Over 100,000 Legal Roles to Be Automated*, Legal IT Insider (Mar. 16, 2016), https://www.legaltechnology.com/latest-news/deloitte-insight-100000-legal-roles-to-be-automated/.

4. Hugh Son, *JPMorgan Software Does in Seconds What Took Lawyers 360,000 Hours*, Bloomberg (Feb. 27, 2017), https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance.

5. Bill Gates, The Road Ahead (1995).

6. Joanna Goodman, Robots in Law: How Artificial Intelligence Is Transforming Legal Services 3 (2016).

7. Julie Sobowale, *Beyond Imagination: How Artificial Intelligence Is Transforming the Legal Profession*, A.B.A. J., Apr. 2016, at 46, 48.

8. Nick Bostrom, Superintelligence: Paths, Dangers, Strategies (2014).

9. Thomas S. Clay & Eric A. Seeger, Altman Weil Inc., 2017: Law Firms in Transition 84 (2017), http://www.altmanweil.com/LFiT2017/.

10. Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, 17 Rich. J.L. & Tech. 11, 43 (2011).

11. Daniel Martin Katz et al., *A General Approach for Predicting the Behavior of the Supreme Court of the United States*, 12 PLoS ONE, 2017, https://doi.org/10.1371/journal.pone.0174698.

12. Goodman, *supra* note 6, at 31.

13. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

14. Goodman, *supra* note 6, at 129 (quoting Rohit Talwar).

15. LexisNexis, Lawyers and Robots? Conversations around the Future of the Legal Industry 3 (2017) (comment of David Halliwell of U.K. law firm Pinsent Masons).

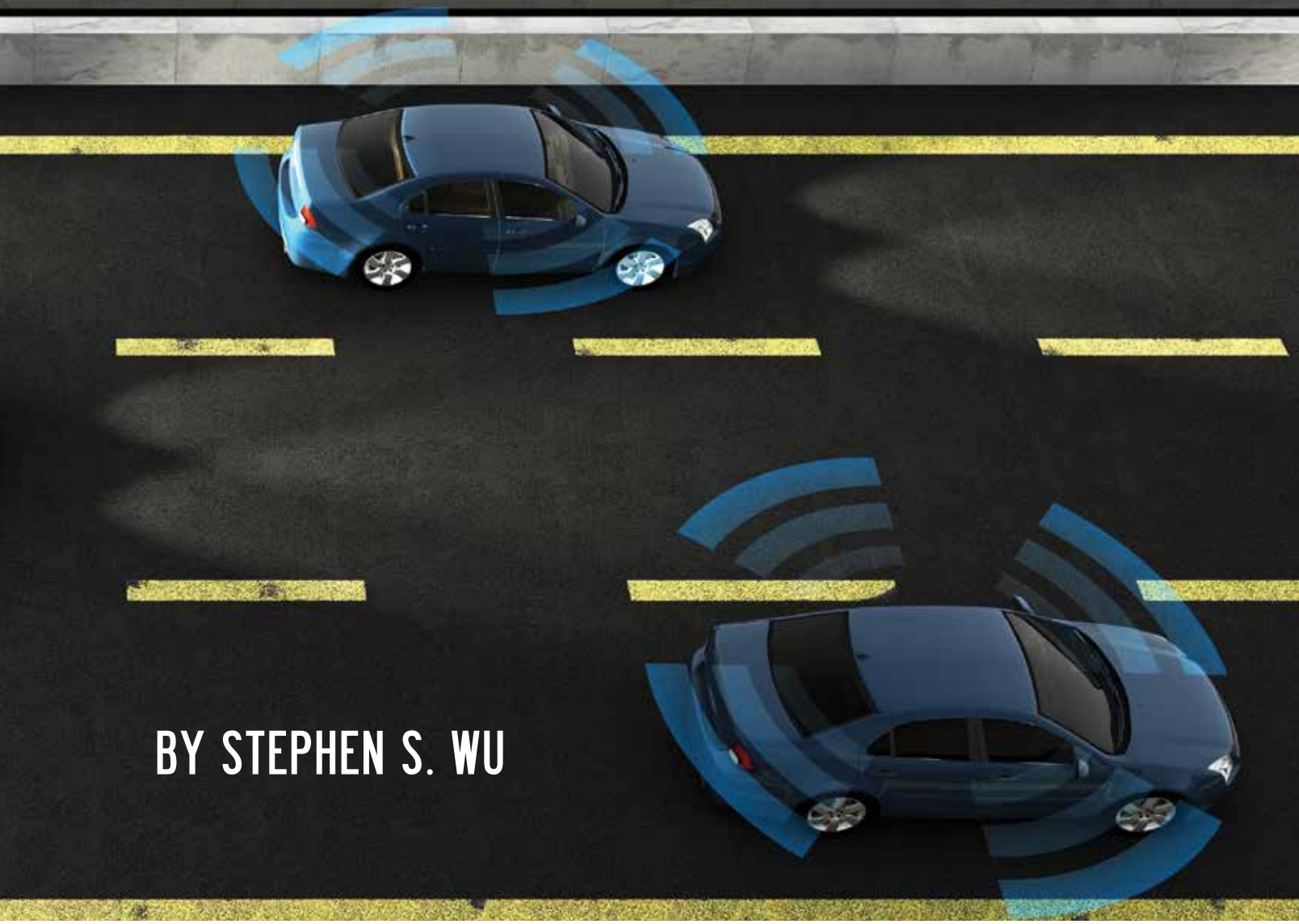
WHEN LAW AND ETHICS COLLIDE WITH AUTONOMOUS VEHICLES

From time to time, busy practicing lawyers face ethical issues of the kind taught in professional responsibility law school classes and continuing legal education courses. However, they do not often discuss the kinds of general ethical issues that academics and professional moral philosophers take up. Recent developments in artificial intelligence and robotics, and autonomous driving in particular, have rekindled interest in ethics throughout the world, and especially in the United States.

Autonomous vehicles (AVs) have captured the imagination of writers in popular media. Living close to the garage where Waymo (the new Google affiliate) houses its AVs in Mountain View, California, I feel like I am living in the AV capital of the world, as I frequently see AVs navigating the streets around my home in Los Altos. Nearby Tesla has deployed a driver assistance system in its cars and intends to deploy fully automated vehicles in two years. Companies are also working on freight truck automation, and their work eventually will result in fully automated trucks.

AV manufacturers will rely on sophisticated algorithms to control AVs. Software implementing such algorithms depends on inputs from sensors, such as light detection and ranging (LiDAR), radar, cameras, and GPS. The software analyzes the AV's location, position relative to the road, and upcoming obstacles. These algorithms then determine the best path to follow and cause the AV throttle, brake, and steering to follow the planned path. A group of moral philosophers has raised ethical questions about these algorithms. In particular, this group asks how AVs should behave when accidents are about to occur. What is the moral way to design AV algorithms? Should they try to preserve the maximum number of lives (assuming they are sophisticated enough to engage in such a calculation)? Or should they avoid doing harm to innocent pedestrians, bystanders, and passengers? Does the manufacturer owe any special ethical duties to the purchaser of the AV or the AV occupants, as opposed to occupants of other vehicles or those outside the AV? Many of the media stories raising these ethical issues rely on the work of Professor Patrick Lin of California Polytechnic State University.

Professor Lin likes to use "thought experiments" to explain ethical dilemmas. Thought experiments are "similar to everyday science experiments in which researchers create unusual conditions to isolate and test desired variables"[1] and are similar to the hypotheticals law professors use to teach legal subjects. Thought experiments can be used to study ethical issues



BY STEPHEN S. WU

involving AV algorithms. Indeed, the last administration's Department of Transportation policy on highly automated vehicles specifically mentions ethical issues in programming AVs: "Manufacturers and other entities, working cooperatively with regulators and other stakeholders (e.g., drivers, passengers and vulnerable road users), should address these situations to ensure that such ethical judgments and decisions are made consciously and intentionally."[2]

Of Trolleys and Autonomous Vehicles

Perhaps the most famous thought experiment is the so-called "trolley problem." As the name suggests, the trolley problem involves a runaway trolley. British philosopher Philippa Foot invented the "trolley problem" and first introduced it in 1967.[3] American philosopher Judith Jarvis Thomson expanded on the trolley problem in a 1985 *Yale Law Journal* comment,[4] which is the more common formulation of the thought experiment: a runaway trolley is heading down the track toward five workers and will soon run over them if no intervention occurs. A spur of track leads off to the right, but there is one worker on the track. A bystander is standing by a switch. If the bystander throws the switch, the trolley will turn onto the spur, saving the five workers, but killing the single worker on the spur.[5]

If the bystander does nothing, the bystander would not be killing anyone. The bystander would merely be "allowing" the five to die. Throwing the switch would involve killing just one person. Some philosophers such as Jarvis Thomson have the view that it is better to maximize the number of lives saved in situations like this. Others such as Foot disagree, saying that it is worse from an ethical standpoint to cause harm than it is to allow harm to happen, even if the consequences are worse. The trolley problem teases out the moral philosopher's dilemma: is it better to throw the switch and save more lives (five versus one), or is it better (for the bystander) to do nothing in order to avoid causing harm to anyone?

Professor Lin has applied the trolley problem to AVs by posing the following thought experiment:

> [Y]ou are about to run over and kill five pedestrians. Your car's crash-avoidance system detects the possible accident and activates, forcibly taking control of the car from your hands. To avoid this disaster, it swerves in the only direction it can, let's say to the right. But on the right is a single pedestrian who is unfortunately killed.[6]

WHEN FACED WITH AN INEVITABLE CRASH, SHOULD AUTONOMOUS VEHICLES BE PROGRAMMED TO SAVE MORE LIVES OR AVOID CAUSING HARM?

News writers have (with or without crediting Professor Lin) repeated this and similar scenarios in numerous recent news articles.[7] Philosophers continue to debate the question of whether it is better to save more lives or avoid doing harm. To the extent there is any consensus, a recent survey showed that philosophers favored throwing the switch in the trolley problem.[8] Thus, if an AV manufacturer hires professional philosophers to advise it on how to design AV algorithms, they are likely to advise the manufacturer to program the AV to steer away from a large group at the cost of running over a single individual.

