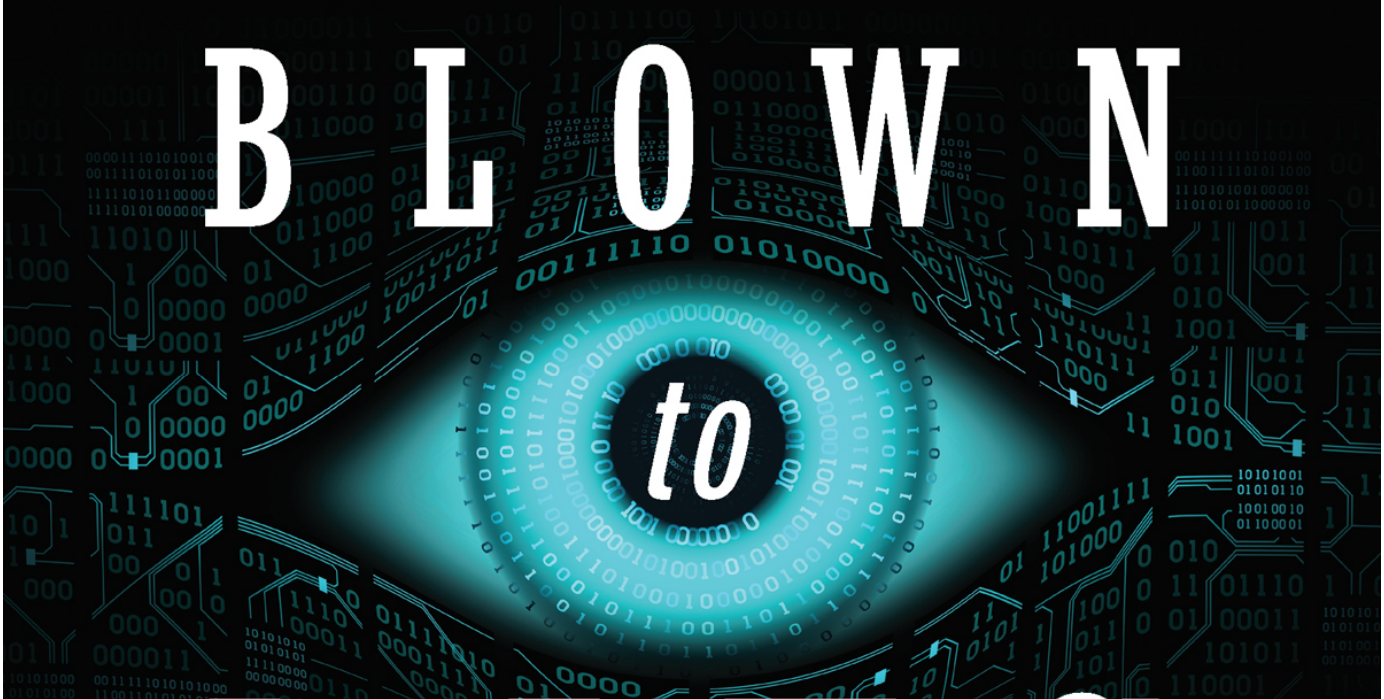
HAL ABELSON • KEN LEDEEN
HARRY LEWIS • WENDY SELTZER

BLOWN to BITS



[YOUR Life, Liberty, and Happiness After the Digital Explosion]

SECOND EDITION

Blown To Bits

Your Life, Liberty, and Happiness After the Digital Explosion

Second Edition

Hal Abelson
Ken Ledeen
Harry Lewis
Wendy Seltzer

CHAPTER 9

The Next Frontier

AI and the Bits World of the Future

In Chapter 1, "Digital Explosion," we met Nicolette, the woman who did not get a second job interview after her first interview was conducted by a computer program. Nicolette's frustration with not knowing the basis for the decision against her is a canonical manifestation of misgivings about the very nature and use of artificial intelligence. In the past, many thinkers have speculated on the potential of intelligent devices and have formulated abstract principles that could be used to guide their performance. In 1950, science fiction writer Isaac Asimov posited three "Laws of Robotics":

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

These simple rules interact in surprising ways, especially in the hands of an expert writer of fiction. But the world has moved on from imaging what might happen to experiencing what does.

We close our tour of the bits world by probing some of the dilemmas and opportunities posed by our breathtaking and yet limited technological success.

Thrown Under a Jaywalking Bus

In Ningbo, in eastern China, executive Dong Mingzhu, the "iron lady[2]" chair of the country's largest air conditioner manufacturer, was featured on a billboard with the caption "lawbreaker." Intersections in China's big cities often feature glittering LED screens. Plenty of them feature advertising, but they are also used for the latest in traffic policing: Live-action photographs combine with facial recognition software to name and shame jaywalkers caught crossing against the light. Dong Mingzhu wasn't a scofflaw. She wasn't even there at the time. The photo displayed there had indeed captured her face crossing the intersection—in an advertisement on the side of a bus.[3]

Ms. Dong escaped with a laugh, but those who don't have her standing or public support can find errors in "artificially intelligent" systems much more damaging. This small event tells us quite a bit about the state of artificial intelligence (AI), about its risks and potential. It portends a future in which machines of all sorts are capable not only of thinking and deciding but of acting without our knowledge or control. Just when we had come to the point in our story where we thought we understood the implications of the digital explosion, we realize that we are entering a whole new world.

How did this happen? At a basic level, the system that caught Ms. Dong worked exactly as it was designed to do: A camera trained on the intersection caught a picture of a person in the crosswalk when the light was against pedestrians crossing, matched the face with a massive database, and—in real time—displayed the infraction for all the world to see. There is quite a lot to explore here. This "system" interacted directly with its environment, watched

what was happening in the intersection, knew when the light was red or green, detected the presence of a person—or at least what it mistakenly thought was a person—matched up the photo against images of everyone it knew, decided what to do next, and took action.

This brief moment, seconds at most, captures many aspects of artificial intelligence, algorithmic decision making, privacy, ethics, fallibility, inherent bias, transparency, and accountability. We will explore all this and more. But first, let's consider some basics: artificial intelligence, machine learning and its cousin deep learning, and algorithmic decision making—where the pieces all come together.

What's Intelligent About Artificial Intelligence?

It has long been a dream of those working with technology to build machines that can learn and ultimately outgrow their programmers. Even Homer, some 28 centuries ago, imagined the divine blacksmith Hephaestus assisted by robots:

> In support of their master moved his attendants. These are golden, and in appearance like living young women. There is intelligence in their hearts, and there is speech in them and strength, and from the immortal gods they have learned how to do things. These stirred nimbly in support of their master.[4]

The concept of AI is often attributed to Alan Turing, the great British mathematician whose 1948 paper on intelligent machinery established a vision for machines that could mimic human behavior in areas as varied as games, language learning, cryptography, and mathematics. In 1950, in a paper titled "Computing Machinery and Intelligence,"[5] Turing introduced the concept of the imitation game (which we now refer to as the Turing test) as a way to measure whether machines could be considered "intelligent."

