

Prologue

What Is Progress?

Every day, we hear from executives, journalists, politicians, and even some of our colleagues at MIT that we are heading relentlessly toward a better world, thanks to unprecedented advances in technology. Here is your new phone. There goes the latest electric car. Welcome to the next generation of social media. And soon, perhaps, scientific advances could solve cancer, global warming, and even poverty.

Of course, problems remain, including inequality, pollution, and extremism around the globe. But these are the birth pains of a better world. In any case, we are told, the forces of technology are inexorable. We couldn't stop them if we wanted to, and it would be highly inadvisable to try. It is better to change ourselves—for example, by investing in skills that will be valued in the future. If there are continuing problems, talented entrepreneurs and scientists will invent solutions—more-capable robots, human-level artificial intelligence, and whatever other breakthroughs are required.

People understand that not everything promised by Bill Gates, Elon Musk, or even Steve Jobs will likely come to pass. But, as a world, we have become infused by their techno-optimism. Everyone everywhere should innovate as much as they can, figure out what works, and iron out the rough edges later.

WE HAVE BEEN here before, many times. One vivid example began in 1791, when Jeremy Bentham proposed the *panopticon*, a prison design. In a circular building and with the right lighting, Bentham argued, centrally positioned guards could create the impression of watching everyone all the

time, without themselves being observed—supposedly a very efficient (low-cost) way of ensuring good behavior.

The idea at first found some traction with the British government, but sufficient funding was not forthcoming, and the original version was never built. Nevertheless, the panopticon captured the modern imagination. For the French philosopher Michel Foucault, it is a symbol of oppressive surveillance at the heart of industrial societies. In George Orwell's *1984*, it operates as the omnipresent means of social control. In the Marvel movie *Guardians of the Galaxy*, it proves to be a flawed design that facilitates an ingenious prison breakout.

Before the panopticon was proposed as a prison, it was a factory. The idea originated with Samuel Bentham, Jeremy's brother and an expert naval engineer then working for Prince Grigory Potemkin in Russia. Samuel's idea was to enable a few supervisors to watch over as many workers as possible. Jeremy's contribution was to extend that principle to many kinds of organizations. As he explained to a friend, "You will be surprised when you come to see the efficacy which this simple and seemingly obvious contrivance promises to be to the business of schools, manufactories, Prisons, and even Hospitals...."

The panopticon's appeal is easy to understand—if you are in charge—and was not missed by contemporaries. Better surveillance would lead to more compliant behavior, and it was easy to imagine how this could be in the broader interest of society. Jeremy Bentham was a philanthropist, animated by schemes to improve social efficiency and help everyone to greater happiness, at least as he saw it. Bentham is credited today as the founder of the philosophy of utilitarianism, which means maximizing the combined welfare of all people in society. If some people could be squeezed a little in return for a few people gaining a great deal, that was an improvement worth considering.

The panopticon was not just about efficiency or the common good, however. Surveillance in factories implied inducing workers to labor harder, and without the need to pay them higher wages to motivate greater effort.

The factory system spread rapidly in the second half of the eighteenth century across Britain. Even though they did not rush to install panopticons, many employers organized work in line with Bentham's general approach. Textile manufacturers took over activities previously performed by skilled

weavers and divided them up more finely, with key elements now done by new machines. Factory owners employed unskilled workers, including women and small children, to perform simple repetitive tasks, such as pulling a handle, for as many as fourteen hours per day. They also supervised this labor force closely, lest anyone slow down production. And they paid low wages.

Workers complained about conditions and the backbreaking effort. Most egregious to many were the rules they had to follow in factories. One weaver put it this way in 1834: "No man would like to work in a power-loom, they do not like it, there is such a clattering and noise it would almost make some men mad; and next, he would have to be subject to a discipline that a hand-loom weaver can never submit to."

New machinery turned workers into mere cogs. As another weaver testified before a parliamentary committee in April 1835, "I am determined for my part, that if they will invent machines to supersede manual labour, they must find iron boys to mind them."

To Jeremy Bentham, it was self-evident that technology improvements enabled better-functioning schools, factories, prisons, and hospitals, and this was beneficial for everyone. With his flowery language, formal dress, and funny hat, Bentham would cut an odd figure in modern Silicon Valley, but his thinking is remarkably fashionable. New technologies, according to this view of the world, expand human capabilities and, when applied throughout the economy, greatly increase efficiency and productivity. Then, the logic goes, society will sooner or later find a way of sharing these gains, generating benefits for pretty much everybody.