The Legal Trolley Problem Dilemma

As a practicing lawyer, I was curious. What would the legal consequences be if an AV manufacturer followed a philosopher's advice and tried to "do the right thing" in trolley problem situations? What would happen if the manufacturer programmed its AVs to steer away from a large group and toward a single individual or small group when they anticipate that a crash is inevitable? For the remainder of this article, I imagine a hypothetical manufacturer (Manufacturer) has implemented just such an algorithm. And I imagine that an accident occurs where the AV steers away from five people (the Five), but at the cost of striking and killing a single individual (the One). I assume that the One was an innocent bystander or pedestrian, rather than a jaywalker or someone engaging in wrongful conduct. I also assume that if the AV had attempted to avoid collision altogether, the AV may have made things worse—it may have killed all six people. I imagine that a representative of the One files a complaint against the Manufacturer.

The most common causes of action in a suit claiming a defect in a product include strict products liability, negligence, breach of warranty, and statutory violations for unfair or deceptive trade practices. With each claim, counsel for the representative of the One would contend that the feature of swerving toward the One made the AV defective. But even worse, the Manufacturer's conduct appears intentional. Indeed, the Manufacturer made a deliberate decision to cause the AV to swerve toward the One (or someone similarly situated to the One). The representative may even assert a cause of action for battery, the essence of which is harmful contact intentionally done.[9] On its face, the representative seems to have a strong case.

The Manufacturer would not have fared any better if it programmed the AV to do nothing, allowing the AV to run over the Five. If the AV killed the Five, representatives of the Five could file suit against the Manufacturer, contending that the Manufacturer had a safer alternative design: it could have programmed the AV to run over the One. Thus, it appears the Manufacturer is in a no-win situation.

Possible Defenses

The Manufacturer might turn to traditional defenses recognized in the law to avoid the dilemma. For instance, it could assert a necessity defense, saying that running over the One was necessary to save lives. Under the necessity doctrine, "it has long [been] recognized that '[n]ecessity often justifies an action which would otherwise constitute a trespass, as where the act is prompted by the motive of preserving life or property and reasonably appears to the actor to be necessary for that purpose.'"[10] The private necessity defense thus serves as a justification for a nongovernmental defendant's conduct where the defendant's act causes harm, but the defendant acted to prevent an even worse harm. However, necessity is likely to be unavailing as a defense for the Manufacturer. In its traditional form, the necessity defense justifies acts of trespass or damage to personal property, but not bodily injury.[11] In our hypothetical case, the AV killed the One and thus does not apply.

Another possible defense is the defense of third persons. Similar to self-defense, the Manufacturer might try to argue that its use of force against the One is justified on order to defend the Five against harm. The Restatement (Second) of Torts provides that an actor can defend any third person from wrongful injury by the use of force.[12] However, the Manufacturer's argument will fail because in our hypothetical case the One was not acting wrongfully. To the contrary, we have assumed that

Stephen S. Wu *(ssw@svlg.com) is a shareholder in Silicon Valley Law Group in San Jose, California, and practices in the areas of information technology, security, privacy, and intellectual property compliance, litigation, transactions, and policies. He served as the 2010–2011 Chair of the ABA Section of Science & Technology Law.* *This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1522240. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF. The author is grateful for the support of the NSF, in addition to Michael Wu for making useful suggestions to improve this article, as well as his editorial assistance.*

the One was an innocent actor. There is no wrongful conduct for the Manufacturer to defend against, and thus the defense does not apply.

A third defense the Manufacturer could try to assert is the "sudden emergency" doctrine, also known as the "imminent peril" doctrine. "[I]f an actual or apparent emergency is found to exist the defendant is not to be held to the same quality of conduct after the onset of the emergency as under normal circumstances."[13] Cases involving the sudden emergency doctrine in the car accident context involve split-second decisions of drivers in a difficult position. The facts of some of these cases sound like real-world trolley problems.[14] The defense recognizes that an actor in such situations cannot be held to the same standard of care as when an actor is calm in normal circumstances. However, the problem for the Manufacturer is that the Manufacturer is considering how to program an AV in the ordinary course of the design process, far from any imminent accident. The sudden emergency doctrine applies only when, *at the time of the actor's conduct causing the accident*, the actor faced a sudden choice between two or more actions. Here, the Manufacturer's programming decision occurred long before the accident. The Manufacturer was not facing a sudden decision. To the contrary, we have assumed that the Manufacturer undertook a careful and deliberate analysis of how to design its AV algorithms and made a choice to program the AV to steer toward the One. No sudden emergency was occurring during the design process. Accordingly, the defense does not apply.

Resolving the Liability Dilemma

Because the traditional defenses offer no protection, the Manufacturer has no easy way out of the liability dilemma. As the law currently stands, I believe the only way for the Manufacturer to limit its legal liability in the trolley problem scenario is to program its AVs to attempt to avoid collision. It should neither steer toward the One nor allow the AV to run over the Five. Rather, it should try to maximize collision avoidance.

I recognize three problems with this approach. First, we have assumed that collision avoidance may make things worse and the AV may end up hurting or killing all six people. Nonetheless, it is more legally defensible and would, as a practical matter, sit better with a jury: the Manufacturer did all it could to save everyone's life. If the accident ended up killing all six, then at least the Manufacturer tried to save lives.

Second, my position is implicitly at odds with the trolley problem thought experiment. I am implicitly rejecting what appears to be a false choice between running over the Five or running over the One.

Finally, I recognize that my choice of collision avoidance of the "legal" solution is not the one philosophers would consider "moral." Law and morality sometimes diverge. Conduct we consider immoral may be legal, and some conduct considered to be morally permissible may be illegal. This is one more case in which law and morality may come to different conclusions. Given the liability dilemma, the only way to immunize the Manufacturer trying to "do the right thing" and allow it to program AVs to steer toward the One is to change the law through legislation or regulations.

Trolley problems are useful starting points for analyzing the ethical issues of programming AVs, if nothing else because they spark discussion among the media and their audience. Some people reject the real-world relevance of the trolley problem, but the principles gleaned from it will aid manufacturers in deciding how to program AVs. More generally, injecting discussions of ethics raises the awareness of ethical dimensions to AV design and manufacturers' decisions, and that is a good thing. ◆

Endnotes

1. Patrick Lin, *Why Ethics Matters for Autonomous Cars*, *in* Autonomous Driving: Technical, Legal and Social Aspects 69, 75 (Markus Maurer et al. eds., 2016).

2. U.S. Dep't of Transp. & Nat'l Highway Traffic Safety Admin., Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety 26 (2016). Likewise, the Trump administration's latest guidance on AVs acknowledges that manufacturers should deliberate concerning ethical issues. U.S. Dep't of Transp. & Nat'l Highway Traffic Safety Admin., Automated Driving Systems 2.0: A Vision for Safety 1 n.1 (2017).

3. Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, Oxford Rev., 1967, at 5, 8.

4. Judith Jarvis Thomson, *The Trolley Problem*, 94 Yale L.J. 1395, 1395 (1985).

5. *See id.* at 1397.

6. Lin, *supra* note 1, at 79.

7. *See, e.g.*, *Why Self-Driving Cars Must Be Programmed to Kill*, MIT Tech. Rev. (Oct. 22, 2015), http://www.technologyreview.com/view/542626/why-self-driving-cars-must-be-programmed-to-kill/.

8. David Bourget & David J. Chalmers, What Do Philosophers Believe? 16 (Nov. 30, 2013), https://philpapers.org/archive/BOUWDP.

9. *See, e.g.*, 5 B.E. Witkin, Summary of California Law § 383, at 599 (10th ed. 2005).

10. *People v. Ray*, 981 P.2d 928, 935 (Cal. 1999).

11. *See* Restatement (Second) of Torts § 263 (Am. Law Inst. 1965) ("One is privileged to commit *an act which would otherwise be a trespass to the chattel of another or a conversion of it,* if it is or is reasonably believed to be reasonable and necessary to protect the person or property of the actor, the other or a third person from serious harm." (emphasis added)).

12. *See id.* § 76.

13. Bill Hollingsworth, Note, *The Sudden Emergency Doctrine in Florida*, 21 U. Fla. L. Rev. 667, 671 (1969).

14. *See, e.g.*, *Myhaver v. Knutson*, 942 P.2d 445, 446 (Ariz. 1997) (en banc) (involving a driver who faced a choice of driving straight into a car driving the wrong way in his lane or crossing the yellow line into oncoming traffic).

ARTIFICIAL INTELLIGENCE AND THE LAW

MORE QUESTIONS THAN ANSWERS?

BY KAY FIRTH-BUTTERFIELD

Artificial intelligence (AI) will be everywhere. It will ensure our world runs smoothly and our every need is met. In its not very intelligent form, it is here already in our cars, smartphones, search engines, and translation and personal assistants; in our homes in the form of robot cleaners and lawnmowers; on the street helping with surveillance, traffic monitoring, and policing; and even in condoms and sex dolls—the list is extensive and growing. AI beats us at the game of Go, even creating moves that we humans have failed to notice in hundreds of years of play; and it wins at poker, a game that requires players to perfect the art of bluff.

Just around the corner are AI-enabled dolls to become our children's real imaginary friends, as well as fully autonomous cars. It is worth noting that although John McCarthy first suggested the idea of autonomous cars as a scientific possibility in the 1960s, the technology has only made enormous strides to make this possible in the last few years.

And yet Vint Cerf, the "father of the Internet," describes AI not as intelligent but as an "artificial idiot."[1] The prize-winning Go-playing computer has no idea it is playing Go. However, AI is excellent at learning, and with enough data to train it and some often, basic instruction, it can learn a new skill—for example, sorting through all the documents in a case and make decisions about discovery, or helping a doctor to diagnose cancer.

What Is Artificial Intelligence?

To understand, we have to realize that AI is not one technology but a range of techniques that give the appearance of intelligence. AI is applied math and statistics at their very best. Techniques such as reinforcement learning, neural nets, deep learning, and more are driving the AI revolution, but they are not—and seem nowhere near—artificial general intelligence (AGI). AGI will be achieved when a computer can perform all the same intellectual activities as a human.