Seventy years after Turing's prescient vision, this form of AI behavior is commonplace. Six-year-old kids say "Alexa, show me a movie of Darth Vader" and expect the device to recognize who they are, understand the request, impose the suitable parental controls, and show the video. We take language comprehension—something completely inconceivable not all that long ago—entirely for granted. Our phones, cars, and TV remotes are capable of understanding continuous speech, placing it in context, and taking appropriate actions.

Machines that mimic human behavior—recognizing speech, understanding language, even passing the Turing test—are AI 1.0. Turing's earliest musings posited the basis of AI 2.0—machine learning. His 1948 report on intelligent machinery explored the notion that machines could learn from experience and modify themselves, much as humans do.

The Turing Test

British computer scientist Alan Turing proposed this test for intelligent machines: A human judge asks questions of two contestants, one computer and one human, communicating only through written messages. If the judge is unable to distinguish the computer from the human, the computer "passes" the test. While the Turing test was written as a thought experiment, it has sparked continued discussion and actual competitions. So far, at least, no computer has fooled a majority of its human judges.

Turing's notion spawned a whole range of chatbots—programs that interact with people through text. The first chat-bot, Joseph Weizenbaum's 1965 Eliza,[6] played a psychotherapist. Eliza convinced some users it was human despite using a simple algorithm that repurposed the question text and threw in some generalities when it couldn't do so. For example:

Men are all alike.

IN WHAT WAY?

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE?

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

Machine Learning: I'll Figure It Out

Programmers write code to tell computers what to do. At least that's the way it has always been. The history of computer science has largely been the history of finding new algorithms to get computers to do increasingly complex and useful things, along with making the machines themselves faster and cheaper. Machine learning (ML) changes all that.

With ML, rather than programmers writing code to do something, the programmers write code that enables the computer to learn. The implications are profound. Before ML, you could reasonably ask how a computer program arrived at a particular result. Not so in this brave new world. The ML code the programmers wrote provided the framework. That piece of software was then exposed to data from which it "learned" how to do the task at hand: translate from Russian to English, decide who should get four years in prison and who

should get eight, buy or sell some stocks, or stomp on the brakes and avoid a car wreck. Most computer software arrives at a definitive result. Give it all the info about what you earned last year, and it will compute what you owe in taxes. ML programs make their best guesses about something they haven't seen before, based on what they have seen in the past.

At heart, ML systems are programs that observe and predict. Is this message spam? Does this photograph depict a cat or a dog? Is the person asking for a loan a good credit risk? Who is the person crossing the street while the light is red? Programmers generate mathematical models, train them on data for which they know the answers, and then apply them to unknown inputs. The trick of machine learning is that you don't start from scratch every time but build a general model, an artificial neural network, that you can give more specific input filters and train to solve new kinds of problems.

For each new task (classify an email as spam or a news story as fake), the developers of an ML solution decide what information should be considered. In the case of email, it might include the sender's email address, subject, a library of key phrases ("make money fast," "grow more hair"), list of known spammers, and more. Next, the system is trained by being allowed to process a set of previously categorized inputs—emails that are known to be good or known to be spam. The software adjusts the weight, a measure of importance, that it gives to each of the characteristics it is considering. In use, that learning process often continues every time the user tells the software that it made the wrong decision.

That's the simplest form of machine learning. Programmers code up the basic rules, and then the software adjusts its discrimination weights after looking at an adequately large set of sample data. While this is conceptually simple, in practice it requires quite powerful machines and very large data sets for training. These issues of scale were largely responsible for the long gap in implementation between Turing's original idea of learning machines and their practical realization.

ML systems operate on structured data. These systems assume that it's possible to get structured input, both for the training data and the actual operation. That's fine for something like classifying email as spam, but it wouldn't work for a self-driving car. Deep learning takes this model one very important step further.

In 1943, five years before Turing's intelligent machines report, Warren McCulloch and Walter Pitts published a paper[7] that described the potential to structure logical decision making in a manner based on a network of neurons, each of which takes inputs, has a threshold weight, and produces an output. This seminal work led, some 60 years later, to neural networks for classification and decision making. Neural networks apply feedback between layers of

nodes to "learn" at several different levels of generality. Importantly, unlike simpler ML systems, neural networks do not need structured data.

Figure 9.1 shows an example from the paper that introduced the model in 1958.[8] Each circle represents a "node"—that is, a set of interconnected neurons that process information before sending it on to nodes to its right or, in the case of rightmost nodes, back to nodes in the previous layer. Neural networks apply feedback between layers of nodes to "learn" at several different levels of generality. Each artificial neuron gets inputs, processes them based on an activation condition, and sends possibly modified output to the next layer or back to a previous layer.

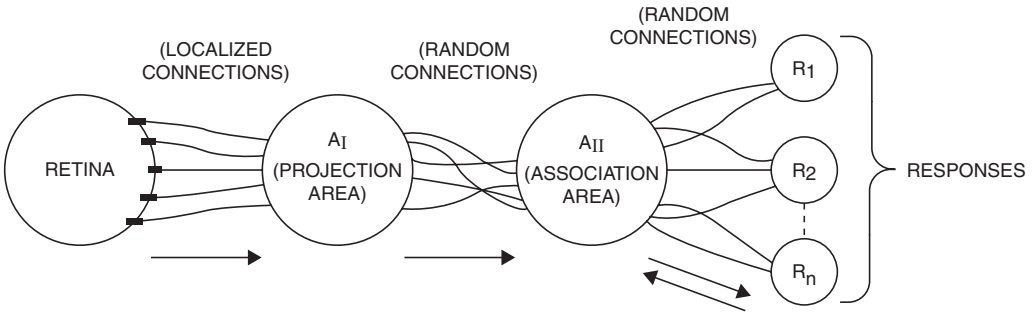


FIGURE 9.1 An artificial neural net modeled on the processing of signals from the retina to the brain.

Deep learning systems, built on artificial neural net platforms, are even more computationally intensive than traditional ML and frequently require special-purpose processors to achieve results in real time. Tesla claims that the AI processor in one of its cars (in 2019) is capable of 144 trillion operations per second—roughly equivalent to the power of 1,000 PCs.

Language translation is a good example of the impact of deep learning. Early attempts at machine translation—from Russian to English, for example—relied on attempts to create a logical model of grammar and vocabulary. As interesting as these linguistic models were, they ultimately failed as a means of doing language translation. Even early on, it was assumed that some form of machine learning would be better. Why not? After all, that's how we all learned our native language. You don't teach a two-year-old to conjugate verbs. You talk to her and correct her mistakes. And when, at six, the same child learns a second language, she doesn't start by learning sentence structure. She repeats the original process.

It took advances in algorithm development and computing power, and it also took a large resource of documents in multiple languages to make neural translation possible. The growth of the Internet made that a reality, and now

we have Google Translate, which currently supports more than 100 languages, from Afrikaans to Zulu.

