FRANK PASQUALE

NEW LAWS OF ROBOTICS

DEFENDING HUMAN EXPERTISE IN THE AGE OF AI

Printed in the United States of America

First printing

Design by Sam Potts

9780674250048(EPUB)
9780674250055(MOBI)
9780674250062(PDF)

The Library of Congress has cataloged the printed edition as follows:

Names: Pasquale, Frank, author.
Title: New laws of robotics : defending human expertise in the age of AI / Frank Pasquale.
Description: Cambridge, Massachusetts : The Belknap Press of Harvard University Press, 2020. | Includes bibliographical references and index.
Identifiers: LCCN 2020012258 | ISBN 9780674975224 (cloth)
Subjects: LCSH: Robotics—Law and legislation. | Artificial intelligence—Law and legislation. | Virtual humans (Artificial intelligence)
Classification: LCC K564.C6 P375 2020 | DDC 343.09/99—dc23
LC record available at https://lccn.loc.gov/2020012258

CONTENTS

Education is the point at which we decide whether we love the world enough to assume responsibility for it, and by the same token save it from that ruin which, except for renewal, except for the coming of the new and young, would be inevitable. And education, too, is where we decide whether we love our children enough not to expel them from our world and leave them to their own devices, nor to strike from their hands their chance of undertaking something new, something unforeseen by us, but to prepare them in advance for the task of renewing a common world.

—*Hannah Arendt,* Between Past and Future

I am speaking of a law, now, understand,
that point at which bodies locked in cages
become ontology, the point at which
structures of cruelty, force, war,
become ontology. The analog
is what I believe in, the reconstruction
of the phenomenology of perception
not according to a machine,
more, now, for the imagination to affix to
than ever before.

—*Lawrence Joseph, "In Parentheses"*

NEW LAWS OF ROBOTICS

1

▪ Introduction

The stakes of technological advance rise daily. Combine facial recognition databases with ever-cheapening micro-drones, and you have an anonymous global assassination force of unprecedented precision and lethality. What can kill can also cure; robots could vastly expand access to medicine if we invested more in researching and developing them. Businesses are taking thousands of small steps toward automating hiring, customer service, and even management. All these developments change the balance between machines and humans in the ordering of our daily lives.

Avoiding the worst outcomes in the artificial intelligence (AI) revolution while capitalizing on its potential will depend on our ability to cultivate wisdom about this balance. To that end, this book advances three arguments that stand to improve all our lives. The first is empirical: right now, AI and robotics most often complement, rather than replace, human labor. The second proposes a value: in many areas, we should maintain this status quo. And the final point is a political judgment: our institutions of governance are actually capable of achieving exactly that outcome. Here is this book's most basic premise: we now have the means to channel technologies of automation, rather than being captured or transformed by them.

These ideas will strike many as commonsensical. Why write an entire book to defend them? Because they have some surprising implications, which should change how we organize social cooperation and deal with conflict. For example, at present, too many economies favor capital over labor and consumers over producers. If we want a just and sustainable society, we must correct these biases.

That correction will not be easy. Ubiquitous management consultants tell a simple story about the future of work: if a machine can record and imitate what you do, you will be replaced by it.[1] A narrative of mass unemployment now grips policymakers. It envisions human workers rendered superfluous by ever-more-powerful software, robots, and predictive analytics. With enough cameras and sensors, this story goes, managers can simulate your "data double"—a hologram or robot that performs your job just as well, at a fraction of your wages. This vision offers stark alternatives: make robots, or get replaced by them.[2]

Another story is possible and, indeed, more plausible. In virtually every walk of life, robotic systems can make labor more valuable, not less. This book tells the story of doctors, nurses, teachers, home health aides, journalists, and others who work *with* roboticists and computer scientists, rather than meekly serving as data sources for their future replacements. Their cooperative relationships prefigure the kind of technological advance that could bring better health care, education, and more to all of us, while maintaining meaningful work. They also show how law and public policy can help us achieve peace and inclusive prosperity rather than a "race against the machines."[3] But we can only do so if we update the laws of robotics that guide our vision of technological progress.

ASIMOV'S LAWS OF ROBOTICS

In the 1942 short story "Runaround," the science fiction writer Isaac Asimov delineated three laws for machines that could sense their environment, process information, and then act.[4] The story introduces a "Handbook of Robotics, 56th Edition," from 2058, that commands:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

A good definition of professions is capacious, and it should include many unionized workers, particularly when they protect those they serve from unwise or dangerous technologies. For instance, teachers' unions have protested excessive "drilling and testing" via automated systems and have promoted their students' interests in many other contexts. Unions that tend toward a path of professionalization—empowering their members to protect those they serve—should have an important role in shaping the AI revolution.

Sometimes it will be difficult to demonstrate that a human-centered process is better than an automated one. Crude monetary metrics crowd out complex critical standards. For example, machine learning programs may soon predict, based on brute-force natural language processing, whether one book proposal is more likely than another to be a best seller. From a purely economic perspective, such programs may be better than editors or directors at picking manuscripts or film scripts. Nevertheless, those in creative industries should stand up for their connoisseurship. Editors have an important role in publishing, exercising judgment, and finding and promoting work that the public may not (now) want, but needs. The same could be said of journalists; even if automated text generation could generate ad-maximizing copy, that hollow triumph should never replace genuine reporting from an authentic, hard-won, human point of view. Professional schools in universities clarify and reexamine standards in media, law, medicine, and many other fields, preventing them from collapsing into metrics simple enough to be automated.

Even in fields that seem most subject to the automation imperative, in areas like logistics, cleaning, agriculture, and mining, workers will play a critical role in a long transition to AI and robotics. Gathering or creating the data necessary for AI will be a demanding task for many. Regulations can make their jobs more rewarding and self-directed. For example, European privacy law empowers drivers to resist the kind of 360-degree surveillance and control that oppresses truckers in the United States.[11] That is not to say that such a dangerous occupation should go unmonitored. Sensors may indicate problems with a driver's reflexes. But there is a world of difference between sensors specifically aimed at safety lapses, and a constant video and audio recording of all actions. Getting the balance right between unnerving and demeaning surveillance and sensible, focused monitoring will be crucial in a wide range of fields.