For lawyers, this lack of definition of AI is a problem. If we are unable to describe something, we cannot legislate or easily draw on the correct existing law when cases come to court in the absence of legislation. Indeed, it is certainly arguable that, as this product is continuously evolving and is unlike any product we have ever seen before, no current legal precedent could apply. Recently, Senator Maria Cantwell (Wash.) proposed a bill that would require the U.S. Department of Commerce to form a committee with an AI focus. According to *GeekWire*, the draft also seeks to create a federal definition of AI as "systems that think and act like humans or that are capable of unsupervised learning," and differentiates between AGI or a system that "exhibits apparently intelligent behavior at least as advanced as a person across the full range of cognitive, emotional, and social behaviors," and "'narrow artificial intelligence,' such as self-driving cars or image recognition."[2] Others

***Kay Firth-Butterfield**, LLM, MA, FRSA (kay.firth-butterfield@weforum.org) is a barrister in the United Kingdom and the Project Head of AI and ML at the World Economic Forum. She is an associate fellow of the Centre for the Future of Intelligence at the University of Cambridge and a senior fellow and distinguished scholar at the Robert S. Strauss Center for International Security and Law, University of Texas, Austin; cofounder of the Consortium on Law and Ethics of A.I. and Robotics, University of Texas, Austin; and vice-chair of the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. She has taught courses on law and policy of emerging technologies and AI at the University of Texas Law School in Austin.*

suggest that we should have "use case" definitions—for example, the way in which Nevada has defined AI for use in autonomous vehicles as "the use of computers and related equipment to enable a machine to duplicate or mimic the behavior of human beings."[3]

Transparency

Currently, the legislation around this technology is principally concerned with data privacy and autonomous vehicles.[4] In Europe, the "home of privacy," the General Data Protection Regulation (GDPR) will come into force in 2018.[5] Among other provisions, the GDPR gives a citizen of the European Union (EU) the right to demand an account of how a decision that affected them adversely was achieved. Thus, if an algorithm was used in the denial of a loan to an EU citizen, that citizen can require the loan company to explain how it came to its decision.

This presents a problem for most of the systems currently known as AI because many develop their decisions within what is termed a "black box." This is not the informative black box flight recorder but rather an opaque one where AI algorithms crunch data to achieve answers. In other words, the scientist feeds data into the computer and the computer uses many iterations of questions and answers to achieve the answer for which it was trained. Using the poker-playing computer as an example, it ran millions of games against itself using three computers powered by supercomputers to achieve its victory. With enough data, computers have the speed, discipline, and endurance to do complex tasks; however, the scientists who design them more often than not do not know how the computers achieve the answer. For the GDPR to work, developers of AI will have to make these systems transparent.

Take, for example, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) developed to assist U.S. judges in sentencing. Under the GDPR, defendants who wish to challenge the fairness of their sentences would ask to see how the computer arrived at its decision. According to a ProPublica study, in the COMPAS model the computer is trained on historic criminal justice data, which may lead to corruption of the data and the subsequent decision—creating racially biased decisions because the historic data encompassed them.[6] In *State v. Loomis*, the judge used the COMPAS tool to assist with sentencing.[7] The Wisconsin Supreme Court rejected Loomis's appeal, saying he would have received the same sentence whether or not the AI was involved.[8] However, the court seemed concerned about the use of COMPAS. Chief Justice Roberts was likewise concerned, when in response to a question regarding AI in the courts he said that AI is already in courtrooms "and it's putting a significant strain on how the judiciary goes about doing things."[9] However, a few months later, the U.S. Supreme Court decided not to hear the writ of certiorari of the Loomis appeal. The question of bias does not end in the way data is collected or cleaned or how it is used in AI—it also comes from the scientists themselves when they are training the AI algorithm. Some suggest that Alexa only comes with a female voice because it was programmed by predominantly white male geeks.

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems has recommended a new IEEE standard to deal with this problem of transparency.[10] P7001 is in the working group phase at the moment and argues that diverse stakeholders need transparency and that transparency is essential to the way we should design embodied AI—for example, autonomous cars and robots. It is argued that accident investigators, lawyers, and the general public need to know what the car or household robot was doing at the time of an accident in order to allocate blame and damages and, most importantly, instill trust in the technology. However, some scientists consider the task of transparency too difficult to achieve. It is their view that, as human beings, we cannot hope to understand complex AI algorithms and thus transparency is illusory because even if we can see what the system is doing we cannot understand it. Instead of human regulation they propose regulation by algorithm. Thus, a car would have standard algorithms to deal with its operation and a "guardian" algorithm to make sure it stays within its set parameters. By way of example, as data is continually collected by the standard algorithm about road users so that it can improve the safety and reliability of driving, the guardian AI would prevent the standard algorithms learning to speed from their collected data about the habits of the human driver. The remaining unsolved question is who will guard the guardians?

Undoubtedly, opaque and transparent systems raise intellectual property (IP), copyright, and patent issues, which will have to be reconciled with legislation or in the courts.

Privacy Concerns

Nor does the problem of data end here. For our current AIs to work they need massive data sets, which is why the *Economist* called data the new oil.[11] If you have data you can create AI, and therefore everything we do is of value to someone; collection and sale of our data is big business. With devices that listen and observe in our home, a once private place has lost its privacy. Some of these devices listen and record and store those recordings the whole time, while others, like Alexa, listen for key "wake up" words. In November 2015, a murder occurred at the home of James Bates. He was accused of the murder, and the prosecutor asked Amazon for any recordings created by Alexa at the time of the death. Amazon refused saying, "Given the important First Amendment and privacy implications at stake, the warrant should be quashed unless the Court finds that the State has met its heightened burden for compelled production of such materials."[12] However, Bates's attorney subsequently obtained copies of recordings from Amazon and released them into evidence. Therefore, this important legal issue has yet to receive a decision.

Additional concerns about privacy are applicable to children. Article 16 of the United Nations Convention on the Rights of the Child gives children a right to privacy.[13] It is difficult to exercise that right, once you have sufficient mental capacity to do so, if your parents—by having devices that listen and record in your home from your birth—have given away your childhood privacy. Indeed, a child might soon have its own monitoring device—an AI-enabled doll to talk to and learn from. Parents or legislators need to be making choices as to what these dolls upload to the cloud or teach their children and, perhaps more importantly, as to their cybersecurity protocol. It is unclear if this will be an option offered by the manufacturers or whether most parents have sufficient knowledge to understand the problem. To prove this point, in the United Kingdom the Purple company recently included a requirement for users of its free Wi-Fi to clean toilets for 1,000 hours; of the thousands who logged on, only one person read the terms and conditions in which this was included.[14] It is for this reason that the German government banned "Cayla," an AI-enabled doll, earlier this year, and in their guidance for autonomous car makers said this about data collected in cars:

> It is the vehicle keepers and vehicle users who decide whether their vehicle data that are generated are to be forwarded and used. The voluntary nature of such data disclosure presupposes the existence of serious alternatives and practicability. Action should be taken at an early stage to counter a normative force of the factual, such as that prevailing in the case of data access by the operators of search engines or social networks.[15]

Another important privacy question is the development of sex robots enabled with AI. These robots will also need to collect data, and that data has to be stored somewhere. The possibility of having one's most intimate secrets hacked must be high, but the real question has to be whether, as demand is principally for female sex robots, this is just another way of continuing sexual assault on women. As we move into the robot age, we may have the opportunity to end the "oldest profession" or perhaps simply enable it to metamorphose. This question becomes all the more pressing when thinking about an AI-enabled child sex doll, which is being produced for pedophiles in Japan. Their developer argues that it stops him, and others, from assaulting human children. However, the importation of this sort of object in the United Kingdom would probably be a crime; a defendant was recently convicted of trying to import a non-AI-enabled sex robot.[16] The Foundation for Responsible Robotics recently published a neutral report in an effort to start these conversations.[17]

Regulating AI

Regulation is often said to stifle innovation, but regulation in this space seems necessary to protect the millions of customers who will buy and use AI-enabled devices. However, it seems unlikely that AI, other than autonomous vehicles, will find its way onto the federal legislative agenda anytime soon. The Kenan Institute of Ethics at Duke University has been considering the idea of "adaptive regulation," which would involve passing a regulation geared toward a specific emerging technology so that developers and users could have some security to guide investment, and revisiting the regulation at an early stage to ensure it was working.

There are a number of efforts to create guidelines for the use of AI. Some initiatives are from industry—for example, IBM has published ethical use guidelines and helped to create the Partnership on AI.[18] Additionally, nonprofits such as AI Global (groupings of geographically localized academics, industry, and government; e.g., AI Austin) and the Future of Life Institute have created guiding principles for the design, development, and use of AI.[19] Additionally, as mentioned, the IEEE has brought some 200 experts together to create standards applicable to work with AI and robotics. In the United Kingdom, there is a British Standard for Robots and Robotic Devices (BS 8611) that provides a guide to the ethical design and application of robots and robotic systems.

However, it seems that the bulk of the law regarding AI will come from judicial decision making, although there may be some regulators who already exist and who could find AI falling within their purview. In a hyper-connected world, we have already seen that cybersecurity is vital. As we extend our dependency on AI, cybersecurity will become ever more vital as AI—better able to adapt to threats faster than humans—will also run the cybersecurity systems. Developers have been using game theory to help teach algorithms about strategic defense. In one scenario, two standard algorithms played a game of collecting "things" but could also attempt to kill one another; they only resorted to trying to do so when there was scarcity of "things." However, when a cleverer algorithm was introduced it immediately killed the weaker two.[20] Regulatory standards can be built on existing ones, such as the U.S. National Institute of Standards and Technology (NIST) standards for cryptography. The Internet of Things (IoT) makes things much less safe as all these devices are a potential access point and many are cheaper to make than to patch. Additionally, as companies produce AI-enabled devices and then go out of business, the burden of security and safety will become greater. For example, will car makers be required to maintain the AI software throughout the lifetime of the car and multiple owners?