Adam Smith, the eighteenth-century founding father of modern economics, could also join the board of a venture capital fund or write for *Forbes*. In his view, better machines would lead to higher wages, almost automatically:

In consequence of better machinery, of greater dexterity, and of a more proper division and distribution of work, all of which are the natural effects of improvement, a much smaller quantity of labour becomes requisite for executing any particular piece of work, and though, in consequence of the flourishing circumstances of the society, the real price of labour should rise very considerably....

In any case, resistance is futile. Edmund Burke, contemporary of Bentham and Smith, referred to the laws of commerce as “the laws of nature, and consequently the laws of God.”

How can you resist the laws of God? How can you resist the unstoppable march of technology? And anyway, why resist these advances?

ALL OF THIS optimism notwithstanding, the last thousand years of history are filled with instances of new inventions that brought nothing like shared prosperity:

- A whole series of technological improvements in medieval and early modern agriculture, including better plows, smarter crop rotation, more use of horses, and much improved mills, created almost no benefits for peasants, who constituted close to 90 percent of the population.
- Advances in European ship design from the late Middle Ages enabled transoceanic trade and created massive fortunes for some Europeans. But the same kinds of ships also transported millions of enslaved people from Africa to the New World and made it possible to build systems of oppression that lasted for generations and created awful legacies persisting today.
- Textile factories of the early British industrial revolution generated great wealth for a few but did not raise worker incomes for almost a hundred years. On the contrary, as the textile workers themselves keenly understood, work hours lengthened and conditions were horrible, both in the factory and in crowded cities.
- The cotton gin was a revolutionary innovation, greatly raising the productivity of cotton cultivation and turning the United States into the largest cotton exporter in the world. The same invention intensified the savagery of slavery as cotton plantations expanded across the American South.
- At the end of the nineteenth century, German chemist Fritz Haber developed artificial fertilizers that boosted agricultural yields. Subsequently, Haber and other scientists used the same ideas to design

chemical weapons that killed and maimed hundreds of thousands on World War I battlefields.

- As we discuss in the second half of this book, spectacular advances in computers have enriched a small group of entrepreneurs and business tycoons over the last several decades, whereas most Americans without a college education have been left behind, and many have even seen their real incomes decline.

Some readers may object at this point: Did we not in the end hugely benefit from industrialization? Aren't we more prosperous than earlier generations, who toiled for a pittance and often died hungry, thanks to improvements in how we produce goods and services?

Yes, we are greatly better off than our ancestors. Even the poor in Western societies enjoy much higher living standards today than three centuries ago, and we live much healthier, longer lives, with comforts that those alive a few hundred years ago could not have even imagined. And, of course, scientific and technological progress is a vital part of that story and will have to be the bedrock of any future process of shared gains. But the broad-based prosperity of the past was not the result of any automatic, guaranteed gains of technological progress. Rather, shared prosperity emerged because, and only when, the direction of technological advances and society's approach to dividing the gains were pushed away from arrangements that primarily served a narrow elite. We are beneficiaries of progress, mainly because our predecessors made that progress work for more people. As the eighteenth-century writer and radical John Thelwall recognized, when workers congregated in factories and cities, it became easier for them to rally around common interests and make demands for more equitable participation in the gains from economic growth:

The fact is, that monopoly, and the hideous accumulation of capital in a few hands, like all diseases not absolutely mortal, carry, in their own enormity, the seeds of cure. Man is, by his very nature, social and communicative—proud to display the little knowledge he possesses, and eager, as opportunity presents, to encrease his store. Whatever presses men together, therefore, though it may generate some vices, is favourable to the diffusion of knowledge, and ultimately promotive of

human liberty. Hence every large workshop and manufactory is a sort of political society, which no act of parliament can silence, and no magistrate disperse.

Electoral competition, the rise of trade unions, and legislation to protect workers' rights changed how production was organized and wages were set in nineteenth-century Britain. Combined with the arrival of a new wave of innovation from the United States, they also forged a new direction of technology—focused on increasing worker productivity rather than just substituting machinery for the tasks they used to perform or inventing new ways of monitoring them. Over the next century, this technology spread throughout Western Europe and then the world.