ML isn't magical. It depends on powerful computers and lots of training data. It can seem like magic, though, because it can cut through tasks that are time-consuming or challenging for humans, And when a result comes out the other end, it often comes without any explanation—just a number, a rating, a score—yes, you get an interview, no you don't; your risk of recidivism is high or low.

Machine Learning and Training Data

Modeling a complex environment for a machine requires lots of labeled training data: lots of pictures of street signs and traffic signals with their meanings spelled out, lots of medical symptoms labeled to distinguish benign conditions from dangerous ones, lots of conversations transcribed and annotated. A great deal of human effort is therefore required to bootstrap the machines, and the data gathering raises numerous questions related to privacy for the data subjects included in training sets, working conditions for the people who do the labeling, and competitive implications of the reliance on big data.

Privacy

Jillian York, a technology and civil liberties activist, was on vacation when a friend contacted her to ask whether she knew her face appeared in an online database of "celebrity images."[9] She didn't, and upon investigation, was surprised to find included in the data set several years' worth of photographs captured by friends and still images extracted from videos. The images were included in the IARPA Janus Benchmark–C data set, a collection of images the U.S. Government's National Institute of Science and Technology (NIST) had made available as part of a public challenge to improve the state of the art in face recognition for "unconstrained in-the-wild face images." The challenge was "intended to drive research and development into face detection, verification, identification, and identity clustering."[10]

York was disturbed that her casual photos had been collected and indexed, but for NIST and IARPA, that was the point. "In-the-wild" meant training and testing recognition on faces with a variety of poses, backgrounds, and settings, so they gathered images that had been posted online, under Creative Commons licenses permitting their reproduction. That copyright permission didn't mean the photos' subjects expected to find themselves in a face ID database, however. Were it not for deep learning's voracious appetite for data, this privacy issue may never have arisen.

Labor

Training an AI system requires human labor to label its data sets: this image has a stop sign; that one shows a red light; that's a normal epithelial cell, this is a malignant one. Some companies distribute this new piece work through Amazon's Mechanical Turk. Others contract with workers who spend their days in call center–like cubicles, circling suspected polyps on colonoscopy videos. Without signing up for any of these jobs, you may have contributed to a training data set by answering the question "which of these photo squares contains a crosswalk?" in the captcha presented on sign-in to a new website. Sometimes this hidden labor comes to light, as when a number of automated voice "assistant" services updated their privacy policies to indicate that sometimes a human might listen to your conversation to help improve system performance.

And then sometimes, unbeknownst to us, systems monitor what we say and do and learn from our interactions. Is it right for corporations to use customers as unpaid labor?

Competition

The need for a great deal of training data means that data processing has economies of scale: Those who can gather more data can learn more from it, and as they improve their services, those services enable them to gather more data from the service's users. Every click of a Google search result has an effect—taking its searcher to a page Google believes will be useful—and a side effect—teaching Google that the page was responsive to the user's query. In the world of bits, scale really matters and provides the potential for insurmountable competitive advantage.

Nearly every Tesla car on the road sends back data that can be used to improve self-driving capabilities. Every time a Tesla driver intervenes when the car is steering to make a correction, data is captured and sent back for analysis. By mid-2020, Tesla had logged 3 billion miles driven on Autopilot.[11] Every mile driven, every automated lane change, every driver intervention provides data to Tesla that are used to improve the software. No other automaker comes close to having this sort of data source.

"Every time the customers drive the car, they're training the systems to be better. I'm just not sure how anyone competes with that."—Elon Musk

Algorithmic Decisions: I Thought Only People Could Do That

For the most part, computers take inputs, do some processing, and produce outputs. Increasingly, however, computers are taking over tasks that involve making decisions. Some are benign, but others can be life altering: which patients should enter hospice care,[12] who should receive a kidney transplant,[13] who gets to have a second-round interview for a job, which convict is likely to commit another crime.[14] Not all algorithmic decision systems rely on some form of machine learning, but many do. The very nature of these systems raises a host of concerns.

Just because a computer made a decision does not mean that it was right. Sometimes the results are funny—and sometimes tragic.

I'll Take Drosophila Biology for $23M

A postdoctoral researcher looking to buy a classic text on fruit fly biology found *The Making of a Fly* listed on Amazon for more than $1 million—$1,730,045.91 (+ $3.99 shipping), to be precise. He found this hard to imagine for a 1992 book (list price $70), when, 19 years later, a used copy could be ordered for as little as $35.54. Over several days, though, the price spiraled ever higher, until one seller was offering the book for $23,698,655.93.[15]

What was going on? Michael Eisen, a research biologist at University of California, Berkeley, found that only two sellers were offering new copies of the book, and as he tracked the book's pricing over a week, he discovered that the two had engaged in an algorithmic cat-and-mouse game. Both were using the same strategy. When one saw the other increasing its price, it increased its price, too. But the numbers were slightly different because the two were aiming at different markets. "bordeebook" aimed to top the market, while "profnath" aimed for a price just barely below the top price. As bordeebook's bot saw the market price shift, it raised its demand to 1.27059 times profnath's price; profnath's bot saw the market jump and, in turn, repriced to offer the book for sale at 99.83% of bordeebook's price, still a significant increase over its own previous price. bordeebook noticed profnath increasing its price and increased its own, and so on. With no human in the loop to check the algorithmic prices, the two bots ran amok, ratcheting up the price beyond what any person would pay for even a very good book (it averages 4.1 stars on Amazon) on Drosophila development.

Amazon isn't the only marketplace where bots can run wild. Consultants offer algorithmic pricing models for eBay auctions and Craigslist sales.

Ride-share services Uber and Lyft set "surge pricing" based on computerized observations that demand exceeds supply—and then they need human intervention or new limit conditions to stop price-gouging during natural disasters or emergencies.[16] Even the stock markets are filled with algorithmic traders. To keep up with other algorithms and gain a competitive advantage, traders program machines to make stock trades at high frequency without any human oversight. But automated trades inspire fears of a "flash crash"—fears that a rapid drop in the markets could be triggered by an input error or a connectivity glitch, setting off a cascade of automated sales and accidentally precipitating widespread economic chaos.

As we deal with buggy automation, we trade speed and efficiency against the risk of compound failures. Our risk tolerance for these algorithmic errors depends on the stakes. While we may find it acceptable to let rare book prices careen skyward until a person notices and pulls them down, we're less likely to tolerate that behavior for staple goods or stock markets, instead demanding regulation that slows the pace and provides oversight.

Algorithmic Justice League

Joy Buolamwini is a researcher at MIT's Media Lab, studying artificial intelligence—in particular, facial recognition. When she began her research, however, she found that her own face wasn't reliably detected as a face.[17] She had to wear a mask to be seen by the systems she was studying. Joy is black, and the programs she was testing had been trained on sets of white faces. The camera didn't see her because it hadn't been programmed with a wide enough range of human appearances.