As to governmental endeavors in the United States, the benefits and problems of AI were considered by the Obama administration in two reports from the Office of Science and Technology, the second focusing on the growing concern that automating our brains might lead to mass unemployment.[21] AI has yet to be taken up as a topic of debate by the Trump administration, but there is a bipartisan working group on AI (led by Congressmen John Delaney (Md.) and Pete Olson (Tex.)),[22] and the Trump administration has voiced its support for the creation of autonomous vehicles. In the political arena, we may have

seen the impact of AI on the decisions of the American public in the presidential race by the focusing of "fake news" and by the minute targeting of voters using AI to build unique profiles of voters from their public records and social media accounts. Individual security of data is often impinged because users believe they have their accounts locked to friends and family, but this is not so. A recent Cambridge University study shows that from 10 Facebook "likes" an AI can know you as well as a work colleague.[23] Additionally, things become more complex as AI can seamlessly change video to insert words into the mouth of the speaker that are entirely different from what was actually said.[24] Recent problems have shown that bad actors can also fool image detection AI—for example, persuading it that a kitten is a computer,[25] or corrupting Microsoft's ill-fated Tay chatbot.[26]

Conclusion

Those who attempt to forecast the future have three chances: to be wrong, to be right, or to be partially right. Undoubtedly, the latter course is the best one to chart. When looking at the future of AI, the rights to data will likely become an increasingly important issue, as well as how the general population learns about AI and what it can do (so that they can safely rear their children and cast their votes). Currently, there is much hype about AI and a paucity of AI scientists outside the major corporations, which could lead to another "AI winter." This is a time of great opportunity to actually shape the way in which humanity survives into the future—we should not waste that opportunity! ◆

Endnotes

1. Frank Konkel, *Father of the Internet: "AI Stands for Artificial Idiot,"* Nextgov (May 9, 2017), http://www.nextgov.com/emerging-tech/2017/05/father-internet-shows-no-love-ai-connected-devices/137697/.

2. Tom Krazit, *Washington's Sen. Cantwell Prepping Bill Calling for AI Committee*, GeekWire (July 10, 2017), https://www.geekwire.com/2017washingtons-sen-cantwell-reportedly-prepping-bill-calling-ai-committee/.

3. Nev. Rev. Stat. § 482A.020.

4. *Ethics Commission Creates World's First Initial Guidelines for Autonomous Vehicles*, Germany.info (June 21, 2017), http://www.germany.info/Vertretung/usa/en/__pr/P__Wash/2017/06/21-AutonomousVehicles.html.

5. Council Regulation 2016/679, 2016 O.J. (L 119) 1 (effective May 25, 2018) [hereinafter GDPR].

6. Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

7. Adam Liptak, *Sent to Prison by a Software Program's Secret Algorithms*, N.Y. Times, May 1, 2017.

8. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

9. *Id.*

10. *See The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems*, IEEE Standards Ass'n, https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html (last visited Oct. 17, 2017).

11. *The World's Most Valuable Resource Is No Longer Oil, but Data*, Economist (May 6, 2017), https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.

12. Eliott C. McLaughlin, *Suspect OKs Amazon to Hand Over Echo Recordings in Murder Case*, CNN (Apr. 26, 2017), http://www.cnn.com/2017/03/07/tech/amazon-echo-alexa-bentonville-arkansas-murder-case/index.html.

13. Convention on the Rights of the Child art. 16, Nov. 20, 1989, 1577 U.N.T.S. 3.

14. Rachel Thompson, *22,000 People Accidentally Signed Up to Clean Toilets Because People Don't Read Wi-Fi Terms*, Mashable (July 13, 2017), http://mashable.com/2017/07/13/wifi-terms-conditions-toilets/#kireRB9KmiqJ.

15. *Ethics Commission Creates World's First Initial Guidelines for Autonomous Vehicles*, *supra* note 4 (guideline 15).

16. *Man Who Tried to Import Childlike Sex Doll to UK Is Jailed*, Guardian, June 23, 2017.

17. Noel Sharkey et al., Found. for Responsible Robotics, Our Sexual Future with Robots (2017).

18. *Transparency and Trust in the Cognitive Era*, IBM THINK Blog (Jan. 17, 2017), https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/.

19. *See* AI Austin, https://www.ai-austin.org/ (last visited Oct. 17, 2017); *Asilomar AI Principles*, Future of Life, https://futureoflife.org/ai-principles/ (last visited Oct. 17, 2017).

20. Joel Z. Leibo et al., *Multi-Agent Reinforcement Learning in Sequential Social Dilemmas*, *in* Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AA-MAS 2017) (S. Das et al. eds., 2017), https://storage.googleapis.com/deepmind-media/papers/multi-agent-rl-in-ssd.pdf.

21. Kristin Lee, *Artificial Intelligence, Automation, and the Economy*, Obama White House (Dec. 20, 2016), https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy.

22. Press Release, Congressman John Delaney, Delaney Launches Bipartisan Artificial Intelligence (AI) Caucus for 115th Congress (May 24, 2017), https://delaney.house.gov/news/press-releases/delaney-launches-bipartisan-artificial-intelligence-ai-caucus-for-115th-congress.

23. *Computers Using Digital Footprints Are Better Judges of Personality than Friends and Family*, Univ. of Cambridge (Jan. 12, 2015), http://www.cam.ac.uk/research/news/computers-using-digital-footprints-are-better-judges-of-personality-than-friends-and-family.

24. Meg Miller, *Watch This Video of Obama—It's the Future of Fake News*, Co.Design (July 18, 2017), https://www.fastcodesign.com/90133566/watch-this-video-an-ai-created-of-obama-its-the-future-of-fake-news.

25. Anish Athalye, *Robust Adversarial Examples*, OpenAI (July 17, 2017), https://blog.openai.com/robust-adversarial-inputs/.

26. James Vincent, *Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less than a Day*, Verge (Mar. 24, 2016), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

BUILDING ETHICAL ALGORITHMS

BY NATASHA DUARTE

This summer, the Supreme Court declined to hear a case about the constitutional rights of a man whose sentencing decision was determined in part by a computer.[1] In *State v. Loomis*,[2] the state used a tool called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) to calculate the "recidivism risk" of the defendant—the likelihood that he would be rearrested for another crime within two years, based on people who shared his characteristics and background. COMPAS asks a series of questions about the defendant and runs the answers through a proprietary algorithm, which provides a recidivism "risk score."[3] A study by ProPublica found that when COMPAS predictions were wrong, they were more likely to incorrectly classify black defendants as high risk and white defendants as low risk.[4]

Whether they are used to target ads or to determine prison sentences, the algorithms behind automated decisions represent more than math. They also represent human choices, such as what data to use to make decisions, what variables to consider, and how much weight to assign to an algorithmic prediction. Each of these choices is value-laden and may lead to different outcomes. For example, crime prediction algorithms are often trained to "learn" patterns from historical arrest records (known as the "training data"). This prioritizes the historical patterns behind law enforcement's decisions about where to patrol and whom to arrest. These tools can also be calibrated to minimize false positives or false negatives, depending on whether a jurisdiction would rather err on the side of keeping "low-risk" people in jail or letting "high-risk" people go free. For better or worse, choices about how to design and use decision-making algorithms are shaping policy, culture, and societal norms.

Industry and government have a responsibility to avoid building and using harmful automated decision-making systems. Several major tech industry players have made public commitments to lead the way on creating ethical standards for machine learning and artificial intelligence.[5] Even smaller companies are beginning to hire chief ethics officers or even assemble entire teams to focus on ethical issues regarding the use of big data, machine learning, and automated decision-making systems. Slowly but surely, institutions that build and use automated decision-making systems are recognizing ethical review as a necessary prerequisite to the large-scale deployment of these systems.

Ethical review of automated decision-making systems is a complex and nuanced process. Yet companies and policymakers do not need to start from scratch when it comes to developing an ethical framework to guide this review. Several frameworks exist that have been or can be adapted to the automated computing context. Ethical principles are only as good as the processes in place to implement and enforce them. Companies need to adopt—and constantly reevaluate—internal processes for ethical design and review. The first section of this article discusses existing ethical frameworks that can be adapted to automated decision-making systems, and the second section is devoted to implementation strategies.

Ethical Frameworks

Several established frameworks provide ethical principles to guide organizations' best practices around technology design and data use. Although these frameworks were not specifically designed for machine learning or artificial intelligence,

they can be adapted to different technologies.

Belmont Report

The Belmont Report[6] was commissioned in the 1970s, prompted by high-profile medical research scandals including the Tuskegee syphilis study[7] and the Stanford prison experiment.[8] The Belmont Report identified three key principles that continue to govern human subjects research: (1) respect for persons, (2) beneficence, and (3) justice. The first principle requires researchers to respect the basic dignity and autonomy of their subjects. Research subjects must be presented relevant information in a comprehensible format and then voluntarily give their consent to participate. "Beneficence" embodies the well-known maxim of "do no harm." It requires researchers to conduct risk-benefit assessments, maximize the possible benefits to research subjects and to the public, and minimize possible harms. Researchers must also assess the specific risks and benefits of including members of vulnerable populations (such as children or pregnant women) in a study. The "justice" principle demands that the benefits and burdens of research are fairly distributed. The report notes that fair distribution does not always mean equal distribution. "[D]istinctions based on experience, age, deprivation, competence, merit and position do sometimes constitute criteria justifying differential treatment for certain purposes."[9]

While the Belmont Report was developed to address medical research on human subjects, its principles are just as salient for big data analysis and automation. For example, consider the concept of informed consent. Imagine a user creates a social media account and agrees to a privacy policy stating that the information she discloses may be used for "research." Has the user consented to allow the company, or outside researchers, to use that information to predict the user's mental health status? Is separate consent required for this type of use, which arguably was not anticipated by the user when she created a profile? Researchers at traditional institutions must address consent issues when they seek institutional review board (IRB) approval to collect information from research subjects. However, big data analysis is often done without IRB approval for several reasons, including the ease of access to publicly available data sets (or data held by companies) and a lack of institutional clarity about whether big data research counts as human subjects research requiring IRB approval.[10]

Natasha Duarte *(Natasha@cdt.org) is a policy analyst at the Center for Democracy & Technology.*

Menlo Report

In 2012, the Menlo Report[11] was commissioned in response to new questions about the ethics of information and communications technology research (ICTR). The Menlo Report identified three factors that make risk assessment challenging in ICTR: "the researcher-subject relationships, which tend to be disconnected, dispersed, and intermediated by technology; the proliferation of data sources and analytics, which can heighten risk incalculably; and the inherent overlap between research and operations."[12] Each of these factors also applies to data-driven automated decision-making systems. Because the data that feeds automated systems is collected and aggregated digitally, data subjects often do not know they are data subjects, and the effects of automated systems can be widely dispersed, difficult to detect, and difficult to connect to one particular system.