We can also design technological transitions that keep human beings in the picture, or at least give them that choice. For example, Toyota has promoted cars with a spectrum of machine involvement, from chauffeur mode (which requires minimal monitoring by a driver) to guardian mode (which focuses the car's computing systems on accident avoidance, while a person helms the vehicle).[12] Planes have had autopilot capacities for decades, but commercial carriers still tend to have at least two people in the cockpit. Even the occasional airline passenger can be grateful that the evangelists of substitutive automation are in no hurry to jettison pilots.[13]

Note, too, that transportation is one of the easier cases for AI. Once a destination is set, there is no argument over the point of a trip. In other service fields, the opposite is true: the customer's or client's mind may change. A class may be too antsy on a beautiful spring day to practice multiplication tables repeatedly. A socialite may call his interior designer, worried that the print they chose for the living room is too brash. A trainer may vacillate, worried that her client is too exhausted to run another minute on a treadmill. In each of these cases, communication is key, as are the human-to-human skills of patience, deliberation, and discernment.[14]

Yes, if thousands of trainers equipped themselves with Google Glass and recorded all their encounters, perhaps some divine database of grimaces and rolled eyes, injuries, and triumphs could dictate the optimal response to a miserable gym goer. But even to begin imagining how to construct such a database—what gets marked as a good or bad outcome, and to what extent—is to understand the critical role people will play in constructing and maintaining a plausible future of AI and robotics. Artificial intelligence will remain artificial because it will always be a product constructed out of human cooperation.[15] Moreover, most recent advances in AI are designed to perform specific tasks rather than to take on entire jobs or social roles.[16]

There are many examples of technologies that make jobs more productive, more rewarding, or both. As the Agency for Digital Italy has observed, “Technology often does not completely replace a professional figure but replaces only some specific activities.”[17] Contemporary law students can barely believe that pre-internet lawyers had to comb through dusty tomes to assess the validity of a case; research software makes that process easier and vastly expands the range of resources available for an argument. Far from simplifying matters, it may make them much more complex.[18] Spending less time hunting down books and more time doing the intellectual work of synthesizing cases is a net plus for attorneys. Automation can bring similar efficiencies to myriad other workers, without mass displacement of labor. And this is not merely an observation. This is a proper goal of policy.[19]

2. Robotic systems and AI should not counterfeit humanity.

From Asimov’s time to the vertiginous mimicry of *Westworld,* the prospect of humanoid robots has been fascinating, frightening, and titillating. Some roboticists aspire to find the right mix of metal bones and plastic skin that can break out of the “uncanny valley”—the queasiness that a humanoid robot evokes when it comes very close to, but does not quite recreate, human features, gestures, and ways of being. Machine-learning programs have already mastered the art of creating pictures of “fake people,” and convincing synthetic voices may soon become common.[20] As engineers scramble to fine-tune these algorithms, a larger question goes unasked: Do we want to live in a world where human beings do not know whether they are dealing with a fellow human or a machine?

There is a critical difference between humanizing technology and the counterfeiting of distinctively human characteristics. Leading European ethicists have argued that “there have to be (legal) limits to the ways in which people can be led to believe that they are dealing with human beings while in fact they are dealing with algorithms and smart machines.”[21] Lawmakers have already passed “bot disclosure” laws in online contexts.

Despite this growing ethical consensus, there are subfields of AI—such as affective computing, which analyzes and simulates human emotion—devoted to making it more and more difficult for us to distinguish between humans and machines. These research projects might culminate in a creation like the advanced androids in the Steven Spielberg film *A.I.*, indistinguishable from a human being. Ethicists debate how such humanoid robots should be designed. But what if they should not be made at all?

In hospitals, schools, police stations, and even manufacturing facilities, there is little to gain by embodying software in humanoid bodies, and plenty to lose. The race to mimic humanity might too easily become a prelude to replacing it. Some people might prefer such replacement in private life, and law should respect such autonomy in intimate realms. But the idea of a society dedicated to advancing it in workplaces, the public sphere, and beyond is madness. It confuses the abolition of humanity with its advance.

This argument may jar or confound technophiles—to reject not merely the substance, but also the premise, not only of Asimov's laws, but also of a vast literature on the future of technology. I hope to justify this conservatism by thinking through, chapter by chapter, the concrete steps we would need to take to get to a science-fictional world of robots indistinguishable from humans. That transition entails massive surveillance of humans, all to create robots designed to fool or allure human beings into treating machines as their equals. Neither prospect is appealing.

The voice or face of another human being demands respect and concern; machines have no such claim on our conscience. When chatbots fool the unwary into thinking that they are interacting with humans, their programmers act as counterfeiters, falsifying features of actual human existence to increase the status of their machines. When the counterfeiting of money reaches a critical mass, genuine currency loses value. Much the same fate lies in store for human relationships in societies that allow machines to freely mimic the emotions, speech, and appearance of humans.

The counterfeiting of humanity is a particular danger as corporations and governments seek to put a friendly face on their services and demands. Google Assistants have wowed the business press by mimicking secretaries making appointments, eerily replicating even the "ums" and "ahs" that punctuate typical phone conversations. These conversational fillers disguise the power of a firm like Google with the hesitation or deference typically expressed by a human's unpolished speech. They cloak a robocall as a human inquiry. For those on the receiving end of the calls, it is all too easy to imagine abuse: a deluge of calls from robotized call centers.

Counterfeiting humanity is not merely deceptive, it is also unfair, giving the counterfeiter the benefit of the appearance of personal support and interest without its reality. As we will see in case after case—of robot teachers, soldiers, customer-service representatives, and more—dissatisfaction and distress at failed imitations of humanity are not merely the result of imperfect technology. Rather, they reflect wise caution about the direction of technology itself.

3. Robotic systems and AI should not intensify zero-sum arms races.

Debates over "killer robots" are a central theater for ethics in international law. A global coalition of civil society organizations is pushing nations to pledge not to develop lethal autonomous weapons systems (LAWS). Several factors now stymie this commendable proposal for technological restraint. Military leaders distrust their counterparts in rival countries. They may hide militarized AI research, advancing in power even as they publicly disclaim any such intent. Rising powers may assert themselves, investing in force projection to match their new economic status, while now-dominant militaries press for more resources to maintain their relative advantage. This is but one of many ways an arms race begins. As AI and robotics enter the picture, the stakes of falling behind one's rivals rise, since emerging technologies promise to be so much more targeted, ubiquitous, and rapidly deployed.

Dovish politicians may commit themselves to a purely defensive posture (reflected in the United States' shift from a Department of War to a Department of Defense in 1949). But defenses can often be repurposed as offensive weapons; think, for instance, of autonomous drones designed to destroy missiles but reprogrammed to assassinate generals. Thus, even protective plans can seem aggressive, as in the case of Ronald Reagan's Strategic Defense Initiative (SDI). Popularly known as Star Wars, SDI would have relied on lasers in space to shoot down Soviet missiles. Had it worked, it would have upset a fragile balance of deterrence (mutually assured destruction via nuclear annihilation). Now, LAWS, Of course, some robots and algorithms will evolve away from the ideals programmed into them by their owners as a result of interactions with other persons and machines (think, for instance, of advanced self-driving cars that evolve as a result of multiple influences).[23] In such cases, there may be multiple potentially responsible parties for any given machine's development and eventual actions.[24] Whatever affects the evolution of such machines, the original creator should be obliged to build in certain constraints on the code's evolution to both record influences and prevent bad outcomes. Once another person or entity hacks or disables those constraints, the hacker is responsible for the robot's wrongdoing.