Instead of just preparing a mask, Joy founded the Algorithmic Justice League to call attention to the problems of biased data sets. Her research has shown algorithms that mis-identify former First Lady Michelle Obama and women members of Congress at a much higher rate than their white male colleagues. Even if an algorithm can be designed without any racist intent or gender bias, if it is shown predominantly white male faces, those biases will color its facial recognition.

Unfortunately, these data biases not only persist but affect real-world deployments. The city of Detroit, which is majority black, has installed thousands of video cameras with facial recognition as part of "Project Green Light," a crime deterrence program.[18] However, one commercially available algorithm has a false positive rate of 1 in 1,000 for black faces, as compared with 1 in 10,000 for white faces. In a city with five times as many black faces as white, if a random set of the city population walks past the cameras, 50 black people will be wrongly identified for every 1 misidentified white person. Anyone who

appears in the state's driver's license database or other city records is liable to be caught up in the dragnet. Is being wrongly tagged for suspicion yet another hazard of "driving while black"?

In summer 2018, the ACLU used Amazon's Rekognition facial recognition application to run photos of sitting members of the U.S. Congress against a mugshot database. Rekognition nabbed 28 possible criminal matches.[19] (Whatever you think of their voting records, none of the congresspeople was pictured among the 25,000 public mugshots in the ACLU database.) Again, the errors were not uniformly distributed: "The false matches were disproportionately of people of color, including six members of the Congressional Black Caucus, among them civil rights legend Rep. John Lewis (D–Ga.)." Amazon responded that setting a higher threshold (rather than using the default) eliminated these errors.[20] That is to say, Amazon suggested that its software was fine but had been used incorrectly. If false positives (identifying a match that really wasn't a match) were a big problem, the ACLU should have set a higher confidence requirement for reporting a match (say, 99% confidence rather than 75% confidence). Of course, that explanation says nothing about the software's apparent propensity to misidentify black people.

Systematic errors like these prompted the city of San Francisco to enact a ban on facial recognition technologies—perhaps better described as face surveillance technologies, since no one objects to the use of face recognition for unlocking one's own cell phone. Civil liberties advocates brought their concerns to the city's board of supervisors, who voted 8–1 for the ban, persuaded that the dangers of discrimination and error outweighed the technology's potential utility.

Black-Box Justice

Algorithms in the justice system help determine whether an accused defendant will be released on bail or held in jail until trial, and they help set the length of a sentence someone will get after pleading guilty.

Investigative journalist Julia Angwin, writing in ProPublica, compared the use of AI in the cases of two people, each accused of $80 thefts.[21] Brisha Borden, age 18, was walking with a friend when they passed a child's bicycle and scooter and took the conveyances for brief rides before dropping them again. Vernon Prater, age 41, was caught after shoplifting $86.53 of tools from Home Depot. As each was booked into jail, a computer program gave a rating of their likelihood of re-offending. Borden, a black woman who had one juvenile misdemeanor charge, was deemed high risk, and Prater, a white man with an armed robbery conviction and another pending charge, was deemed low risk.

The computer program providing these re-offense risks asked a set of background questions about the defendants and delivered a number. But later events didn't match the computed predictions. Prater, the white man fingered by the computer as low risk, was re-arrested for a major electronics theft and is serving eight years in prison. Borden, identified as high risk, has avoided further legal trouble. However, the friend arrested along with her reports that the earlier arrest has made it difficult for her to find employment.

Programs such as Northpointe's COMPAS are used around the country as inputs to judges' sentencing decisions or bail hearings. The programs themselves are opaque: Fed a series of inputs, they produce a number but no explanation of their reasoning or opportunities to give feedback on their results. When defense attorneys ask for more detail to help their clients argue against mistakes, they are told that the programs are commercial trade secrets that can't be revealed.

Large-scale analyses show inconsistencies in the programs' application. In Angwin's analysis at Pro Publica, COMPAS was systematically biased against African Americans, marking them as more likely to re-offend than white defendants in otherwise similar circumstances. It will take further investigation to determine whether that bias was introduced deliberately in the programming, whether it was embedded in data used to train the system, or whether it emerged from data patterns that give an incomplete prediction for the future. The program itself is a black box: We see only its inputs and outputs, not what happens inside. But it becomes embedded in the justice system. It may save time, and people may truly believe its provider's claims of improved decision making, but it also can deflect responsibility for making hard decisions: "It wasn't my choice; the computer told me so."

Opacity is a fundamental characteristic of deep learning-based decision systems. These are systems that learned from experience, that formulated their own decision rules, and, most often, that have no mechanism for explaining how they arrived at their conclusions.

AI systems, by their very nature, are black boxes. The conclusions they draw, the classifications they make, the judgments they render result not from an algorithm defined and implemented by humans but rather from an accumulation of knowledge gained through observation. An opaque process cannot be seen to be fair both to the accused and to the observing public. Neither the accused nor the public can take guidance on how to stay on the right side of the law, as they can from reasoned judicial decision. When not even its creators know why an algorithm reached its conclusions or can explain what factors would change its outcome, and yet its product is given weight in a judicial proceeding, those seeking to challenge algorithmic results are denied due process.

What's Next

AI 2.0 solutions that utilize deep learning have the potential to be transformative beyond what we have already seen from the digital explosion. Their capacity to synthesize complex information and render judgments in real time presents the double-edged sword of opportunity and risk. A host of questions remain to be explored in the next years.

Responsibility

Elaine Herzberg was walking her bicycle across the road in Tempe, Arizona, on a dark Sunday night in 2018 when she was struck and killed by an autonomous Uber vehicle, becoming the first reported American death caused by a self-driving car.[22] In the days following, Uber engineers and law enforcement investigators pored over the data to try to determine what had gone wrong. The self-driving car was designed to avoid collisions but had failed to do so here, with fatal result. Was the problem in the software or hardware? In the sensors, computer vision, processing, or actuated response? And who was responsible—the software developers, the car company, the car owner?

Many of the car's components recorded data, logging inputs and outputs in a manner similar to an aircraft black-box flight recorder. Those logs could show whether sensors detected an object and whether a command to apply the brakes had issued. At some of the intermediate stages, however, questions of interpretation become more difficult: If the car "saw" an object but failed to identify the object as a person for whom it should stop (even if it had to brake suddenly), where might the recognition have erred?[23] Much of this software's operation is discontinuous: Two very similar scenes can present widely varied appearance to AI, and a small error or difference between test and real-world conditions can have dramatic consequences. Enumerating all the possibilities in testing, or even recording enough information for an after-the-fact audit, can be daunting.

In Uber's collision with Ms. Herzberg, a National Transportation Safety Board investigation found that an emergency braking system had been disabled, and the human "backup driver" didn't react fast enough to stop the car.