The Menlo Report builds on the principles articulated in the Belmont Report but accounts for the additional challenges presented by information technology. For example, "In the ICTR context, the principle of Respect for Persons includes consideration of the computer systems and data that directly . . . impact persons who are typically not research subjects themselves."[13] The report added a fourth principle calling for consideration of law and public interest. This principle asks researchers to engage in legal due diligence, transparency, and accountability.

ACM Software Engineering Code of Ethics and Professional Practice

While the Belmont and Menlo Reports apply specifically to research, the Association for Computing Machinery (ACM) published a code of ethics in 1992 that applies generally to the practice and profession of software engineering. The ACM Software Engineering Code of Ethics and Professional Practice (ACM Code)[14] recognized engineering as a profession and acknowledged that "software engineers have significant opportunities to do good or cause harm" and to enable or influence others to do good or cause harm. Among other things, the ACM Code requires engineers to act in the public interest, even when serving their clients or employers. Under the ACM Code, engineers' responsibility to the public requires them to ensure that any software produced by engineers is safe, passes appropriate tests, and is ultimately in the public good; to disclose any potential danger to the user, the public, or the environment; to be fair and avoid deception concerning the software; and to consider issues of physical disabilities, allocation of resources, economic disadvantage, and other factors that can diminish access to the benefits of the software. Engineers must also report "significant issues of social concern" to their employers or clients. In turn, the ACM Code prohibits employers from punishing engineers for expressing ethical concerns.

Fair Information Practice Principles

The Fair Information Practice Principles (FIPPs)[15] are internationally recognized as the foundational principles for responsible collection, use, and management of data, and they continue to serve as guiding principles in the era of big data.[16] There are many iterations of the FIPPs, but they were first codified in 1980 by the Organisation for Economic Co-operation and Development (OECD) in its Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.[17] Compliance with the FIPPs means minimizing the amount of data collected and the length of retention; ensuring that data is collected for a specified purpose and used only in contexts consistent with that purpose (unless additional consent is requested and given); giving individuals control over and access to their personal information, including the right to consent to its collection and use and to correct or delete it; and securing data through the use of encryption, de-identification, and other methods.

Common Principles

Together, these existing frameworks can provide a roadmap to guide ethical review of big data analytics, automated decision-making systems, and artificial intelligence. Below are some examples of how these common principles can apply to automated decision-making systems. This is not a comprehensive list of considerations to guide ethical review, and many of these questions do not have clear right or wrong answers, but they provide insight into key concepts and approaches.

Individual Autonomy and Control

Does the automated system reflect individuals' privacy choices? For example, does it target ads by inferring sensitive characteristics—such as race, gender, or sexual orientation—that an individual has intentionally obscured or declined to disclose?

Beneficence and Risk Assessment

Are useful insights that companies glean from data passed on to data subjects in helpful ways? For example, information that is not obviously health related may be used to predict individuals' propensity for certain health conditions.[18] Whether these predictions should be communicated to individuals, and how best to do so, is a complex question requiring rigorous evaluation of the risks and benefits of disclosure.

Justice or Fairness

Are any groups or populations—especially protected or vulnerable classes—over- or underrepresented in the training data? Does the automated system rely on existing data or make assumptions about certain groups in a way that perpetuates social biases? Does the system create disparate outcomes for different groups or populations? Is the risk of error evenly distributed across groups?

Transparency and Accountability

Are the claims a company makes about its automated decision-making systems truthful and easy for users to understand? Do explicit and implicit statements allow users (or parties who contract to use the system) to form accurate expectations about how the system functions, its accuracy, and its usefulness?

Data Governance and Privacy

Is the data used to train a system accurate, complete, and up to date? Was the collection of data limited to only information needed to perform a specific function or solve a specific problem? Was

personal information effectively deleted when it was no longer necessary or relevant? Were adequate steps taken to ensure that training data were not linked or reasonably linkable to an individual?

Professional Judgment
Are engineers trained and encouraged to spot and raise ethical issues? Are there specific mechanisms through which engineers and other employees can report ethical issues? Are there incentives (or disincentives) for doing so?

Implementing Ethical Review of Automated Decision-Making Systems

Creating a set of ethical principles or guidelines is a good start, but a review process must accompany it. The technology sector is beginning to recognize that conducting ex ante ethical reviews of automated decision-making systems is imperative, though industry has been slow to develop and share processes for doing so. This may be because these systems are still seen as new and experimental. The uncertainty surrounding the technology is all the more reason to shore it up with sound ethical risk assessment procedures. Even with ethical review, automated decision-making systems (like any technology) will have unintended consequences, some of them harmful. Sound internal processes can put companies in a better position to detect, remedy, and learn from harmful outcomes and avoid replicating them.

The technology at issue may be new, but the need for businesses to adopt internal processes to promote social good is not. Over the past few decades, industry and civil society have engaged in the development of process-based frameworks for putting human rights and corporate social responsibility principles into practice. These frameworks are useful for informing how companies can implement ethical principles into the design of automated decision-making systems before they are deployed in the wild.

UN Guiding Principles on Business and Human Rights

In 2011, the United Nations (UN) Human Rights Council published its Guiding Principles on Business and Human Rights,[19] based on the "Protect, Respect, and Remedy" framework developed by UN Special Representative John Ruggie. The guidance includes operational principles for businesses, including (1) making a publicly available policy commitment to respect human rights; (2) assessing actual and potential human rights impacts that the business may cause or contribute to; (3) assigning responsibility for addressing human rights impacts to the appropriate people within the business; (4) tracking the effectiveness of the business's response to human rights issues through qualitative and quantitative indicators, drawing on feedback from both internal and external sources, including affected stakeholders; (5) reporting publicly on how the business addresses human rights impacts, "particularly when concerns are raised by or on behalf of affected stakeholders"; and (6) providing remedies for adverse impacts.

Ranking Digital Rights Corporate Accountability Index

Since 2015, Ranking Digital Rights has evaluated companies' respect for freedom of expression and privacy using its Corporate Accountability Index.[20] The index includes measures of internal implementation mechanisms such as (1) employee training, (2) whistleblower programs, (3) impact assessments, (4) stakeholder engagement, (5) grievance and remedy mechanisms, and (6) public disclosure of implementation processes.

GNI Implementation Guidelines

The Global Network Initiative (GNI) Implementation Guidelines for the Principles on Freedom of Expression and Privacy provide details for how technology companies can protect and advance free expression and privacy rights.[21] The guidance includes (1) oversight by the board of directors of the company's human rights risk assessments, reporting, and response; (2) employee training; (3) impact assessments, including specific guidance building on the UN framework; (4) ensuring that business partners, suppliers, and distributors also comply with human rights principles; (5) management structures for integrating human rights compliance into business operations; (6) written procedures and documentation; (7) grievance and remedy mechanisms; (8) whistleblowing mechanisms; (9) multi-stakeholder collaboration; and (10) transparency.

Applying Human Rights Implementation Guidance to Automated Decision-Making Systems

The frameworks for implementing human rights principles share a set of common processes that can also support ethical design and use of automated decision-making systems. These common processes are (1) public commitments; (2) employee training; (3) risk assessment; (4) testing; (5) grievance and remedies mechanisms; (6) transparency measures, such as public reporting of ethical review processes or results; and (7) oversight. Here are examples of how these processes could be adapted to the automated decision-making context:

Public Commitments
Companies should make public commitments to uphold ethical principles in the design, training, review, testing, and use of the automated systems they build. These commitments should include a statement of the company's ethical principles.

Employee Training
Companies should develop rigorous ethical training for employees and consultants who build automated systems—particularly engineers and product developers—based on real-world scenarios. Training should be specialized according to the type of company, its mission, and the products it creates, and should train engineers to become adept at spotting ethical issues on their own.

Risk Assessment
Companies should develop comprehensive risk assessments designed to anticipate potential negative or disparate impacts of automated decision-making systems on all individuals and groups likely to be impacted. Potential risks include but are not limited to privacy and data security, discrimination, loss of

important opportunities (particularly in the contexts of finance, housing, employment, and criminal justice), and loss of autonomy or choice.

It is important that any risk-benefit assessments be realistic about the potential benefits of an automated decision-making system. For example, claims that collecting more data will help solve complex problems should have a fact-based rationale and should not automatically override concerns about data minimization and privacy.

Companies should conduct individualized risk assessments for each vulnerable population that could be affected by the outcomes of an automated system. The assessments should take into account historical marginalization, disparities in access to resources and justice, and other factors that might lead to harmful disparate impacts.

Risk assessments should also anticipate potential uses and abuses of an automated system by third parties. For example, a company selling an automated decision-making system to a government entity must assess the risk that the government entity will use the system in ways that the company may not have intended or in ways that create disparate impacts or violate rights.

Testing

Automated decision-making systems should be tested for limitations, such as disparate impacts on minority groups. The possibilities and best practices for testing (e.g., whether comprehensive testing before deploying an algorithm in the wild is practicable) may vary across tools and contexts. In some cases, access to the algorithm and testing by outside experts may be the best way to ensure fairness.

Grievance and Remedies

Companies should have clear and transparent mechanisms in place to receive and respond to grievances from individuals who believe they have been harmed by an automated decision-making system. For example, if a social media user believes her content was wrongfully flagged and removed by an automated tool for violating the platform's terms of service, she should be able to report the issue to the company and receive an explanation of why the content violated the terms of service and an opportunity to appeal the decision.

Transparency

Companies should be open about their ethical review processes and mechanisms when practicable. Sharing these processes can help create industry guideposts, facilitate discovery and mitigation of adverse impacts, and foster trust between companies and the public.

Oversight

Several companies have recently taken the important step of hiring a chief ethics officer and—even better—assembling a team dedicated to ethical assessment and review. These teams should have full access to information about projects and the authority to recommend changes to or even halt projects that do not meet ethical standards.