For a concrete application of this principle, consider a chatbot that gradually learns certain patterns of dialogue from interactions on Twitter. According to some news accounts, Microsoft's AI chatbot, Tay, quickly adopted the speech patterns of an unhinged Nazi sympathizer after only a few hours on Twitter.[25] Microsoft did not program that outcome, but it should have known that it was a danger of exposing a bot to a platform notorious for its poor moderation of harassment and hate speech. Moreover, to the extent that the chatbot did log where the malign influences came from, it could have reported them to Twitter, which, in some better version of itself, could have taken some action to suspend or slow the flood of abuse coming from troll accounts and worse.

Regulators will need to require responsibility-by-design (to complement extant models of security-by-design and privacy-by-design). That may involve requiring certain hard-coded audit logs, or licensing practices that explicitly contemplate problematic outcomes.[26] Such initiatives will not simply regulate robotics and AI *post hoc*, but will influence systems development by foreclosing some design options and encouraging others.[27]

▪ ▪ ▪

Each of these new laws of robotics, promoting complementarity, authenticity, cooperation, and attribution, rests on a theme that will animate the rest of our exploration: the critical distinction between technology that replaces people and technology that helps them do their jobs better. The point of the new laws is to develop policies that capitalize on human strengths in fields such as health and education, and to take advantage of human limits to bound the scope and intensity of conflict and regimentation in social life.

AI researchers have long aimed to create computers that can sense, think, and act like humans. As far back as the 1960s, roboticists at MIT were developing robot sentries to relieve soldiers of the boring and dangerous duty of standing guard at vulnerable sites.[28] But there is another way to think about the sentry robot—not as AI replacing troops, but as one more tool to increase soldiers' effectiveness as defenders. An army does not necessarily need to requisition more soldiers to monitor emerging threats. It can instead develop sensors and computers designed to act as a second set of eyes and ears, rapidly processing threat levels and other data to better inform soldiers' actions. This goal, deemed "intelligence augmentation" (IA), has informed the projects of many internet pioneers.[29] It is also a mainstay of modern warfare, as drone pilots handle a rich array of sensor data to make life-and-death decisions about aerial bombings.

Though sometimes blurry, the distinction between AI and IA is a critical one for innovation policy. Most parents are not ready to send their children off to robot teachers. Nor should their children be taught that teachers will eventually be replaced by machines perfectly personalized to their learning styles. There are many visions of robotics in education that are far more humane. For example, schools have already experimented successfully with "companion robots" that help young students by drilling them on vocabulary words and asking them questions about what they just learned. Looking like animals or fanciful creatures rather than people, these robots do not challenge the distinctiveness of humanity.

Researchers are finding that in many contexts, IA results in better service and outcomes than either artificial or human intelligence working alone. Assistive AI and robotics could be a godsend for workers, freeing more hours for rest or leisure. But in any modern market economy, there are economic laws that tilt the scale toward AI and against IA. A robot does not demand time off, a fair wage, or health insurance. When labor is viewed primarily as a cost, its fair pay is a problem, which machines are supposed to solve. Robotics revolutionized manufacturing by replacing assembly-line workers. Now, many business experts want similar technological advances to take over more complex work, from medicine to the military.

Caught up in these managerialist enthusiasms, too many journalists have discussed "robot lawyers" and "robot doctors" as if they are already here. This book will show that such descriptions are overblown. To the extent that technology transforms professions, it has tended to do so via IA, not AI. Submerged beneath breathless headlines about "software eating the world," there are dozens of less spectacular instances of computation helping attorneys or doctors or educators to work faster or better.[30] The question now for innovation policy is where to sustain this predominance of IA, and where to promote AI. This is a problem we must confront sector by sector, rather than hoping to impose a one-size-fits-all model of technological advance.

Conversations about robots usually tend toward the utopian ("machines will do all the dirty, dangerous, or difficult work") or the dystopian ("and all the rest besides, creating mass unemployment"). But the future of automation in the workplace—and well beyond—will hinge on millions of small decisions about how to develop AI. How far should machines be entrusted to take over tasks once performed by humans? What is gained and lost when they do so? What is the optimal mix of robotic and human interaction? And how do various rules—from codes of professional ethics to insurance policies to statutes—influence the scope and pace of robotization in our daily life? Answers to these questions can substantially determine whether automation promises a robot revolution or a slow, careful improvement in how work is done.

▪ ▪ ▪

Why should we be especially concerned with robots and AI, as opposed to the ubiquitous screens and software that have already colonized so much of our time? There are two practical reasons. First, the physical presence of a robot can be far more intrusive than any tablet, smartphone, or sensor; indeed, those technologies can simply be embedded into robots.[31] No flat-screen could reach out and restrain a misbehaving child or a defiant prisoner, modifying and repurposing present technology of crowd control for new forms of discipline. But a robot could.

Even if uptake of robots is slow or limited, AI threatens to supercharge the techniques of fascination and persuasion embedded into tech ranging from mobile apps to video poker.[32] As human-computer interaction researcher Julie Carpenter observes, “Even if you know a robot has very little autonomy, when something moves in your space and it seems to have a sense of purpose, we associate that with something having an inner awareness or goals.”[33] Even something with as little animation as a robot vacuum cleaner can provoke an emotional response. The more sensors record our reactions, the richer the veins of emotional data for more sophisticated computers to mine.[34] Every “like” is a clue to what engages us; every lingering moment on a screen is positive reinforcement for some database of manipulation. Miniaturized sensors make surveillance mobile, unraveling efforts to hide. Indeed, the decision to shield oneself from sensors may be one of the most revealing activities one can engage in. Moreover, processing capacities and data storage could put us on a path to a dystopia where everything counts, and anything a student does may be recorded and backed up to inform future evaluations.[35] By contrast, a student in an ordinary school may encounter different teachers every year, starting off with a relatively clean slate each time.[36]

None of these troubling possibilities is destined to happen, though, which raises a second reason to focus on robotics policy now. As robots enter highly regulated fields, we have a golden opportunity to shape their development with thoughtful legal standards for privacy and consumer protection. We can channel technology through law.[37] Robots need not be designed to record every moment of those whom they accompany or supervise. Indeed, robot supervision itself could seem sufficiently oppressive that we require some human monitoring of any such system (as a robotic South Korean prison mandated for its mechanical guards). When robots are part of a penal system, a broad, rich debate on prison policy and on the relative merits of retribution and rehabilitation should inform any decision to deploy them. Indeed, one of the main points of the new laws of robotics is to warn policymakers away from framing controversies in AI and robotics as part of a blandly general “technology policy,” and toward deep engagement with domain experts charged with protecting important values in well-established fields.