According to data obtained from the self-driving system, the system first registered radar and LIDAR observations of the pedestrian about 6 seconds before impact, when the vehicle was traveling at 43 mph. As the vehicle and pedestrian paths converged, the self-driving system software classified the pedestrian as an unknown object, as a vehicle, and then as a bicycle with varying expectations of future travel path. At 1.3 seconds before impact, the self-driving system determined that an emergency braking maneuver was

needed to mitigate a collision. According to Uber, emergency braking maneuvers are not enabled while the vehicle is under computer control, to reduce the potential for erratic vehicle behavior. The vehicle operator is relied on to intervene and act. The system is not designed to alert the operator.[24]

The car's computer vision systems saw a moving object, but, uncertain what kind of object, couldn't confidently predict whether it—she—would be in the car's path. By the time the system identified the pedestrian walking a bicycle and realized the car would hit her, its only corrective move was one that had been disabled. Designers of the system apparently made earlier choices that put the system in this bind. For example, although they could have instructed the car to slow at any hint of danger, that would slow the trip; unexpected braking could introduce a greater risk of being hit from behind.

Habituation

As we design artificially intelligent systems, we face choices of what parameters to set and what risks to take. Paradoxically, as systems get better, the choices may become more explicit. A human driver reacts instinctually (or fails to react) to an object darting into the road; an autonomous vehicle must be programmed to anticipate such obstacles and also to choose among a series of imperfect alternatives that might include braking hard (at some risk to its occupants and cars behind), swerving into oncoming traffic (at risk to occupants and the opposite lane), or veering into a crowded sidewalk. Calculating the risks and their potential harms seems both necessary and unfair: How can we compare human lives?

Trolley Problem

Philosopher Philippa Foot first described the "trolley problem" in a 1967 article.[25] She asks us to consider the choice facing "the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed." Suppose his trolley is running straight toward the five, with no room to apply the brakes, but he could throw a switch to divert to a spur track on which only one is working. Should he throw the switch? Must he?

This thought experiment, and a variety of similarly gruesome choices (should you push the fat man off the overpass to stop the train? kill a healthy person if her organs can help five sick people to live?), can help illuminate the factors that go into our thinking about "fairness" or, as our hair-splitting

get ever more tangled, remind us that intuitive judgments may be clouded by elements that should be irrelevant to a principled fairness.

The trolley problem has obvious appeal to those thinking about autonomous vehicles, where the vehicle might be in the pre-switch position, caught between alternatives in which any choice causes harm: Should the car protect its passenger or another car with two passengers? Its passenger or a pedestrian? Its owner's wallet against damage to the car? Some observers note that each of these decision situations comes only as the last step in a long chain of events and suggest that we should break out of the trolley problem framing altogether by changing the parameters of driving: Build interlock switches to keep the trolley off the track entirely when repairs are underway. Instead of putting autonomous vehicles on the road with human-driven cars, give them separate lanes and software that coordinates the actions of the fleet. Any of these lenses refocuses questions of systemic design and justice: Whose interests are considered and given priority, and who might be overlooked?

Transparency: Why Did You Do That?

And so we come back to Nicolette, who didn't get the job after the computer interviewed her. There are many similar examples. Christian Sandvig and Karrie Karahalios want to help us understand algorithmic systems by auditing them. They have tried, for example, preparing Facebook profiles with slight variations to compare the advertisements shown to each.[26] Will someone who describes herself as a single childless woman see the same housing ads as someone who gushes about her young children? As a man?

When presented with black-box AI, or even with a fully disclosed algorithm whose workings are opaque, we can try to understand its operation through audits. An after-the-fact audit examines the inputs and outputs, pulling out individual data points and aggregate results to look for anomalies or unexpected behavior.

Because their tests involve creating possibly false or adversarial information to interact with computer systems, however, Sandvig and Karahalios face challenges under the Computer Fraud and Abuse Act (CFAA) if their "fake profiles" go against the terms of service of one of the platforms they seek to test. The black box has rules against testing how it works. Landlords argued that auditors were trespassing when they came to test an apartment they weren't planning to rent (though they lost their suit). As Karahalios and Sandvig fight to enforce the right to conduct live-audit tests for possible machine discrimination, they are seeking to enshrine a similar protection for assessment of the digital environment. When does the right to examine a system's bias outweigh the system operator's right to limit access?

Is Better Good Enough?

Elaine Herzberg's death was a tragedy. Even though the Tempe police concluded that the accident was unavoidable,[27] Uber terminated its testing of self-driving cars as a result of this accident.[28] That is entirely understandable. We naturally respond to individual deaths in this way, and there is broad concern about the overall safety of autonomous vehicles even when accidents don't involve fatalities. How, though, should we think about statistical advantages? Consider the reported results for Tesla cars with Autopilot:

> In the 1st quarter, we registered one accident for every 4.68 million miles driven in which drivers had Autopilot engaged. For those driving without Autopilot but with our active safety features, we registered one accident for every 1.99 million miles driven. For those driving without Autopilot and without our active safety features, we registered one accident for every 1.42 million miles driven. By comparison, NHTSA's most recent data shows that in the United States there is an automobile crash every 479,000 miles.[29]

The makers of the HireVue automated screening system similarly argue that their system works better than the humans it replaces. Granting that the system offers no explanation for its decision, the company claims that:

> Decades of research have shown that traditional interviews are full of implicit and explicit bias, and tremendous inconsistency. The HireVue approach has been proven to be measurably more accurate at predicting performance than human evaluators and is audited, tested, retrained, and audited again to ensure that there is no adverse impact.[30]

So, which is it—biased or not? Are we fundamentally better off with a system whose bias can be evaluated, even if it cannot be interrogated?

The Future of Work

We have already seen the tremendous impact on productivity that has resulted from the digital explosion. We no longer have typing pools, and travel agents are all but extinct. The accounts payable department is a tiny fraction of the size it once was, as technology has taken over a wide range of tasks.

Until the advent of AI 2.0—of machine learning systems capable of perceiving, understanding, and interacting with the physical world—the majority of the impact on labor has been limited to information-intensive tasks. That is about to change, as we see computer systems that are capable of learning,

capable of self-improvement, and capable of making seemingly complex decisions in real time.

One of the most common jobs in the United States is "driver"—trucks, buses, taxis, tractors, fork lifts, Uber, and Lyft. How long before autonomous vehicles of one sort or another take over these jobs? And what will happen in all sorts of other professions—in tax preparation, reading X-rays, customer service, and more?

The Role of Regulation

AI, in general, and deep learning systems that render opaque judgments in particular, present a critical need for intelligent regulation.

Asimov's laws of robotics have given birth to 1,000 alternatives and descendants. Even standards just for algorithmic fairness and transparency are numerous. The U.S. Association for Computing Machinery proposed the following principles as a starting point:[31]

1. Awareness: Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.

2. Access and redress: Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

3. Accountability: Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.

4. Explanation: Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.

5. Data Provenance: A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over

privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.