Case Study: Airbnb's Inclusion Team

During the past year, the home-sharing company Airbnb has made a number of internal changes aimed at reducing discrimination on its platform. After struggling with accounts of racial discrimination against would-be guests (those seeking accommodations), the company reviewed its practices and policies to see what structural changes might help reduce racial bias on the platform.[22] One of those changes was the creation of a permanent team of engineers, data scientists, researchers, and designers whose sole function is to advance inclusion and root out bias. A key function of this team is assessing what information in a would-be renter's profile—such as photo and name—might trigger a racially biased decision to reject the renter, and whether highlighting other information could mitigate that bias. The team is experimenting with reducing the prominence of profile photos and highlighting objective information like reservation details, reviews of would-be guests from previous hosts, and verifications. For guests who do not have reviews or verifications, the team is exploring how it can use design to improve messaging and social connections to build trust between hosts and would-be guests. The company has also overhauled its processes for receiving and responding to discrimination complaints, making it easier for users to flag and get help for potential instances of discrimination.

In this case, Airbnb is tackling individual human bias rather than algorithmic bias, but the lessons it has learned apply to automated decision-making systems as well: mitigating bias requires an internal acknowledgment of concerns, dedicated personnel, constant evaluation, and thoughtful internal mechanisms. Airbnb has acknowledged that there is no single solution to addressing bias and discrimination on its platform, and it is still experimenting with different approaches.

Conclusion

Automated decision-making systems are more than neutral lines of code. They are agents of policy, carrying out the values embedded in their design or data. They get their values from the choices made by the humans who create them—whether those value choices are conscious or not. A well-developed ethical review process is not a silver bullet for preventing unfair outcomes. But it is necessary if we hope to build systems that promote democracy, equality, and justice. ◆

Endnotes

1. *Loomis v. Wisconsin*, 137 S. Ct. 2290 (2017).

2. 881 N.W.2d 749 (Wis. 2016).

3. *See* Petition for Writ of Certiorari, *Loomis*, 137 S. Ct. 2290 (No. 16-6387), http://www.scotusblog.com/wp-content/uploads/2017/02/16-6387-cert-petition.pdf; Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

4. Angwin et al., *Machine Bias*, *supra* note 3; Julia Angwin & Jeff Larson, *Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say*, ProPublica (Dec. 30, 2016), https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say.

5. John Markoff, *How Tech Giants Are Devising Real Ethics for Artificial Intelligence*, N.Y. Times, Sept. 1, 2016, https://www.nytimes.com/2016/09/02/technology/artificial-intelligence-ethics.html.

6. Nat'l Comm'n for the Protection of Human Subjects of Biomedical & Behavioral Research, The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research (1979) [hereinafter Belmont Report], https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html.

7. *The Tuskegee Timeline*, Ctrs. for Disease Control & Prevention, https://www.cdc.gov/tuskegee/timeline.htm (last updated Aug. 30, 2017).

8. Stanford Prison Experiment, http://www.prisonexp.org/ (last visited Oct. 17, 2017).

9. Belmont Report, *supra* note 6.

10. *See, e.g.*, Kalev Leetaru, *Are Research Ethics Obsolete in the Era of Big Data?*, Forbes (June 17, 2016), https://www.forbes.com/sites/kalevleetaru/2016/06/17/are-research-ethics-obsolete-in-the-era-of-big-data/.

11. David Dittrich & Erin Kenneally, U.S. Dep't of Homeland Sec., The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research (2012), https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/menlo_report_actual_formatted.pdf.

12. *Id.* at 5.

13. *Id.* at 7.

14. *Software Engineering Code of Ethics and Professional Practice*, Ass'n for Computing Machinery (1999), http://www.acm.org/about/se-code.

15. *See, e.g.*, Org. of Econ. Co-operation & Dev., The OECD Privacy Framework 70–72 (2013) [hereinafter OECD Privacy Framework], http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

16. *See* Robert Gellman, Fair Information Practices: A Basic History (Apr. 10, 2017) (unpublished manuscript), https://bobgellman.com/rg-docs/rg-FIPshistory.pdf.

17. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, Org. of Econ. Co-operation & Dev. (1980), http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm. These guidelines were updated in 2013. OECD Privacy Framework, *supra* note 15.

18. Deepika Singhania, *This Startup Uses AI to Predict Lifestyle Disease Risks*, YourStory (Apr. 20, 2017), https://yourstory.com/2017/04/tissot-signature-innovators-club-march-winner/.

19. U.N. Human Rights Council, Guiding Principles on Business and Human Rights, U.N. Doc. HR/PUB/11/04 (2011), http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

20. *2017 Indicators*, Ranking Digital Rts., https://rankingdigitalrights.org/2017-indicators/ (last updated Sept. 14, 2016).

21. *Implementation Guidelines for the Principles of Freedom of Expression and Privacy*, Global Network Initiative, http://globalnetworkinitiative.org/sites/default/files/Implementation-Guidelines-for-the-GNI-Principles_0.pdf (last visited Oct. 17, 2014).

22. Airbnb hired Laura Murphy to lead the review, and Murphy assembled a cross-department internal team as well as outside experts. Together, they reviewed aspects of Airbnb such as how hosts and guests interact, the company's written policies, enforcement of the policies and response to complaints of discrimination, and the (lack of) diversity on the Airbnb team. Laura W. Murphy, Airbnb's Work to Fight Discrimination and Build Inclusion (2016), http://blog.atairbnb.com/wp-content/uploads/2016/09/REPORT_Airbnbs-Work-to-Fight-Discrimination-and-Build-Inclusion.pdf.

WHY CHANGES IN DATA SCIENCE ARE DRIVING A NEED FOR QUANTUM LAW AND POLICY, AND HOW WE GET THERE

BY APRIL F. DOSS

The growing complexity of issues at the intersection of technology, privacy, security, and law have brought us to a turning point similar to the revolution in scientific thinking that has taken place over the past 50 years. For centuries, scientists believed that the laws of physics described by Isaac Newton were adequate to explain the workings of the universe. By and large, for most purposes they still are. But at the edges of our understanding—in the subatomic scale and across the vastness of the universe—Newtonian physics broke down. Something in those laws didn't hold true. The evidence we were collecting made clear that those theories were inadequate to explain new scientific questions we were facing.

At the intersection of law and technology today we are facing a similar revolution. It is now possible to collect data that is so granular—subatomic particles of information, if you will—on such a massive, grand, and continuous scale—data collection and analysis that matches the scale of the universe—that our traditional approaches to law and policy struggle to make sense of what these advances mean for privacy and technology, and leave real doubts about whether law and policy can keep up. In the face of these challenges, we need a new approach, something I think of as "quantum policy." Allow me to relate a few examples of these current challenges, and explain what quantum policy could mean.

Quantum physics asks us to believe two apparently contradictory things simultaneously: that light can be both a particle and a wave, and that it can be both at the same time; or that bits in a computer can register both one and zero at the same time. In the days of Newtonian physics, propositions like these would have felt like something out of *Alice in Wonderland*, an exercise in believing impossible things. Today, though, we know that Newtonian physics cannot explain subatomic behavior or the cosmos. And that realization led to the development of a new field of physics. It helps to remember that these new theories were driven by necessity: quantum physics came about because the laws of physics we had relied on before were no longer sufficient to explain the way the universe worked.

We are at a similar turning point when it comes to applying traditional law and policy to the questions raised by algorithms, big data, and cybersecurity and privacy law. Law and policy are straining under the pressures imposed by technological advances, and a Newtonian approach is no longer sufficient to answer the questions that are pressing to be clarified, or to keep pace with the widespread scope of rapid and hard-to-predict changes.

What do I mean by a Newtonian approach to law and policy? It is one in which we are content with gradual, incremental change, where we continue to rely on the slow accretion of precedent, where we are content with having critical legal issues decided by disparate cases that take years to wend their way through multiple jurisdictions before arriving at a critical mass of new law that is achieved through an organic evolution. It means continuing to require that there be a case in controversy, refusing to allow courts to offer advisory opinions. It means that once a precedent has been established, it is extremely hard to overturn—the inertia becomes almost insurmountable.

Technology development works differently. Beta versions, user acceptance testing, and minimum viable products are the watchwords of the day, and failing fast to support rapid improvement in iterations is key.

A simple example is one that my intellectual property (IP) colleagues often point to: the timeline for obtaining traditional patent protection for a new invention has made patents on software increasingly obsolete. If the window for a new software product or technique to be cutting edge shrinks to a mere 12 to 18 months, then it may no longer make sense to wait for a patent to issue before bringing the product to market. By the time patent protection is obtained, the product itself would be obsolete. Of course, this does not apply to all IP issues, and not all the time. But this is a real and genuine concern among technology developers and business owners who are struggling to fit their more agile business model into a framework of traditional, and often slow, legal processes.

It isn't possible, of course, to map a precise one-for-one comparison between the evolution of hard science and law and policy. Instead, I am offering up quantum physics as a conceptual analogy for the ways in which we might approach the challenge of modernizing law and policy.

According to quantum mechanics, light can be a wave and a particle at the same time, and both properties can be leveraged. Relying on duality, rather than resisting it, leads to quantum gains. Quantum mechanics allows us to zoom in, to understand the behavior of things at almost unfathomably micro levels, leading to advances in miniaturization that were previously

April F. Doss *(aprilfdoss@gmail.com) was formerly the head of intelligence law for the National Security Agency (NSA), and the chair of the cybersecurity and privacy practice at the law firm Saul Ewing. She currently serves as senior minority counsel for the Senate Select Committee on Intelligence (SSCI). The views expressed here are her own, and not those of the NSA, SSCI, or any other organization.*

unimaginable. Quantum mechanics also explains the behavior of the universe at macro levels. It fills in the questions raised by Einstein's theories of relativity, and allows us to zoom out and take stock of the heavens in ways that we could not have predicted only a short time ago.

Law and policy need to embrace some analogous approaches. One of these is the ability to embrace, rather than resist, duality: not just balancing rights or ideals that seem to conflict, but encouraging them to thrive at the same time, and understanding that sometimes outcomes are probabilistic.