Cynics may scoff that such values are inherently subjective and that they are destined for obsolescence in an ever more technological society. But communities of scholars and consultants focused on science, technology, and human values have shown that anticipatory ethics can inform and influence technology design.[38] Values are designed into technology.[39] Canadian, European, and American regulators have already endorsed privacy-by-design as a basic principle for developers.[40] Rules like that should apply a fortiori to sensor-laden technology, which can move freely to maximize its ability to record images and sound. For example, in the same way that many video cameras have a red light indicating that they are recording, robots should feature an equivalent indicator when they record persons around them. AI-driven data should be subject to strict limits on collection, analysis, and use.[41]

Technologists may counter that it is too early to regulate robotics. Let problems develop and only then move to counter them, say the partisans of laissez-faire. But quietism misses the mark. All too often in high technology fields, industry says it is *never* a good time to regulate. When troubling new business practices emerge, would-be regulators are accused of strangling an “infant industry.” Once the practices are widespread, the very fact of their ubiquity is offered as proof that consumers accept them. For every argument offered for legal intervention, there is a pat strategy of deflection based on bromides and platitudes. “Is there really a problem?,” “Let’s wait and see,” and “Consumers want it” are all offered as all-purpose rationales for inaction, played like trump cards in a game of whist.[42]

A wait-and-see attitude ignores the ways in which technology, far from being independent of our values, comes to shape them.[43] Robotic companions for children in online charter schools would not merely reflect (or distort) current values regarding what kinds of socialization are owed to the young. They would also shape those values of those generations, inculcating them with a sense of what types of moments are private and what are fair game for potentially permanent recording. These mores should not simply result from whatever is most profitable that their owners want to "feed" them. In the realm of pure computing unconnected to social consequences, that right might be respected. All manner of irresponsible speech is permitted in the name of free expression; software programmers can assert a similar right to enter data into programs without regard to its social consequences. But as soon as algorithms—and especially robotics—have effects in the world, they must be regulated and their programmers subject to ethical and legal responsibility for the harms they cause.[53]

PROFESSIONALISM AND EXPERTISE

Who gets to decide what this responsibility entails? A smooth and just transition will demand both old and new forms of professionalism in several key areas. The concept of expertise commonly connotes a mastery of a certain body of information, but its actual exercise involves much more.[54] For those who conflate occupational duties with mere knowledge, the future of employment looks grim. Computers' capacity to store and process information has expanded exponentially, and more data about what individuals do during their workday is constantly accumulating.[55] But professionalism involves something more complex: a recurrent need to deal with conflicts of values and duties, and even conflicting accounts of facts.[56] That makes a difference to the future of work.

For example, imagine that you are driving down a two-lane road at forty-five miles per hour, cruising home. You see a group of children walking home from school about a hundred yards ahead. Just as you're about to pass by them, an oncoming eighteen-wheeler swerves out of its lane and is about to hit you head on. You have seconds to decide: sacrifice yourself, or hit the children so you can avoid the truck.

I like to think that most people would choose the nobler option. As the automation of driving advances, such self-sacrificial values can be coded into vehicles.[57] Many cars already detect whether a toddler in a driveway is about to be run over by a driver with a blind spot. They even beep when other vehicles are in danger of being bumped. Transitioning from an alert system to a hardwired stop is technically possible.[58] And if that is possible, so is an automatic brake that would prevent a driver from swerving for self-preservation at the expense of many others.

But the decision can also be coded the other way—to put the car occupants' interests above all others. Although I do not think that that is the correct approach, the correctness of the approach is beside the point for our purposes. The labor question addresses how engineers, regulators, and marketers, as well as government relations and sales professionals, work together to shape human-computer interactions that respect the interests of everyone affected by automation, while also respecting commercial imperatives. There are few one-shot problems in design, marketing, and safety. As technology advances, users adapt, markets change, and new demands are constantly arising.

The medical profession has long been faced with such dilemmas. Doctors' jobs are never limited to merely taking care of the cases before them. Obliged to understand and monitor risks and opportunities that are constantly shifting, doctors must keep track of where medicine is headed, staying current about studies that either confirm or question mainstream medical knowledge. Consider even a decision as trivial as whether to give an antibiotic to a patient with a sinus infection. A good primary-care doctor must first decide whether the drug is clinically indicated. Doctors may take subtly different positions on how robust their obligation is to conserve antibiotic prescriptions to slow the evolution of resistant microbes. They also need to keep track of the prevalence of potential side effects of antibiotics, such as the sometimes devastating infections caused by *Clostridium difficile*—and the varying likelihood of such effects on different types of patients. Patients have some awareness of all these things when they visit a physician, but they are not responsible for coming to a correct decision or melding all these judgment calls into a recommendation for a particular case. That is the professional's role.

For true believers in the all-encompassing power of big data, predictive analytics, algorithms, and AI, the "brains" of robots can hack their way around all these problems. This is a tempting vision, promising exponential technological progress to raise living standards. But is it a realistic one? Even systems based purely in the digital realm—such as search-engine algorithms, high frequency trading, and targeted advertising—have proven in numerous cases to be biased, unfair, in-accurate, or inefficient.[59] Information is much harder to capture accurately in the wild, and there are disputes over what should be measured in the first place. Stakes rise considerably higher when algorithmic systems are empowered as the brains of robots that can sense their environment and act upon it. Meaningful human control is essential.

Nor is such human control only necessary in fields such as medicine, which has a long history of professional self-governance. Even in transport, professionals will have critical roles for decades to come. However fast robotic driving advances, the firms developing it cannot automate the social acceptance of delivery drones, sidewalk wagons, or cars. As legal expert Bryant Smith has observed, lawyers, marketers, civil engineers, and legislators must all help prepare society as a whole for the widespread deployment of such technologies.[60] Governments need to change their procurement policies, both for vehicles and infrastructure. Local communities must make difficult decisions about how to manage the transition, since stop lights and other road features optimized for human drivers may not work well for robotic vehicles, and vice versa. As Smith observes, "Long-term assumptions should be revisited for land-use plans, infrastructure projects, building codes, bonds, and budgets."[61]

The labor required for such a transition will be vast and diverse.[62] Security experts will model whether vehicles without human passengers pose special risks to critical infrastructure or to crowds. Terrorists do not need to recruit a suicide bomber if they can load a self-driving car with explosives. Public health experts will model the spread of infectious disease if such vehicles include strangers. Legislators are already grappling with the question of whether to require such vehicles to revert control to a person upon request or to give that control to police when they demand it.[63] I used the ambiguous term "person" in the last sentence because we still do not have a good term for occupants of a semi-autonomous vehicle. Both law and norms will shape that new identity over time.[64]

None of these decisions should be made solely—or even predominantly—by the programmers and corporations developing algorithms for self-driving cars. They involve governance and participation by a much wider range of experts, ranging from urban-studies scholars to regulators to police and attorneys. Negotiations among affected parties are likely to be protracted—but that is the price of a democratic and inclusive transition toward new and better technology. And these are only a few of the ethical, legal, and social implications of a widespread transition to self-driving cars.[65]

Nevertheless, some futurists argue that AI obviates the need for professions. With a large enough set of training data, they argue, virtually any human function can be replaced with a robot. This book takes the exact opposite view: to the extent our daily lives are shaped by AI and machine learning (often run by distant and massive corporations), we need more and better professionals. That is a matter of affirming and extending the patterns of education and licensure we already have in fields like medicine and law. And it may require building entirely new professional identities in other critical sectors where both wide public participation and expertise are essential.