Most people around the globe today are better off than our ancestors because citizens and workers in early industrial societies organized, challenged elite-dominated choices about technology and work conditions, and forced ways of sharing the gains from technical improvements more equitably.

Today we need to do the same again.

The good news is that incredible tools are available to us, including magnetic resonance imaging (MRI), mRNA vaccines, industrial robots, the internet, tremendous computational power, and massive amounts of data on things we could not measure before. We can use these innovations to solve real problems—but only if these awesome capabilities are focused on helping people. This is not the direction in which we are currently heading, however.

Despite what history teaches us, the predominant narrative today has shifted back toward something remarkably close to what was prevalent in Britain 250 years ago. We are living in an age that is even more blindly optimistic and more elitist about technology than the times of Jeremy Bentham, Adam Smith, and Edmund Burke. As we document in Chapter 1, people making the big decisions are once again deaf to the suffering created in the name of progress.

We wrote this book to show that progress is never automatic. Today's "progress" is again enriching a small group of entrepreneurs and investors, whereas most people are disempowered and benefit little.

A new, more inclusive vision of technology can emerge only if the basis of social power changes. This requires, as in the nineteenth century, the rise of counterarguments and organizations that can stand up to the conventional wisdom. Confronting the prevailing vision and wresting the direction of technology away from the control of a narrow elite may even be more difficult today than it was in nineteenth-century Britain and America. But it is no less essential.

Artificial Struggle

Nothing has been written on this topic which can be considered as decisive—and accordingly we find everywhere men of mechanical genius, of great general acuteness and discriminative understanding, who make no scruple in pronouncing the Automaton a *pure machine*, unconnected with human agency in its movements, and consequently, beyond all comparison, the most astonishing of the inventions of mankind.

—EDGAR ALLAN POE, “Maelzel’s Chess Player,” 1836 (italics in original)

The world of the future will be an ever more demanding struggle against limitations of our intelligence, not a comfortable hammock in which we can lie down to be waited upon by our robot slaves.

—NORBERT WIENER, *God and Golem, Inc.*, 1964

In its special report on the future of work in April 2021, the *Economist* magazine took to task those worrying about inequality and dwindling job opportunities for workers: “Since the dawn of capitalism people have lamented the world of work, always believing that the past was better than the present and that the workers of the day were uniquely badly treated.”

Fears about AI-driven automation are particularly overblown, and “popular perceptions about the world of work are largely misleading.” The report proceeded to provide a clear restatement of the productivity bandwagon: “In fact, by lowering costs of production, automation can create more demand for goods and services, boosting jobs that are hard to automate. The economy may need fewer checkout attendants at supermarkets, but more massage therapists.”

The report’s overall assessment: “A bright future for the world of work.”

The management consulting company McKinsey expressed a similar conclusion in early 2022 as part of its strategic partnership with the annual World Economic Forum in Davos:

For many members of the world’s workforces, change can sometimes be seen as a threat, particularly when it comes to technology. This is often coupled with fears that automation will replace people. But a look beyond the headlines shows that the reverse is proving to be true, with Fourth Industrial Revolution (4IR) technologies driving productivity and growth across manufacturing and production at brownfield and greenfield sites. These technologies are creating more and different jobs that are transforming manufacturing and helping to build fulfilling, rewarding, and sustainable careers.

The *Economist* and McKinsey were articulating views of many tech entrepreneurs and experts that concerns about AI and automation are exaggerated. The Pew Research Center surveyed academics and technology leaders, and reported statements from more than a hundred of them, with the overwhelming majority stating that although there were downsides, AI would bring widespread economic and societal benefits.

According to the prevailing perspective, there may be some disruption along the way—for example, in terms of jobs lost—but such transition costs are unavoidable. In the words of one of the experts quoted by the Pew Research Center, “In the coming 12 years AI will enable all sorts of professions to do their work more efficiently, especially those involving ‘saving life’: individualized medicine, policing, even warfare (where attacks will focus on disabling infrastructure and less in killing enemy combatants and civilians).” The same person also conceded, “Of course, there will be

some downsides: greater unemployment in certain ‘rote’ jobs (e.g., transportation drivers, food service, robots and automation, etc.).”