6. Auditability: Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.

7. Validation and Testing: Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

But how do we get policymakers, designers, engineers, and even consumers to take such principles seriously when motives for profit, market share, innovation, and simple excitement are pressing on everyone to get the product built and out the door without, in the words of Norbert Wiener, the father of Cybernetics, "exert[ing] the full strength of our imagination to examine where the full use of our new modalities may lead us"[32]?

Optimism for the Future

Despite all the inherent risks and complexities, we should recognize that AI and ML have vast potential for good. Just a few examples of the solutions in development now are accelerated pharmaceutical development, increased crop yields, reduced automotive injuries and deaths, lower-cost health care, better fraud and crime detection, and improved manufacturing efficiencies.

AI in general and machine learning in particular create the potential for exponential acceleration in the capabilities of digital systems as they learn to improve themselves, much as we humans have done for millennia.

Bits Lighting Up the World

So what happens after the digital explosion? To be sure, we are today nowhere near "after." More bits than ever before are now being produced, analyzed, stored, and used as training data for systems that consume and generate more bits. We are still near the beginning of the explosion. But it is not too soon to try to view it as a whole.

In Greek mythology, Prometheus stole Zeus's fire and brought it from Olympus to Earth, along with the useful arts of civilization. Zeus retaliated for

Prometheus's trickery by visiting upon humanity the ills and evils that beset us. We have been trying to make the best of things ever since.

The Prometheus myth is about technology. Technology, like fire, is neither good nor bad; its value depends on how we use it. And once we start using a technology, society itself changes. As William Yeats wrote, "All changed, changed utterly: A terrible beauty is born."[33]

Information technologies spark a special kind of fire. Bits are the atomic particles of the information flames. With our information tools, we can do things, both good and bad, that we could not have done unassisted. For better or worse, these technologies enable us to think, reason, create, express, debate, compromise, learn, and teach in ways never before possible. They connect people across physical space, both in pairs and in groups. They extend the reach of our voices and the range of our hearing. They also amplify our capacity to frighten, harass, and hate other people and to misrepresent ourselves to others. They enable us to earn and to spend money without going anywhere and also to steal money from the comfort of our homes.

So central was Prometheus, the fire-bringer, to the Greek conception of humanity that in later retellings of the myth, he is credited with creating the human species itself. What changes to society will information technologies yield, in a decade or two, when the ongoing digital explosion has unimaginable power?

We don't know, of course. But if things go on changing as they are changing today, there are likely to be dramatic changes to three distinctive aspects of human culture: our sense of personal identity and privacy, our capacity for free speech, and the creativity that drives human progress.

Privacy and Personhood

As the digital explosion was beginning, the struggle over privacy seemed to be a war. Individuals wanted to protect themselves from invasive forces. Institutions, both corporations and government, wanted the benefit of information that individuals would rather not reveal. In actual practice, as us versus them, good versus bad, or individuals versus institutions.

In the digital explosion, as technologies improved, data gathering became easier and less annoying. Modest incentives induced individuals to sacrifice their personal privacy, often before they understood what they were giving up. Relatively few people today worry about stores keeping track of their purchases. Even without loyalty cards, a credit card swipe together with bar code scans at the cash register link a customer's name to his preferences in candy and condoms. You have to give up many conveniences to protect your privacy, and most people are not willing to do it.

The next generation may not even see the loss of privacy as a sacrifice. Socrates said that the unexamined life is not worth living, but people who have grown up with social networks may find life fully exposed to public view simply normal. As Sun Microsystem CEO Scott McNealy quipped, "You have zero privacy anyway. Get over it."

But getting over it is not so simple when social interactions happen through the computer screen. When most personal interactions were face-to-face or over the telephone, we mistrusted people claiming to represent our bank and trusted people we felt we had gotten to know. In the electronic world, we do the opposite: We trust our bank's website with large sums of our money, but we have to be reminded that close electronic friends may be impostors. Where is the border for children between the personal and the public? Will we need laws about fraudulent friendships?

As electronic privacy becomes lost in the cloud of bits and as caution gives way to social networking, what societal structures will break down? What will evolve to replace them? Society as we know it functions because of a web of trusting relationships between parties who are independently responsible for their own actions. What will replace that if the concept of personal identity becomes meaningless? Will the very notions of privacy and identity be destroyed in the explosion?

What Can We Say, and Who Will Be Listening?

The digital explosion revolutionizes human communication. Earlier technologies for disseminating text, spoken words, and images also changed the world—but all included choke points. A million eyes might have read your book, but only if you could get it published. You might have discovered a scandal that would bring down a government, but only if you could get a newspaper to expose it to public view. A million ears might have heard your speeches, but only if you could control a radio station.

No longer are speakers bound by the whims of those who control the loudspeakers and printing presses. In the United States, anyone can say anything, without permission from church or state, and be audible to millions. No one has to listen, but it is easy to put the message where millions can hear it.

An Earlier Information Revolution

Victor Hugo said of printing in *The Hunchback of Notre Dame:* "It is the mother of revolution. It is the mode of expression of humanity which is totally renewed; it is human thought stripping off one form and donning another; it is the complete and definitive change of skin of that symbolical serpent which since the days of Adam has represented intelligence."

And yet there is a cost. It is not a financial cost; after all, it costs nothing to spread the word via email or Twitter or Facebook or YouTube. The cost is that the speaker relies on many intermediaries to handle the messages, and so there are many opportunities for snooping, eavesdropping, filtering, and censoring, not to mention the dilemma of knowing the source and reliability of the information you receive. The choke points have multiplied and become more diffuse, but they have not disappeared.

The very technological miracles that have created the communication revolution have also created a Big Brother revolution. Massive surveillance in China has been automated based on face recognition coupled with monitoring of cell phone signals.[34] U.S. law enforcement has embraced Clearview AI's face surveillance capabilities with equal enthusiasm.[35] With the success of speech recognition and language understanding, we have to expect that it is already feasible to monitor every voice communication passing over telephone wires or the Internet by means of the technological equivalent of a human listener. Machines will be waiting attentively for someone to say the "wrong" thing—whatever that is deemed to be.

Governments eavesdrop to protect national security, political opposition, and public morality. Communication companies want to listen to what their networks are being used for so that they can tailor their service to the content in the most profitable way—a soft form of corporate censorship, in which unwanted communications are slowed down or made costly. Service providers want to listen in so they can add advertising to the content they deliver.

In spite of the unimaginable expansion of communications over the past quarter century, the jury is still out on whether speech will be freer or less free in the future than it was in the past, even in the United States, with its uncompromising First Amendment. And like the tree falling in the forest, of what use will free speech be if no one is listening? The dramatic pluralism of our information sources threatens to create a society where no one learns anything from people with whom they disagree. It is simply too easy for people to decide whom they want to hear and to ignore everyone else. Will the digital explosion in fact make information more limited?

In spite of the unimaginable expansion of communications over the past quarter century, the jury is still out on whether speech will be freer or less free in the future than it was in the past.

A Creative Explosion or a Legal Explosion?