TWO CRISES OF EXPERTISE

Asserting humans' value as a form of expertise may be jarring to some readers. At present, the most popular argument against AI's encroachment on the governance of workplaces and cities is a democratic appeal. AI's critics argue that technical experts in topics like machine learning and neural networks are not diverse enough to represent the persons their technology affects.[66] They are too removed from local communities. The same could be said of many other experts. There is a long and distinguished history of activists complaining about aloof doctors and professors, incomprehensible lawyers, and scientists detached from the quotidian problems of everyman. Confronted about economists' predictions of disastrous consequences from Brexit, British politician Michael Gove asserted that "people in this country have had enough of experts."[67] As that sentiment fuels populist campaigns around the world, there is a deepening chasm between politics and expertise, mass movements and bureaucratic acumen, popular will and elite reasoning.

Commenting on such trends, the sociologist Gil Eyal argues that expertise is a way of talking about the "intersection, articulation, and friction between science and technology on the one hand, and law and democratic politics on the other."[68] This is indeed a venerable tension in administration, where bureaucrats must often make difficult decisions implicating both facts and values. For example, raising or reducing pollution limits is a decision with medical consequences (for the incidence of lung cancer), economic impact (on the profitability of enterprise), and even cultural significance (for the viability of, say, mining communities). Eyal focuses on a democratic challenge to pure technocratic decision-making on each of those fronts.

This book examines a different, and distinct, challenge to expertise—or, more precisely, a clash of forms of expertise. Well-credentialed economists and AI experts have asserted that their ways of knowing and ordering the world should take priority almost everywhere, from hospitals to schools, central banks to war rooms. Few are as blunt as a former technology company CEO who remarked to a general, "You absolutely suck at machine learning. If I got under your tent for a day, I could solve most of your problems."[69] But the general theme of many books on AI-driven automation and economic disruption is that the methods of economics and computer science are *primus inter pares* among other forms of expertise. They predict (and help enact) a world where AI and robotics rapidly replace human labor, as economics dictates cheaper methods of getting jobs done. On this view, nearly all workers will eventually share the fate of elevator operators and horse-and-buggy drivers, just waiting until someone with adequate data, algorithms, and machinery replaces us.

To be sure, there are areas where economics and AI are essential. A business cannot run without covering its costs; a self-checkout lane will fail if its scanner program cannot recognize product bar codes. But the questions of whether a business should exist or a cashier should be replaced with a robot kiosk cannot be answered by economics or computer science alone. Politicians, communities, and businesses decide, based on complex sets of values and demands. These values and demands cannot simply be reduced to equations of efficiency and algorithms of

Monetary Fund, experts sternly warn that tens of millions of jobs are about to be replaced by robots.[75] Focused on our role as producers, this is a discussion framed by gloom and urgency. Field after field, it seems, is set to be automated—first routine tasks, then more professional roles, and then even the work of coding itself once a "master algorithm" has been found.[76] Reporting on this literature can be apocalyptic. "Robots to steal 15 million of your jobs," blasted the *Daily Mail,* trumpeting a study touted by Bank of England governor Mark Carney.[77] While estimates of job loss vary widely, the economic literature's drumbeat is unmistakable: every worker is at risk.

Simultaneously, economists celebrate the cheapening of services. The model of economic progress here is eerily similar to the one featured in automation narratives. Leaders in the health care and education sectors are supposed to learn from the successes of the assembly line in manufacturing, as well as data-driven personalization in the internet sector. Dialectically templatized and personalized approaches to health and education are supposed to make hospitals and schools cheaper and eventually make the best of their services available to all.[78]

Combine "robots are taking all the jobs" dystopianism with "ever-cheaper services" utopianism, and you have a bifurcated vision of our economic future. The workplace is destined to become a Darwinian hellscape, where employees are subordinate to machines that record their every movement to develop robotic replicants. The only consolation comes after hours, when the wonders of technology are supposed to make everything cheaper.

This model of miserable workers and exultant consumers is not merely troubling—it is unsustainable. Taken individually, a reduction in labor costs seems like a good thing. If I can replace my dermatologist with an app and my children's teachers with interactive toys, I have more money to spend on other things. The same goes for public services; a town with robot police officers or a nation with drone soldiers may pay less taxes to support their wages and health care. But doctors, teachers, soldiers, and police are all potential purchasers of what others have to sell. And the less money that they have, the less money I can charge them. In classical economic terms, the great worry here is deflation, a self-reinforcing spiral of lower wages and prices.

Even in the most self-interested frame, the "cost" of goods and services to me is not a pure drain on my well-being. Rather, it is a way of reallocating purchasing power to empower those who helped me (by creating whatever I am buying) to eventually help themselves (to perhaps purchase what I make or do). To be sure, a universal basic income would make up some of the purchasing power of those put out of work by robotics. But it is unrealistic to expect redistribution to do anything near the work of "pre-distribution" in assuring some balanced pattern of economic reward. Most democratic electorates have been cutting the relative tax liability of the richest for decades.[79] Robotization is unlikely to change that dynamic, which unravels ambitious plans for redistributing wealth.

Regarding the economy as an ongoing ecology of spending and saving, and as a way of parceling out power over (and responsibility for) important services, gives us a better perspective on the robotics revolution. Traditional cost-benefit analysis tends to dictate a rapid replacement of humans by machines, even when the machines' capabilities are substandard. The lower the cost of a service, the greater its benefits appear by comparison. But once we understand the benefits of cost itself—as an accounting of effort and an investment in persons—the shortcomings of this simplistic, dyadic view of the economy becomes clearer. The penultimate chapter further develops the benefits of the costs of programs and policies recommended in the rest of this book.

PLAN OF THE BOOK

Too many technologists aspire to rapidly replace human beings in areas where we lack the data and algorithms to do the job well. Meanwhile, politicians have tended toward fatalism, routinely lamenting that regulators and courts cannot keep up with technological advance. The book will dispute both triumphalism in the tech community and minimalism among policymakers in order to reshape public understanding of the state's role in cultivating technological advance. I offer policy analysis that shows the power of narrative and qualitative judgment to guide the development of technology now dominated by algorithmic methods and quantitative metrics. The point of the book is to distill accumulated knowledge from many fields and vantage points and present it to the public for its use. Ideally, it will lay a foundation for what Alondra Nelson calls "anticipatory social research," designed to shape, and not merely respond to, technological advance.[80]

The translation of tasks into code is not a purely technical endeavor. Rather, it is an invitation to articulate what really matters in the process of education, caregiving, mental-health care, journalism, and numerous other fields. Although there is a temptation to simply set forth quantifiable metrics of success in all those fields and to optimize algorithms to meet them (whether via trial and error, crunching past data, or other strategies), the definition of what counts as success or failure in such fields is highly contestable. Basing a decision on one metric excludes all others. No one can be "driven" by data in general. Particular data matter, and the choice of what to count (and what to dismiss as non-representative) is political.