But we should not worry too much about these downsides, for we have the same tech entrepreneurs to ease the burden with their philanthropy. As Bill Gates articulated at the 2008 World Economic Forum, these successful people have an opportunity to do good while doing well for their businesses, by helping the less fortunate with new products and technologies. He declared that “the challenge is to design a system where market incentives, including profits and recognition, drive the change,” with the goal of “improving lives for those who don’t fully benefit from market forces.” He dubbed this system “creative capitalism” and set the philanthropic goal for everybody “to take on a project of creative capitalism in the coming year” as a way of alleviating the world’s problems.

We will argue in this chapter that this vision of almost inexorable benefits from new technology, including intelligent machines, led by talented entrepreneurs is an illusion—the AI illusion. Like Lesseps’s conviction that canals would benefit both investors and global commerce, it is a vision rooted in ideas, but it receives a further boost because it enriches and empowers elites corralling technology toward automation and surveillance.

Even the framing of digital capabilities in terms of intelligent machines is an unhelpful aspect of this vision. Digital technologies are general purpose and can be developed in many different ways. In steering their direction, we should focus on how useful they are to human objectives—what we will call “machine usefulness.” Encouraging the use of machines and algorithms to complement human capabilities and empower people has, in the past, led to breakthrough innovations with high machine usefulness. In contrast, infatuation with machine intelligence encourages mass-scale data collection, the disempowerment of workers and citizens, and a scramble to automate work, even when this is no more than so-so automation—meaning that it has only small productivity benefits. Not coincidentally, automation and large-scale data collection enrich those who control digital technologies.

From the Field of AI Dreams

People are right to be excited about advances in digital technologies. New machine capabilities can massively expand the things we do and can

transform many aspects of our lives for the better. And there have also been tremendous advances. For example, the Generative Pre-trained Transformer 3 (GPT-3), released in 2020 by OpenAI, and ChatGPT released in 2022 by the same company, are natural-language processing systems with remarkable capabilities. Already trained and optimized on massive amounts of text data from the internet, these programs can generate almost human-like articles, including poetry; communicate in typical human language; and, most impressively, turn natural-language instructions into computer code.

Software programs have a simple logic. A program, or algorithm, is a recipe that instructs a machine to take a prespecified set of inputs and perform a set of step-by-step computations. For example, Jacquard's loom took several punched cards as its input and activated an elegantly designed mechanical process, which moved a beam and wove cloth to produce the designs specified in the cards. Different cards created distinct designs, some of them strikingly complex.

Modern computers are referred to as "digital" because the inputs are represented in discrete form, taking one of a finite set of values (most commonly as zeros and ones). But they share with Jacquard's loom the general principle that they implement exactly the sequence of computations or actions that are specified by a programmer.

What about artificial intelligence? Unfortunately, there is no commonly agreed upon definition. Some experts define artificial intelligence as machines or algorithms demonstrating "intelligent behavior" or "high-level capabilities," although what these are is often open to debate. Others provide definitions motivated by programs such as GPT-3, equating intelligent machines with those that have goals, observe their environment, obtain other inputs, and attempt to achieve their objectives. For example, GPT-3 receives distinct goals in different applications and tries to accomplish them as successfully as possible.

Whatever the exact definition of modern machine intelligence, it is clear that new digital algorithms are being applied widely to every domain of our lives. Rather than attempting to arbitrate between different definitions of machine intelligence, we will use "modern AI" to capture the currently prevailing approach in this domain.

Applying digital technologies to the production process—for example, with numerically controlled machinery—long predates modern AI. The

major computing breakthroughs of the past seventy years came from finding ways of performing tasks using software in areas such as document preparation, database management, accounting, and inventory control. Software can also create new production capabilities. In computer-assisted design, it improves the precision and ease with which workers perform design tasks. It makes the work of cashiers and other consumer-facing employees potentially more productive. As we emphasized in Chapter 8, it also enables automation.

To be automated by traditional software, a task needs to be “routine,” meaning that it must involve predictable steps that are implemented in a defined sequence. Routine tasks are performed repetitively, embedded in a predictable environment. For example, typing is routine. So are knitting and other simple production tasks that involve a significant amount of repetitive activity. Software has been combined with machinery that interacts with the physical world to automate various routine tasks, exactly as Jacquard intended, and modern numerically controlled equipment, such as printers or computer-assisted lathes, regularly accomplish this. Software is also an integral part of robotics technology used extensively for industrial automation.