Law and policy need the ability to zoom in, to tackle legal and policy questions at a level of detail that many practitioners resist. How often have you heard, or perhaps said, "I'm not technical" when talking about a complex question with a client? Lawyers should no longer allow themselves the intellectual sloppiness of saying that. It doesn't take a computer science degree to embrace the intellectual challenge of understanding what data objects are, or how analytics work, or to consider the territoriality and jurisdictional questions raised by actions and information that traverse the world's communications networks. We shirk our duties when we shy away from trying to understand, at a layman's level, the technologies that are shaping our world.

Law and policy also need to zoom out, to know when it's useful to describe black holes—to look at the macro effects of an analytic, for example—instead of focusing on the bits of data like grains of sand.

We need to recognize the limits of traditional approaches to law and policy, and to look for appropriate ways to bring to bear concepts like minimum viable product, rapid iteration, and failing fast and improving often, which have been key to technological advances in the private sector. The idea of failing at anything is anathema to lawyers: partly because so many of us are ego-driven (often to unhealthy degrees), and also because we value the stability that comes with predictability in the law, and we want to avoid actions that could bring about unintended societal injustice, whether to an individual or group.

However, the slow pace of legal and policy developments in key areas such as big data and data analytics, cybersecurity, and privacy mean that in fact we are already failing to provide the certainty, guidance, and resolution of issues that justice demands. If we are honest, it shouldn't be hard to recognize the enduring wisdom in the words of William Gladstone or William Penn that "justice delayed is justice denied."[1]

A more creative, rapid, and innovative approach to these issues could allow the law to move forward more quickly—but only if it is done in a way that has appropriate safeguards built in to counterbalance the risks of uncertainty in outcomes, unintended consequences, and erroneous judgments.

Subatomic Particles and the Universe of Conclusions: Quantum Policy for Big Data Analytics

A number of years ago, when I was at the National Security Agency (NSA), I had the privilege of leading the group that was charged with developing a new approach to vetting the legal and policy ramifications of big data analytics. I would like to be able to say that a team of smart lawyers and intelligence oversight officers were able to come up with innovative solutions, but that wouldn't be accurate. Instead, what happened was that a cross-disciplinary team emerged: one that included software developers, intelligence analysts, people with expertise in data tagging and in platforms, and yes, people who were steeped in the specifics of the policy, legal, and oversight regimes.

We knew that we needed to create a framework that would be flexible enough to accommodate the wide range of analytics that would undoubtedly emerge over time. It needed to be agile enough to make decisions quickly—speed is often critical in intelligence operations, and we wanted to give people confidence that new ideas could be run through the vetting process quickly enough to meet mission needs. It needed to be adaptable over time, able to incorporate best practices as we learned them. It needed to include a definitive registry of decisions, allow periodic re-review when analytics or

legal or policy requirements changed, and be scalable so that the review process could keep pace with a demand signal that was likely to grow.

With all of these requirements in mind, we assumed that risk-based tiering made sense. We knew we would need a decision matrix, something that could be translated into codable yeses and nos. We would need an iterative approach: evaluating analytics to create the matrix, training people on the matrix once it existed, and adding new rules to the matrix every time a new analytic review prompted a new rule. This cycle of constant updates would allow our decision-making processes to continue to grow and mature. And all of this would only be possible if we created a standing framework of human interaction, documentation, and thresholds for decision making that could be repeatable, scalable, and timely, and could grow.

The traditional approach of having one of our in-house clients bring us a problem would barely have worked for vetting a single cloud analytic—the intelligence analysts could have articulated the intended mission outcome, but probably could not describe with precision the ways the data would interact with each other or provide a comprehensive overview of the legal and privacy protections built into the computing platform where the analytics would run. The technology team could tell us how the data would interact, but were not as well positioned to gauge what the impact would be on intelligence analysis if an analytic were tweaked in a particular way at our suggestion. Relatively few lawyers have the technical and operational expertise to fully evaluate the legal implications of a complex analytic involving data governed by many different sets of rules. And analytic decision making required a repeatable approach that could tackle not just the legal advice but other dimensions of the decision as well: policy approval, resource commitment, and integrating legal advice with technical review to ensure that an analytic—while it might be given a green light by the lawyers—wouldn't be run until someone else had made sure that it wouldn't end up crashing its platform.

In other words, Newtonian approaches to legal review were not enough to deal quickly or well with assessing how complex analytics would operate on small bits of data, or to address the outcomes they would drive when they were run at a large scale.

With all of that in mind, the analytics vetting team crafted a novel, interdisciplinary way of reviewing analytics, codifying the results, creating a repeatable framework, and ensuring that the ecosystem for analytic vetting could continue to evolve and change as the people using the framework learned more. In creating this analytic vetting framework, this small, interdisciplinary team took a quantum leap forward into new ways of managing complexity, volume, scale, iteration, and a host of other issues that arose with the challenge of big data analytics. In other words, quantum policy had just taken hold.

Simultaneous Belief in Two Seemingly Contradictory Things

Anyone watching national security and privacy debates can see the ways in which black-and-white thinking obscures the complexity of the issues we face. On the national security side, zealous advocates with mental blind spots sometimes fail to acknowledge that many law enforcement and electronic surveillance programs raise genuine privacy impacts. While many people might believe that some privacy intrusions are worth the increased security benefit (I count myself as among that group, having testified publicly in favor of the renewal of Foreign Intelligence Surveillance Act (FISA) section 702[2]), it would be wrong to pretend that no privacy impacts exist; the better question is whether those privacy impacts are justified by the national security gain. Similarly, some privacy advocates disregard the real benefits of national security, going so far as to praise leaks of sensitive information and lionize the leakers. This could not be more short-sighted. When the privacy community celebrates theft of data from government systems and the unauthorized release of classified information, they undermine the credibility of the important privacy values that they are trying to speak for.

When it comes to private sector activities, many privacy advocates applaud companies that refuse to cooperate with the government for law enforcement or national security purposes. National and multinational companies use their opposition to judicial warrants as a marketing tool. While multinational companies face genuinely complex dilemmas over whether and how to comply with the vast array of laws across all of the jurisdictions where they do business, privacy advocates are mistaken to think of these companies as acting purely from altruism, when the same companies carry out data collection and analysis schemes that are more comprehensive, intrusive, and unfettered than anything that is typically lawful for the government to do in Western democracies.

Cybersecurity puts the paradox in stark relief. Most privacy advocates want to see a high level of security for systems holding personal data. But in order to achieve a high level of cybersecurity, it is often necessary to implement some degree of network and user activity monitoring—which has the effect of being privacy-intrusive. Today, many companies capture detailed system, network, and user log information and run complex analytics on it, perhaps combining that data with other behavioral indicators, in order to detect and assess insider threat. Yet doing so necessarily comes at a privacy cost.

Modern technology provides us with a nearly endless supply of seemingly contradictory positions, of situations in which the on-the-one-hand and on-the-other discussion seems to go on without end. We need approaches to law and policy that will help us harmonize those tensions, rather than simply thinking of one thing versus another; we need to be able to embrace them both.

What to Do When, under Existing Laws, the Behavior of the Universe Is Hard to Predict

Although cybersecurity is a challenge for everyone, what you see depends on where you sit, and cybersecurity legal risk looks different, and less definitive, from the private sector perspective than from the government perspective.

First, in the private sector, you can get sued. It doesn't matter whether you are a multinational corporation or a small

business that happens to hold personal information—like the Social Security numbers of your employees, or the credit card numbers of your customers. Most government entities do not share that concern. In the classic government case, there is a premium on confidentiality of information, and there is high concern over integrity and availability. While both government and private entities can face devastating impacts from the loss of secrets, the government can rarely be sued. Private entities have to think about all of the cybersecurity strategies that government entities use, and they also need to incorporate an additional set of risk mitigation tools in order to manage their cybersecurity risk—tools like insurance coverage, contractual language, limitations on liability, and representations and warranties.

Second, in cybersecurity lawsuits is there is no clearly defined standard of care. Frequently, a company cannot be confident that the cybersecurity measures it has taken will be deemed to have been "enough" in the face of a breach. As a result, cybersecurity risk assessments now permeate every aspect of commercial life, from mergers and acquisitions, to cross-border data transfers of personnel and customer information, to complying with the patchwork of data breach notification laws in this country alone.[3]

Further complicating matters, companies that have been hacked often feel as though they are victimized twice: first, as the victim of a computer crime; and second, as the victim of a federal or state enforcement action or regulatory probe seeking, with perfect 20/20 vision, to determine whether their cybersecurity preparedness had been "reasonable," despite the lack of clearly defined standards. Private sector entities know it is unlikely that law enforcement will be able to reach attribution for a cyberattack—much less make indictments or arrests. And so they are reluctant to provide the government with information that could help identify cybersecurity threat trends. I often encouraged my clients to report cyberattacks to law enforcement. But there is a kernel of truth in those reservations my clients expressed: we treat cybercrime differently from other crimes. We expect the police to keep the city streets generally safe; we know that a bank that reports a theft to the police is unlikely to be held responsible for failing to stop the thief. At the same time, we know that the government cannot ensure cybersecurity for the private sector; we are not even sure that it should try. What we are left with is an odd tension in which hacking is a crime, but the victim shoulders part of the blame. (And if the victim is fined, those penalties are likely to be paid to the state, rather than as restitution to the second order of victims, such as customers or patients whose information has been breached.) We suspect that some kind of public-private partnership has to be central to solving the cybersecurity conundrum, but we are having an awfully hard time figuring out how to get there.

What Will the Future Bring?

The trends in technology and networking are driving us toward more widespread harvesting of information and increasingly complex ways of using it that range from consequential to comical. In just the past year in the private sector, here are a few examples of things we have seen:

- In California, an employee sued her company for requiring her to install an app on her phone that would track her location even when she was off work.
- Complex data analytics are being used to influence decisions by parole boards, and also to prioritize the text messages that come into a crisis hotline.
- A committee in the U.S. House of Representatives approved legislation that would allow employers to require employees to undergo genetic testing and grant employer access to those results along with other health information.[4]
- Litigation was brought against the manufacturer of an app-controlled sex toy, alleging invasion of privacy because the company did not tell its customers that it was harvesting data from the app about what settings were used, when, and how often.
- We know that there are widespread insecurities in the Internet of Things, along with widespread adoption of in-home devices from smart refrigerators to audio-powered personal in-home assistants. In Germany, children's toys with cameras and artificial intelligence (AI) were recalled because they were capable of spying.
- Biometrics are increasingly used for identification, and people are volunteering to have implanted microchips that can be used for everything from tracking the time they spend at work to opening security doors and presenting their mass transportation passes.