Among AI ethicists, tension has developed between pragmatists (who focus on small and manageable reforms to computational systems to reduce their discriminatory or otherwise unfair judgments) and futurists, who worry about the rise of out-of-control systems and self-improving AI (which could, it is feared, rapidly grow "smarter," or at least more lethal, than their human creators). The pragmatists tend to dismiss the futurists as haunted by phantoms; the futurists think of the pragmatists' concerns as small bore. I believe each side needs the other. The horrific outcomes predicted by futurists are all the more likely if we do not aggressively intervene now to promote transparency and accountability in automated systems. But we are unlikely to take on that hard task if we fail to reckon with the fundamental questions about human nature and freedom that the futurists are asking.

These questions are not new. For example, in 1976, computer scientist Joseph Weizenbaum asked, "What human objectives and purposes may not be appropriately delegated to computers? . . . The question is not whether such a thing can be done, but whether it is appropriate to delegate this hitherto human function to a machine."[81] But the queries

"Can robots be better than humans?" or "When should humans not use robotics?" are incomplete. Almost anyone, in any job, is already using some level of automation on a continuum between simple tool and human-replacing AI. A better frame is, "What sociotechnical mix of humans and robotics best promotes social and individual goals and values?"

In a series of case studies, I answer that question concretely by making the case that AI supplementing, rather than replacing, human expertise realizes important human values. Chapters 2, 3, and 4 describe what that process might look like in health care, education, and media, focusing on the first new law of robotics: the need for technology to complement, rather than replace, existing professionals.

I am, on the whole, optimistic about the prospects for complementary automation in health and education. Patients and students by and large demand human interaction.[82] They realize that however advanced AI becomes, it is enormously helpful to obtain guidance on how to use it from experts who study the reliability of various sources of knowledge daily. Even more importantly, in so many care or learning contexts, human relations are intrinsic to the encounter. Robotic systems can provide technical support, improving judgments and developing entertaining and engaging drills. Perhaps rural and disadvantaged areas will demand them as substitutes for now-absent professionals. But this dictate of necessity is far from an exemplary labor policy. It is particularly troubling when it comes to mental health care for vulnerable populations.

When there is a nurse, teacher, or doctor at the point of contact of AI—to mediate its effects, assure good data collection, report errors, and do other vital work—there is far less chance of a grimly deterministic future in which we are all poked and prodded into learning or wellness by impersonal machines. Professionals in health and education also owe clear and well-established legal and ethical duties to patients and students. These standards are only beginning to emerge among technologists. Thus, in the case of media and journalism—the focus of Chapter 4—a concerted corrective effort will be necessary to compensate for what is now a largely automated public sphere.

When it comes to advertising and recommendation systems—the lifeblood of new media—AI's advance has been rapid. Reorganizing commercial and political life, firms like Facebook and Google have deployed AI to make the types of decisions made by managers at television networks or editors at newspapers—but with much more powerful effects. The reading and viewing habits of hundreds of millions of people have been altered by such companies. Disruption has hit newspapers and journalists hard. And it has been terrifying for some vulnerable groups, including minorities targeted for harassment. The only way to stem an epidemic of fake news, digital hate campaigns, and similar detritus is to bring more responsible persons back in to guide the circulation of online media.

Whereas Chapter 4 focuses on AI's failures in judging the value of news, Chapter 5 describes the perils of using AI to judge people. Computation is playing an ever larger role in hiring and firing, as well as in the allocation of credit and the treatment of debt. It is also creeping into security services. I warn against the rapid adoption of robotic policemen and security guards. Even predictive policing, a wholly software-driven affair supervised by officers, has proven controversial thanks to its frequent reliance on old and biased data. Machines sent out on their own to patrol neighborhoods or rustle panhandlers off sidewalks are even more troubling. Nor are many other applications of AI in civil law ready for widespread adoption. They promote a rule of machines over persons, which sacrifices human dignity on the altar of efficiency.

Chapter 6 takes this concern about power to traditional and online battlefields. Debate on lethal autonomous weapon systems has taken on a familiar structure. Abolitionists call for a ban on killer robots, realists reject that approach, and reformers occupy a middle ground by proposing regulation short of an outright ban. Abolitionists and reformers are now engaged in spirited arguments about the value of each side's approach. But the two groups' strategies may eventually harmonize. Reformers acknowledge that some types of weapons systems are so dangerous that they should never be built. Abolitionists concede that some defensive uses of automation (particularly in cyberwarfare) are necessary for national security.

2

▪ Healing Humans

There are two AI dreams in medicine. The first is utopian, straight out of science fiction novels. Care robots will spot and treat any disease, instantly. Nanobots will patrol our veins and arteries, busting clots and repairing damaged tissues. Three-dimensional printed organs, bone, and skin will keep us all looking and feeling young well into our eighties and nineties. With enough luck, even brains can be uploaded for perpetual safekeeping, with robotic bodies sleeving indestructible minds.[1]

Whatever its long-term merits, that sci-fi vision is far, far off—if it ever arrives at all. More realistic medical futurists are still ambitious, but they offer more attainable visions. They recognize the critical role that human empathy plays in care, that human insight contributes to diagnoses, and that human dexterity adds to surgery. They by and large embrace the first new law of robotics, promoting a future where AI is primarily aiding (rather than replacing) doctors and nurses. That is wise, reflecting a realistic assessment of the current state of technology and data.[2]

Unfortunately, even many realists tend to stumble when it comes to policy and law. They see health care systems through a primarily economic lens, lamenting their expense and inefficiencies. They advocate deregulation to spur innovation and budget limits to force cost cutting. But what we really need in medical technology policy is more responsibility to collect and use the best data—not less. And we need to invest in the cutting edge of medical practice, rather than simply assuming that hospitals and doctors will come up with ever more ingenious ways of doing more with less.[3]

Science fiction writers dream of a day when a combination of apps and robots can take care of all our medical needs. But this is not the current path of leading medical technological developments—nor should policymakers intervene to make it our lodestar. The stakes of medical care, as well as the psychological stress of illness, counsel in favor of a lasting human presence in the deployment of health-sector automation. While economic imperatives will pressure hospitals and insurers to substitute software for therapists and bots for nurses' attention, professional associations should ensure that cost considerations are balanced against the many virtues of direct human involvement in care.