But only a small fraction of human tasks is truly routine. Most of the things our species do involve some amount of problem solving. We deal with new situations or challenges by coming up with solutions that draw analogies on the basis of past experience and knowledge. We employ flexibility when the relevant environment changes constantly. We rely heavily on social interaction, such as communication and explanation or simply the camaraderie that many coworkers and customers enjoy in the process of economic transactions. Collectively, we are a pretty creative species.

Customer service, for example, requires a combination of social and problem-solving skills. There are tens of thousands of problems a customer may encounter, some that are rare or entirely idiosyncratic. It is relatively easy to help a customer who has missed a flight and would like to take the next available plane. But what if the traveler has ended up in the wrong airport or now needs to fly to a new destination?

Modern AI approaches have been used to extend automation into a broader range of routine tasks, such as bank-teller services. Pre-AI automation—for example, using automated teller machines (ATMs)—was

extensive by the 1990s, with a focus on simple tasks, such as dispensing cash. Depositing checks was only partially automated. ATMs accepted deposits, and magnetic-ink-character-recognition technology was used to sort checks according to their bank code and bank account number. But humans were still necessary for other routine tasks, such as recognizing handwriting, organizing accounts, and monitoring overdrafts. Based on more recent advances in AI-based handwriting-recognition and decision-making tools, checks can now be processed without human involvement.

More significantly, the ambition of AI is to expand automation to nonroutine tasks, including customer service, tax preparation, and even financial advice. Many of the tasks involved in these services are predictable and can be automated straightforwardly. For example, information from wage and tax statements (such as the W-2 form in the United States) can be scanned and automatically entered in the relevant fields to compute tax obligations, or the relevant information about deposits and balances can be provided to a bank customer. Recently, AI has ventured into more complex tasks as well. Sophisticated tax-preparation software can query users about expenses or items that look suspicious, and customers can be presented with voice-activated menus to categorize their problem (even if this often works imperfectly, ends up shifting some of the work to users, and causes longer delays as customers wait for a human to provide the necessary help).

In robotic process automation (RPA), for example, software implements tasks after watching human actions in the application's graphical user interface. RPA bots are now deployed in banking, lending decisions, e-commerce, and various software-support functions. Prominent examples include automated voice-recognition systems and chatbots that learn from remote IT-support practices. Many experts believe this kind of automation will spread to myriad tasks currently performed by white-collar workers. *New York Times* journalist Kevin Roose summarizes RPAs' potential as follows: "Recent advances in A.I. and machine learning have created algorithms capable of outperforming doctors, lawyers and bankers at certain parts of their jobs. And as bots learn to do higher-value tasks, they are climbing the corporate ladder."

Supposedly, we will all be the beneficiaries of these spectacular new capabilities. The current CEOs of Amazon, Facebook, Google, and Microsoft have all claimed that AI will beneficially transform technology in

the next decades. As Kai-Fu Lee, former president of Google China, puts it, “And like most technologies, AI will eventually produce more positive than negative impacts on our society.”

The evidence does not fully support these lofty promises, however. Although talk of intelligent machines has been around for two decades, these technologies started spreading only after 2015. The takeoff is visible in the amount that firms spend on AI-related activities and in the number of job postings for workers with specialized AI skills (including machine learning, machine vision, deep learning, image recognition, natural-language processing, neural networks, support vector machines, and latent semantic analysis).

Tracking this indelible footprint, we can see that AI investments and the hiring of AI specialists concentrate in organizations that rely on tasks that can be performed by these technologies, such as actuarial and accounting functions, procurement and purchasing analysis, and various other clerical jobs that involve pattern recognition, computation, and basic speech recognition. However, the same organizations also lower their overall hiring substantially—for example, reducing their postings for all sorts of other positions.

Indeed, the evidence indicates that AI so far has been predominantly focused on automation. Moreover, claims that AI and RPAs are expanding into nonroutine, higher-skilled tasks notwithstanding, most of the burden of AI automation to date has fallen on less-educated workers, already disadvantaged by earlier forms of digital automation. Nor is there any evidence that lower-skilled workers are benefiting from AI applications, although obviously the people who run these firms see some gain for themselves and their shareholders.