In the same letter quoted in Chapter 1, Thomas Jefferson wrote, "He who receives an idea from me, receives instruction himself without lessening mine;

as he who lights his taper at mine, receives light without darkening me." Will the digital explosion be used to enlighten the world or to create illusions and to blind us to the truth?

What would Jefferson have to say about the viral spread of misinformation through online social networks? Four years after the 2016 U.S. elections, social media companies and governments everywhere have found no simple means to balance the right to express political views with the reality that interesting lies spread quickly and unexciting truths can take a long time and immense labor to reveal. And not all misinformation is political or commercial. When disease spreads rapidly, it can be hard for reliable information to penetrate where unfounded conspiracy theories have taken root. During the novel coronavirus epidemic that began in 2019, anti-vaccination misinformation spread before the first vaccine trials had even begun. Only a year earlier, 83 people, mostly young children, died on the tiny island of Samoa from the entirely preventable disease measles because an anti-vaccination campaign had reduced vaccination rates on the island to 34%.[36]

Manipulation of speech is not the only form of technologically enabled control of information. Patent and copyright laws in the United States were designed to promote individual creativity in the interest of the progress of society. The law struck a balance between providing financial incentive to the creator and high social benefit to the population at large. The term for which artists and inventors maintained exclusive control over their creations was designed to be long enough to provide a financial return and short enough to provide an incentive for continued creativity. And there was a high threshold on what could be protected at all so that the system did not encourage lawyerly inventiveness rather than artistic and engineering creativity.

As mechanical tools have been supplanted by information-processing tools, and all manner of writing, music, and art have gone digital, the rules of the game have changed. The parties that receive the strongest protections are now major corporations rather than the original creators or the ultimate consumers. At a time when information technology promises disintermediation—getting rid of intermediaries—those intermediaries are becoming more powerful, not less.

Unexpected—and unintended—consequences attend any speech-limiting regulation. In the midst of the Covid-19 crisis, copyright takedown notices were issued to Google to remove informative articles from search results. Google purged a news story about two visitors to Vietnam who had become sick—information that could have helped others learn that they had been exposed to the virus. The story appeared on a Vietnamese government–affiliated news website, naming the hotel, bars, and restaurants the tourists visited and urging readers who patronized the establishments to take precautions. Someone who wanted this information removed backdated a blog

post with the identical information and then complained that the news story infringed his copyright. Google removed search links to the original story even though the fake was dated more than four months before the tourists visited. The blog with the fake post included only seven others, all cited in copyright complaints filed with Google.[37]

The legal power of strong intermediaries protecting their economic interests has increased at the same time as new technologies have empowered the creators to reach their consumers directly.

Similar tensions are visible in the world of invention. The power of the incumbent radio and television broadcast industries to exclude newcomers from the airwaves restrains both speech and invention, limiting radio communications and keeping useful devices off the market. And there is good reason to ask if the same pattern is repeating itself in the domains of information search—where Google is dominant—and social networking, with Facebook's overwhelming position in large parts of the world.

Will the United States move toward being an information democracy or an information oligarchy? Whose hands will be on the controls that regulate the way we produce and use bits in the future?

A Few Bits in Conclusion

The worldwide bits explosion is lighting up the world (see Figure 9.2).[38] Most of the illumination today is in Europe and North America, but it is growing brighter almost everywhere. There is no physical reason it can't continue to grow. Bits are not like oil or coal. Their production requires almost no raw materials and only tiny amounts of electricity. They flow through glass fibers in astonishing numbers, and they radiate through space, over short distances and long. With our cameras and computers, we produce them at will, in unintelligibly larger numbers every year. Existing dark spots—North Korea, for example—may remain black for a time, but eventually even these regions may glow brighter. And all that data and thought-stuff, all those atoms of light, can be captured and stored electronically for eternity.

The explosion happened through technological inventions supported by political and economic freedoms. Gutenberg laid the foundation when he invented the printing press, and Morse's telegraph, Bell's telephone, and Edison's phonograph were all precursors. Claude Shannon was the bits Prometheus. After the Second World War, his mathematical insights lit the flame of communication and computing technologies, which have now illuminated the earth with bits.

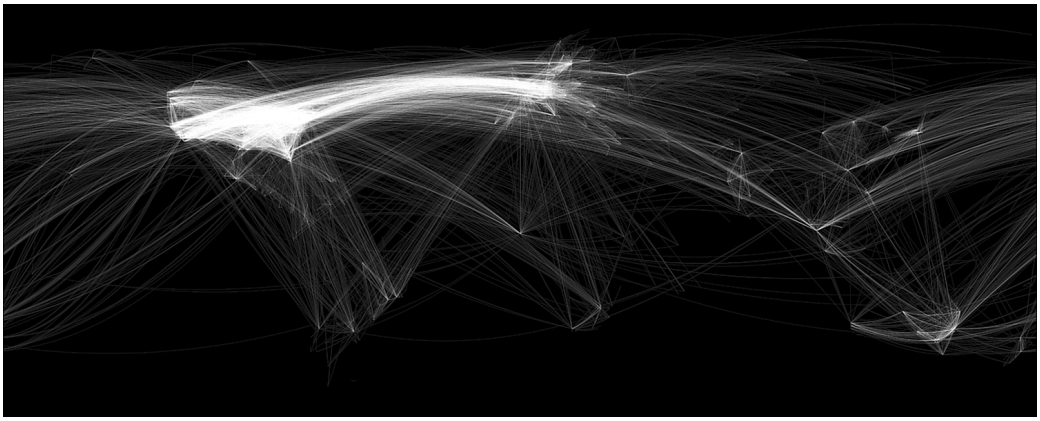


FIGURE 9.2 A map of the world, showing the number of Internet connections between routers. At present, the United States and Europe are heavily interconnected. If the volume of data transmissions were depicted instead (giving more prominence, for example, to areas with heavily used Internet cafés), Africa, Asia, and South America might show more prominently.

The bits explosion is not over. We are in the middle of it. But we don't know whether it will be destructive or enlightening. The time for deciding who will control the explosion may soon be past. Bits are still a new phenomenon—a new natural resource whose regulatory structures and corporate ownership are still up for grabs. The legal and economic decisions being made today—not just about bits but about everything that depends on bits—will determine how our descendants will lead their lives. The way the bits illuminate or distort the world will shape the future of humanity.

Endnotes

1 Isaac Asimov, *I, Robot* (Gnome Press, 1950).

2 Russell Flannery, "2017 Forbes China 100 Top Businesswomen List," *Forbes*, February 6, 2017.

3 Xinmei Shen, "Facial Recognition Camera Catches Top Businesswoman 'Jaywalking' Because Her Face Was on a Bus," *Abacus*, November 22, 2018, https://www.abacusnews.com/digital-life/facial-recognition-camera-catches-top-businesswoman-jaywalking-because-her-face-was-bus/article/2174508.