DECIDING WHEN TO SEEK HEALTH CARE

Imagine waking up with a piercing stomach pain. Is it appendicitis? Bloating? A pulled muscle? Stomach pain is one of the hardest differential diagnoses for even seasoned emergency doctors. Abdominal agony could result from any one of dozens of conditions, ranging from the trivial to the life threatening.[4] Even a slight risk of a disastrous outcome would seem to counsel in favor of a trip to the hospital to get some professional advice—stat.

For the wealthy or well insured, the decision can be an easy one. For others, though, it could be ruinous to seek help. In the developing world, medical bills may threaten families' ability to meet basic needs.[5] In the United States, millions are either uninsured or underinsured. A single trip to the ER can cost well over $10,000; even a false alarm can run into the thousands of dollars once tests, physician fees, and other charges are taken into account. Even for those with adequate insurance and ample financial resources, a hospital visit poses the risk of unnecessary tests, exposure to viruses, and hours of inconvenience.

For many, the first place to look for information about sudden symptoms is Google. And for many years, Google saw medical searches—like those prompted by a sudden pain in the middle of the night—as little different than other searches. As long as they had enough "Google juice" (that mysterious mix of relevance and authority that powers content to the top of search results), sites of dubious reliability might mix with information from established doctors or medical schools. The burden was on Google users to separate the wheat from the chaff, discerning the credibility of sites.

By 2016, the company had revised its approach.[6] It collaborated with experts at the Mayo Clinic to vet information that appeared in common health searches.[7] Type in "headache on one side," and above or next to a standard list of search results, there appear a series of boxes, each briefly describing a possible classification for the headache. Pick any one of them (say, "tension headache"), and you find another box, again attributable to Google itself. It describes in very brief terms whether the condition is common, how common it is in various age groups, and what types of medical intervention might be helpful.

These new Google results are a heartening sign for artificial intelligence in health care. They do not reflect a company dead set on replacing the expertise of doctors with big data and algorithms. Rather, professionals are invited in, to help devise structured approaches to health care information and the health system itself. Similarly, IBM has shifted the marketing of its Watson system in health care and law, billing it as more of a helper for than a replacement of doctors.[8] When I talked to a representative of IBM's Watson team in 2017, he said that they promote a vision of augmented, not artificial, intelligence. As even the firms with the most to gain from AI marketing shift toward an IA (intelligence augmentation) approach, the dream of a wholly automated diagnostic tool may soon seem more anachronistic than futuristic. There will always be a place for domain experts to evaluate the accuracy of AI advice, and to assess how well it works in the real world.

AI'S CORE COMPETENCE: AVOIDING COMMON ERRORS

Doctors are expert pattern recognizers. We expect a dermatologist to tell us whether a mole is malignant or merely a beauty mark; we suffer colonoscopies to give gastroenterologists a chance to spot (and excise) polyps. Yet even the best doctor can make a mistake, and average physicians may become bored or distracted at critical moments. Thanks to AI, we may greatly reduce those types of errors, saving thousands of lives per year.

The method depends on massive amounts of data. A database may include labeled images of millions of different abnormalities that eventually became cancerous, as well as millions that did not. As we might search on Google for websites matching a query, a computer can rapidly compare images of your colon or skin with those in the database. Ideally, machines learn to spot "evil digital twins"—tissue that proved in the past to be dangerous, which is menacingly similar to your own.[9]

This machine vision—spotting danger where even experienced specialists might miss it—is far different from our own sense of sight. To understand machine learning—which will come up repeatedly in this book—it is helpful to compare contemporary computer vision to its prior successes in facial or number recognition. When a facial recognition program successfully identifies a picture as an image of a given person, it is matching patterns in the image to those in a preexisting database, perhaps on a 1,000-by-1,000-pixel grid. Each box in the grid can be identified as either skin or not skin, smooth or not smooth, along hundreds or even thousands of binaries, many of which would never be noticeable by the human eye. And even more sensing is available. Medical images may also encode data at the pixel or voxel (3-D pixel) level that map to what our hands, nose, or ear might sense—and far more.

Pattern recognition via machine vision found early commercial success with banks, which needed a way to recognize numbers on checks (given the wide variety of human handwriting). With enough examples of written numbers and computational power, this recognition can become nearly perfect. Thus, machine vision is "superhuman" in many respects, in terms of both "ingesting" data and comparing those data to millions of other images. A dermatologist might use a heuristic to diagnose melanoma (such as ABCDE, for asymmetry, borders that are irregular, color that is varied, large diameter, and evolving), or his or her experience of past cancerous and non-cancerous moles. A sufficiently advanced AI can check any of those ABCDE parameters against other moles with exceptional precision—so long as the data are accurate. Moreover, as sensors grow more advanced, AI may find unexpected sources of insight on what distinguishes malignant from benign moles.

Machine vision has “subhuman” aspects as well, and can exhibit surprising fragility.[10] Most of its applications in medicine now are “narrow AI,” focused on a particular task and that task alone. Narrow AI for detecting polyps, for example, might “see” a problem polyp that no gastroenterologist would, but it might also be incapable of recognizing other abnormalities that it was not trained to detect. Joint work in diagnostics—involving both an AI program and a doctor—is more valuable than either working alone.[11]

Physicians train for years, but medical knowledge never stops advancing. It is no longer humanly possible to memorize every potential interaction between drugs—particularly in complex cases, where patients may be taking twenty or more medications. Pharmacists can play a role in stopping bad outcomes, but they, too, can overlook unusual problems.[12] Integrated into electronic health records, clinical decision support software (CDSS) is an early form of AI that can help physicians avoid terrible outcomes.[13] CDSS “monitors and alerts clinicians of patient conditions, prescriptions, and treatment to provide evidence-based clinical suggestions.”[14] There is already evidence that CDSS reduces errors.[15] Yet even in this relatively straightforward area of information provision, programmers, managers, and engineers did not simply impose CDSS onto medical practice. Law has played a major role in its diffusion, including government subsidies to support such systems. The threat of malpractice lawsuits (for doctors) or corporate liability (for hospitals) counsels in favor of adopting CDSS; however, courts have also recognized that professional judgment cannot be automated, and they are loath to make failure to follow a machine recommendation an automatic trigger for liability if there are sound reasons for overriding the CDSS.[16]

Ongoing regulation will be critical here to assure that patients will have the benefit of cutting-edge technology, without burdening doctors and nurses with information overload. Several authors have chronicled the problem of “alert fatigue.”[17] Human-computer interaction experts are working to assure a better balance between alerts and subtler reports about potential problems. The ideal CDSS software should be neither overbearing nor merely a quiescent watcher of practitioners. It can only fulfill its promise if its messages are continually calibrated to hold Microsoft responsible for ransom notes written as an MS Word document, which is a blank slate. Nor are parents responsible for the crimes of their adult children, who are independent entities.