Private sector litigation over data breaches keeps growing: suits against companies by individual customers and patients; derivative lawsuits by shareholders against directors and officers for failing to ensure effective cybersecurity measures were in place; and regulatory or enforcement actions taken by governments. Insurance companies have started underwriting policies for property damage and personal injury that result from a cyberattack with impacts in the physical realm: when a car's steering system is hijacked, or a dialysis machine is shut down, or traffic lights are switched to show green in all directions, or an app that's supposed to control a home cooktop is hijacked by a malicious user to turn on a stove, in turn causing a fire that burns down a house.

Consumers still say they want privacy, and they also still value the benefits that come from commercial products that harvest their information for digital personal assistants and in-home security systems, more accurate driving directions, the fun of locating friends online, interactive children's toys, or other AI. Many consumers do not read privacy notices or understand them, and governments struggle to decide where the line is between appropriate consumer protection and economic-strangling paternalism.

Through it all, law and policy will struggle to keep up, because the pace of technology change is limited only by ingenuity and imagination, and to a much lesser extent by the laws of electrical

engineering and physics. In other words, recent years have shown us a world in which the challenges of big data, privacy, and information security grow more complicated, more multidimensional, and more interconnected. And those challenges are greatest when dealing with the extraordinarily detailed types of information now available about nearly everyone, and in dealing with the macro-level conclusions being drawn from analytics that are assessing that information.

How Quantum Policy Can Help

Just as quantum physics paved the way for technologies from which the entire world benefits, quantum policy can help support the effort to democratize data science, privacy, and related technology.

From its inception, the Internet has been a great democratizing force, making unprecedented volumes of information available to millions of people all over the world, often for free (or included in the price of a cell phone service plan). Free webmail services made global communication cheap, easy, and practically instant for billions of people around the globe who would never previously have considered sending international letters or telegrams. Voice over Internet Protocol (VoIP) connections made real-time conversations with voice and video available to billions of people who never could have afforded international long-distance phone calls. Web hosting, blogs, YouTube, you name it—nearly everything necessary for the exchange of ideas and commerce have been put into place by the innovation engine of the Internet, at a price that makes these tools accessible on a scale that could never have been imagined before. Data has never been more ubiquitous or available. At the same time, it is impossible to escape the potential negative impacts: the same AI that promises rapid innovation, the same genetic information that promises new medical treatments, and the same convenience of personalized traffic recommendations can also be used to bring about unprecedented invasions of privacy. And the same access to information that supports democracy can also be subverted to cement the iron rule of authoritarian regimes.

The profusion of data and ways to manipulate it, the paradox of privacy in an era of living online, the question over who can misuse information and how, or how to think about consumer choice in a time when we freely give away data but do not always understand how it is used—all of these dilemmas are forcing us to acknowledge the limits of our current legal and policy approaches to them. If we do not modernize our approaches to law and policy, they will continue to lag sorely behind the pace of technological change, with real-life consequences—some of them unintended—for individuals, organizations, and governments. Here are a few examples of what quantum policy could look like:

- We need to admit that we cannot put the tech genie back in the bottle or reseal Pandora's box; legal originalism has limited effectiveness in judicial review of these kinds of matters.
- We need to avoid the temptation to use scaremongering, or else we will miss the opportunity for clear and rational discussions about the ways technology can be privacy-protective.
- We need to educate the public without paternalism or condescension.
- We need to teach technology in law schools, teach privacy in technology schools, and increase the use of cross-disciplinary teams.
- We need to take rigorous approaches to data risk, being sure we understand where the biggest data really is, and assign accountability to both the government and the private sector accordingly.
- We need to pursue cost-effective data security, realigning penalties to harm.
- We need to consider duality: For example, if the encryption and going-dark debate falters because of black-and-white thinking, we should consider alternatives outside of that binary box. Perhaps encryption and back doors are not the only ways to achieve the goals of keeping data private and secure, while allowing the government to access it for a legitimate purpose.
- We need to focus privacy policy on the end goals and sensitivity of the information, rather than focusing on how it was acquired (commercial purchase, government warrant, etc.), and adopt policies that avoid unnecessarily placing the United States at a disadvantage against other nations, either in national security or economic terms.
- We should consider ways to implement test cases and incubator environments for legal and policy evolution.

When it comes to dealing with legal and policy issues for emerging technologies, we are still largely living in a Newtonian age. If we do not make the leap to quantum policy, our entire ecosystem of jurisprudence, litigation, legislation, intellectual property, and privacy rights will suffer as a result. Quantum physics did not evolve overnight, and neither will quantum policy. But by looking at examples of innovations that have worked, and by assembling cross-disciplinary teams to think in creative ways about the challenges that face us, we can move toward solutions that allow the law, and lawyers, to keep up. ◆

Endnotes

1. The quote has been variously attributed to both.

2. *Section 702 of the Foreign Intelligence Surveillance Act: Hearing Before the H. Comm. on the Judiciary*, 115th Cong. 20 (2017) (statement of April F. Doss, Partner, Saul Ewing LLP).

3. At last count, 48 states had their own data breach laws. These laws have different definitions of an actionable breach; they impose different timelines for notifying victims; some require notifying state government agencies and others do not; and they have different requirements for the information to be included in consumer and regulator notifications.

4. Sharon Begley, *House Republicans Would Let Employers Demand Workers' Genetic Test Results*, PBS NewsHour (Mar. 11, 2017), http://www.pbs.org/newshour/rundown/house-republicans-let-employers-demand-workers-genetic-test-results/.

IN MEMORIAM

CHARLES RAY "CHAS" MERRILL

BY STEPHEN S. WU AND MICHAEL S. BAUM

Charles Ray "Chas" Merrill, 76, passed away on April 25, 2017, in Denver, Colorado. In his practice, he served as a partner in McCarter & English's Newark, New Jersey, office for many years. After graduating from Harvard Law School in 1965, he began his career in the area of tax law, estate planning, and corporate law. He received a tax LLM from New York University in 1970. However, he later moved into information security and electronic commerce law. He served as an active member of the Section's Information Security Committee (ISC) in the 1990s and 2000s before he retired in 2006.

In his work for the Section, he was a pioneer, an intellect, and a patient mentor. He provided wise counsel to ISC leadership, and was a dear friend to many in the ISC. In the earliest days of the development of e-commerce—when it was still called electronic contracting—Chas quickly distinguished himself as a thought leader in the frenetic world of digital signature law and policy. Chas was a co-reporter and key leader in producing two of the Section's path-breaking publications on public key infrastructure: *Digital Signature Guidelines* and *PKI Assessment Guidelines*.

As the ISC leader for these two publications, Chas maintained a steady (and often paternal) hand in leading the many (sometimes vexing) drafting initiatives. He often exercised his unique personal charm and humor to help maintain a shared mission and organize diverse legal and nonlegal professionals into teams that produced sections of the *Digital Signature Guidelines* and *PKI Assessment Guidelines*. These two publications were the heart of the ISC's work plan for many years, and both publications had a worldwide influence on digital signature laws and industry practices for many years.

Chas also helped the ISC present programs at the RSA Conference, the world's leading information security conference. For instance, he acted as the judge in the very first mock trial that the Section cosponsored with the RSA Conference in 2007. Audience members raved about the mock trial program in their reviews.

Chas had a special ability to inspire and lead teams of people, an activity that sometimes resembled "herding cats." His leadership also included an over 20-year involvement as an adult Scout leader for the Boy Scouts of America. He received the Silver Beaver Award from his local council, the highest award a council can bestow on an adult leader.

Denley Chew, an attorney working at the Federal Reserve Bank of New York, wrote, "Chas was an early pioneer of this new emerging field of 'information security law' who really helped make the ISC into what it is today—intellectual, social, constructive—and for that I think we will always be grateful *and* aspirational."

Chas's work is for the ages. He was a true gentleman who will be sorely missed.

Stephen S. Wu *(ssw@svlg.com) served as the 2010–2011 Chair of the ABA Section of Science & Technology Law.* ***Michael S. Baum*** *(michael@secureav.com) served as the founding Chair of the Section's Information Security Committee and Electronic Commerce Division.*

Section of Science & Technology Law
American Bar Association
321 North Clark Street
Chicago, Illinois 60654-7598
PC545800001401

Nonprofit
Organization
U.S. Postage
PAID
American Bar
Association

CONNECTWITHSCITECH

#SCITECH

Each year, thousands of members count on SciTech to put the tools they need at their fingertips to navigate rapidly changing science and technology law issues:

THREE INFORMATION-PACKED PERIODICALS:

- *The SciTech Lawyer*, our glossy quarterly magazine with cutting-edge coverage
- *SciTech e-Merging News*, our quarterly electronic newsletter with timely practice perspectives and Section activities/opportunities
- *Jurimetrics*, our quarterly electronic law review published by the Section and the Center for Law, Science & Innovation of the Sandra Day O'Connor College of Law at Arizona State University

---

- Unlimited access to over 20 free hot-topic committees and list serves
- Free access to Section articles via the SciTech e-Archive and ABA Web Store
- Members-only discounts on Section books (save 10% or more) and CLE programs (save $100 off the public rate, $50 off the ABA rate)
- Exclusive career and business development resources
- Free access to our podcast archive on emerging issues
- Selected chapters from preeminent SciTech books (delivered electronically) and access to chapter archive
- Two free webinars/teleconferences on hot topics (notices sent electronically)
- Limited-time special offers

CALENDAROFEVENTS

November 9, 2017
ASCENDING STARS OF ABA SCITECH:
HOT TOPICS IN 60 MINUTES
Teleconference (Free to SciTech)

November 16, 2017
MEDICAL DEVICES AND DIGITAL HEALTH:
IDENTIFYING OPPORTUNITIES AND MANAGING RISKS
New York, NY

November 17, 2017
DRIVING THE NEXT GENERATION:
LEGAL UPDATE ON SELF-DRIVING CARS
CLE Webinar

February 1–2, 2018
SCITECH MEETINGS AND EVENTS
AT ABA MIDYEAR MEETING
Vancouver, BC