4 Homer, *The Iliad of Homer* (University of Chicago Press, 2011).

5 A. M. Turing, "Computing Machinery and Intelligence," *Mind* 49, no. 236 (October 1950): 433–460.

6 Weizenbaum, Joseph. "ELIZA-a Computer Program for the Study of Natural Language Communication between Man and Machine." *Communications of the ACM* 9, no. 1 (1966): 36–45. https://doi.org/10.1145/365153.365168.

7 Warren McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5, no. 4 (1943): 115–133, https://link.springer.com/article/10.1007/BF02478259

8 Frank Rosenblatt, "The Perceptron," *Psychological Review* 65, no. 6 (1958).

9 Max Harlow and Madhumita Murgia, "Who's Using Your Face? The Ugly Truth About Facial Recognition," April 19, 2019, https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e.

10 "Face Challenges," National Institute of Science and Technology, June 10, 2015, https://www.nist.gov/programs-projects/face-challenges.

11 Karpathy, Andrej, Presentation at Scaled Machine Learning Conference, April 2020, https://youtu.be/hx7BXih7zx8

12 Kris Newby, "Compassionate Intelligence: Can Machine Learning Bring More Humanity to Health Care?" Stanford Medicine, Summer 2018, https://stanmed.stanford.edu/2018summer/artificial-intelligence-puts-humanity-health-care.html.

13 Loren Larsen, "HireVue Assessments and Preventing Algorithm Bias," HireVue, June 22, 2018, https://www.hirevue.com/blog/hirevue-assessments-and-preventing-algorithmic-bias.

14 Noel L. Hillman, "The Use of Artificial Intelligence in Gauging Risk of Recidivism," *The Judges' Journal*, January 1, 2019, https://www.americanbar.org/groups/judicial/publications/judges_journal/2019/winter/the-use-artificial-intelligence-gauging-risk-recidivism/.

15 Michael Eisen, "Amazon's $23,698,655.93 Book About Flies," It Is Not Junk, April 22, 2011, http://www.michaeleisen.org/blog/?p=358; John D. Sutter, "Amazon Seller Lists Book at $23,698,655.93 -- plus Shipping," CNN, April 25, 2011, http://www.cnn.com/2011/TECH/web/04/25/amazon.price.algorithm/index.html.

16 Victor Lyckerson, "Uber Agrees to Limit Surge Pricing During Emergencies, Disasters," *Time*, July 8, 2014, https://time.com/2967490/uber-agrees-to-limit-surge-pricing-during-emergencies-disasters/; Mike Isaac, "Uber Reaches Deal with New York on Surge Pricing in Emergencies," *Bits Blog*, July 8, 2014, https://bits.blogs.nytimes.com/2014/07/08/uber-reaches-agreement-with-n-y-on-surge-pricing-during-emergencies/.

17 Steve Lohr, "Facial Recognition Is Accurate, If You're a White Guy," *The New York Times*, February 9, 2018, https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.

18 Amy Harmon, "As Cameras Track Detroit's Residents, a Debate Ensues Over Racial Bias," *The New York Times*, July 8, 2019, https://www.nytimes.com/2019/07/08/us/detroit-facial-recognition-cameras.html.

19 Jacob Snow, "Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots," American Civil Liberties Union, July 26, 2018, https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28.

20 Ry Crist, "Amazon's Rekognition Software Lets Cops Track Faces: Here's What You Need to Know," CNET, March 19, 2019, https://www.cnet.com/news/what-is-amazon-rekognition-facial-recognition-software/.

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

22 Daisuke Wakabayashi, "Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam," *The New York Times*, March 19, 2018, https://www.nytimes.com/2018/03/19/technology/uber-driver-lessfatality.html.

23 Aarian Marshall and Alex Davies, "Uber's Self-Driving Car Didn't Know Pedestrians Could Jaywalk," *Wired*, November 5, 2019, https://www.wired.com/story/ubers-self-driving-car-didnt-know-pedestrians-could-jaywalk/.

24 "Preliminary Report, Highway HWY18MH010," National Traffic Safety Board, accessed August 25, 2020, https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf.

25 Philippa Foot, "The Problem of Abortion and the Doctrine of the Double Effect," *Oxford Review*, no. 5 (1967): 5–15.

26 Christian Sandvig et al., "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms," paper presented to "Data and Discrimination: Converting Critical Concerns into Productive Inquiry," a preconference at the 64th Annual Meeting of the International Communication Association, May 22, 2014, Seattle, WA, https://www.semanticscholar.org/paper/Auditing-Algorithms-%3A-Research-Methods-for-on-Sandvig-Hamilton/b7227cbd34766655dea10d0437ab10df3a127396?p2df.

27 Uriel J. Garcia and Katrina Bland, "Tempe Police Chief: Fatal Uber Crash Likely 'Unavoidable' for Any Kind Of Driver," *The Arizona Republic*, March 20, 2018.

28 Carolyn Said, "Uber Puts the Brakes on Testing Robot Cars in California After Arizona Fatality," *San Francisco Chronicle*, March 27, 2018. https://www.sfchronicle.com/business/article/Uber-pulls-out-of-all-self-driving-car-testing-in-12785490.php#:~:text=Uber%20plans%20to%20end%20all,Motor%20Vehicles%20to%20the%20company.&text=Uber's%20permit%20will%20expire%20Saturday%2C%20the%20letter%20said.

29 Tesla Motors, "Tesla Vehicle Safety Report," https://www.tesla.com/VehicleSafetyReport.

30 Loren Larsen, "HireVue Assessments and Preventing Algorithm Bias,"

31 "Statement on Algorithmic Transparency and Accountability," Association for Computing Machinery, U.S. Public Policy Council, January 12, 2017, https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf.

32 *Science* 06 May 1960:Vol. 131, Issue 3410, pp. 1355-1358. https://science.sciencemag.org/content/131/3410/1355.

33 William Butler Yeats, "Easter 1916," https://www.poetryfoundation.org/poems/43289/easter-1916.

34 Paul Mozur and Aaron Krolik, "A Surveillance Net Blankets China's Cities, Giving Police Vast Powers," - *The New York Times*," accessed May 14, 2020, https://www.nytimes.com/2019/12/17/technology/china-surveillance.html.

35 Kashmir Hill, "The Secretive Company That Might End Privacy as We Know It," *The New York Times*, December 17, 2019, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facialrecognition.html.

36 Renee DiResta, "Health Experts Don't Understand How Information Moves," *The Atlantic*, May 6, 2020, https://www.theatlantic.com/ideas/archive/2020/05/health-experts-dont-understand-how-information-moves/611218/.

37 Andrea Fuller et al., "Google Hides News, Tricked by Fake Claims," *Wall Street Journal*, May 15, 2020, https://www.wsj.com/articles/google-dmca-copyright-claims-takedown-online-reputation-11589557001.

38 Chris Harrison, Human–Computer Interaction Institute, Carnegie Mellon University, www.chrisharrison.net/projects/InternetMap/high/worldBlack.png.