When they assert that they are not responsible for their creations, leading developers of AI benefit from both the "blank slate" and "independent entity" metaphors. Given a decade of research on algorithmic accountability, neither justification should immunize such firms. We all now know that algorithms can harm people.[28] Moreover, lawyers have grappled with the problem of malfunctioning computers for decades, dating back at least to the autopilot crashes of the 1950s and the Therac-25 debacle of the 1980s (when a software malfunction caused tragic overdoses of radiation).[29]

Nevertheless, some proposals would severely diminish the role of courts in the AI field, preempting their traditional role in assigning blame for negligent conduct. Others would kneecap federal regulatory agencies, leaving it up to judges to determine remedies appropriate for accidents. Even if such legal "reforms" never happen, firms could limit or shift their liability via infamous terms of service that consumers "agree to" as contracts. Finally, free expression absolutists argue that when AI is simply "saying" things about persons rather than doing things to them, it should be treated like free speech and be immune from lawsuits. Advocates for these four horsemen of irresponsibility—sweeping preemption, radical deregulation, broad exculpatory clauses, and opportunistic free expression defenses—argue that AI will develop rapidly only if inventors and investors are free from the threat of lawsuits.

Bewitched by the promise of innovation, policymakers may be tempted to sweep away local laws in order to give industry leaders an immediate, very clear picture of their legal obligations.[30] Or they may "empower" AI users to contract away their rights to sue. The perverse case for contractual sovereignty here is that my right to give away my rights advances my autonomy. A less strained, utilitarian rationale is that is that citizens need to give up some rights so that AI can flourish.

Even if liability shields are needed to spur some innovation, they must not be absolute. As Wendy Wagner has observed, tort litigation is critical to exposing information that may be blocked from regulators.[31] When regulation is harmonized internationally or for a nation, more local entities should also be empowered to develop their own standards for the level of risk they are willing to accept from new technology.[32] While that granular litigation and regulation moves forward, higher-level authorities have the resources and time frame to map out broad trends in technological development and to solicit expert advice. For example, the US National Committee on Vital and Health Statistics (where I began a four-year term in 2019) offers policymakers expert advice on how data are optimally collected, analyzed, and used. That advice is critical because in well-ordered societies, regulators help shape technology's development (and do not merely react to it once it is created).[33]

Moreover, courts and legislatures should be wary of exculpatory clauses, limiting when consumers can sign away their rights. Judges have frequently been unwilling to recognize such clauses in the medical context, reasoning that patients are vulnerable and lack the information necessary for a truly informed choice.[34] We all stand in a similar position of vulnerability when it comes to most robotics and AI, since we are almost never privy to the data and codes behind them. Even where exculpatory clauses are allowed, there is still an important role for courts to play in policing unfair terms.[35] And there are certain types of causes of action that should be preserved, whatever terms contracting parties are willing to agree to.

In order to assess risks responsibly, both vendors and users need accurate accounts of the data used in AI (inputs) and data on its performance (outputs). No one should be allowed to contract away the right to inspect those data when AI causes harm.[36] The next section describes how regulators can help assure better inputs of data to AI-driven innovation and how they can promote quality outputs from such technology.

WHO WILL TEACH THE LEARNING HEALTH CARE SYSTEM?

We might once have categorized a melanoma simply as a type of skin cancer. But that is beginning to seem as outdated as calling pneumonia, bronchitis, and hay fever "cough." Personalized medicine will help more oncologists gain a more sophisticated understanding of a given cancer as, say, one of a number of mutations. If they are properly combined, compared, and analyzed, digitized records could indicate which combination of chemotherapy, radioimmunotherapy, surgery, and radiation has the best results for that particular subtype of cancer. That is the aspiration at the core of "learning health care systems," which are designed to optimize medical interventions by comparing the results of natural variations in treatments.[37]

For those who dream of a "Super Watson" moving from conquering *Jeopardy* to running hospitals, each of these advances may seem like steps toward cookbook medicine implemented by machine. And who knows what's in the offing a hundred years hence? In our lifetime, what matters is how all these data streams are integrated, how much effort is put into that aim, how participants are treated, and who has access to the results. These are all difficult questions, but no one should doubt that juggling all the data will take skilled and careful human intervention—and plenty of good legal advice, given complex rules on health privacy and human subjects research.[38]

To dig a bit deeper in radiology: the imaging of bodily tissue is rapidly advancing. We've seen the advances from X-rays and ultrasound to nuclear imaging and radiomics.[39] Scientists and engineers are developing ever more ways of reporting what is happening inside the body. There are already ingestible pill-cams; imagine much smaller, injectable versions of the same.[40] The resulting data streams are far richer than what came before. Integrating them into a judgment about how to tweak or entirely change patterns of treatment will take creative, un-systematizable thought. As radiologist James Thrall has argued,

> The data in our . . . information system databases are "dumb" data. [They are] typically accessed one image or one fact at a time, and it is left to the individual user to integrate the data and extract conceptual or operational value from them. The focus of the next 20 years will be turning dumb data from large and disparate data sources into knowledge and also using the ability to rapidly mobilize and analyze data to improve the efficiency of our work processes.[41]

Richer results from the lab, new and better forms of imaging, genetic analysis, and other sources will need to be integrated into a coherent picture of a patient's state of illness. In Simon Head's thoughtful distinction, optimizing medical responses to the new volumes and varieties of data will be a matter of *practice,* not predetermined *process.*[42] Both diagnostic and interventional radiologists will need to take up difficult cases anew, not as simple sorting exercises.

Given all the data streams now available, one might assume that rational health policy would deepen and expand the professional training of radiologists. But it appears that the field is instead moving toward commoditization in the US.[43] Ironically, radiologists themselves have a good deal of responsibility here; to avoid nightshifts, they started contracting with remote "nighthawk" services to review images.[44] That, in turn, has led to "dayhawking" and to pressure on cost-conscious health systems to find the cheapest radiological expertise available—even if optimal medical practice would recommend closer consultations between radiologists and other members of the care team for both clinical and research purposes. Government reimbursement policies have also failed to do enough to promote advances in radiological AI.[45]

Many judgment calls need to be made by imaging specialists encountering new data streams. Presently, robust private and social insurance covers widespread access to radiologists who can attempt to take on these challenges. But can we imagine a world in which people are lured into cheaper insurance plans to get "last year's medicine at last year's prices"? Absolutely. Just as we can imagine that the second tier (or third or fourth or fifth tier) of medical care will probably be the first to include purely automated diagnoses.

Those in the top tier may be happy to see the resulting decline in health care costs overall; they are often the ones on the hook for the taxes necessary to cover the uninsured. But no patient is an island in the learning health care system. Just as ever-cheaper modes of drug production have left the United States with persistent shortages of sterile injectables, excluding a substantial portion of the population from high-tech care will make it harder for those *with* access to such care to understand whether it's worth trying.[46] A learning health system