Reassuringly, AI does not appear to be advancing so much that it will create mass joblessness. Like the industrial robots we discussed in Chapter 8, current technology thus far can perform only a small set of tasks, and its impact on employment is limited. Nevertheless, it is heading in a direction that is biased against workers and is destroying some jobs. Its most major likely impact is to further lower wages for many people, not create a completely workless future. The problem is that although AI fails in most of what it promises, it still manages to reduce the demand for workers.

The Imitation Fallacy

Why, then, all this emphasis on machine intelligence? What we should care about is whether machines and algorithms are useful to us. For example, according to most definitions, the global positioning system (GPS) may not be intelligent because it is based on the implementation of a straightforward search algorithm (the A* search algorithm, first devised in 1968). Yet GPS devices do provide a tremendously useful service to humans. Almost no expert would classify pocket calculators as intelligent, but they perform tasks that most humans would find impossible (such as quickly multiplying two seven-digit numbers).

Instead of fixating on machine intelligence, we should ask how useful machines are to people, which is how we define machine usefulness (MU). Focusing on MU would guide us toward a more socially beneficial trajectory, especially for workers and citizens. Before developing this case, however, we should understand where the current focus on machine intelligence comes from, which takes us to a vision articulated by the British mathematician Alan Turing.

Turing was fascinated by machine capabilities throughout his career. In 1936 he made a fundamental contribution to the question of what it means for something to be “computable.” Kurt Gödel and Alonzo Church had recently tackled the question of how to define the set of computable functions, meaning the set of functions whose values can be calculated by an algorithm. Turing developed the most powerful way of thinking about this question.

He imagined an abstract computer, now called a Turing machine, that can carry out computations according to the inputs specified on a possibly infinite tape—for example, instructions to implement basic mathematical operations. He then defined a function to be computable if such a machine could compute its values. A machine is said to be a universal Turing machine if it can compute any number that can be calculated by any Turing machine. Notably, if the human mind is in essence a very sophisticated computer and the tasks that it performs are within the class of computable functions, then a universal Turing machine could replicate all human capabilities. Before World War II, however, Turing did not venture into the question of whether machines could really think and how far they could go in performing human tasks.

During the war, Turing joined the top-secret Bletchley Park research facility, where mathematicians and other experts worked to understand encrypted German radio messages. He devised a clever algorithm—and designed a machine—to speed up the breaking of enemy ciphers. This then helped British intelligence to quickly decipher encrypted communications that the Germans had presumed to be unbreakable.

After Bletchley, Turing took the next step in his prewar work on computation. In 1947 he declared to a meeting of the London Mathematical Society that machines could be intelligent. Undeterred by the hostile reactions of participants, Turing continued to work on the problem. In 1951 he wrote: “‘You cannot make a machine think for you.’ This is a commonplace that is usually accepted without question. It will be the purpose of this paper to question it.”

His seminal 1950 paper, “Computing Machinery and Intelligence,” defines one notion of what it means for a machine to be intelligent. Turing imagined an “imitation game” (now called a Turing test) in which an evaluator engages in a conversation with two entities, one human and one machine. By asking a series of questions communicated via a computer keyboard and screen, the evaluator attempts to tell which one is which. A machine is intelligent if it can evade detection.

No machine is currently intelligent according to this definition, but one could turn it into a less categorical ranking of machine intelligence. The better a machine can imitate humans, the more intelligent it is. To make this operational, one can define the notion of “human parity” at a task, which would be achieved if a machine can perform that task at least as well as humans. Then, the more tasks a machine can reach human parity in, the more intelligent it is.

Turing’s own thoughts on this subject were subtler. He understood that passing this test might not mean true thinking capacity: “I do not wish to give the impression that I think there is no mystery about consciousness. There is, for instance, something of a paradox connected with any attempt to localise it.” Despite this reservation, the modern field of AI followed in Turing’s footsteps and focused on artificial intelligence, defined as machines acting autonomously, reaching human parity, and subsequently outperforming humans.

Boom and Mostly Bust

Fascination with machine intelligence often leads to exaggeration. The eighteenth-century French innovator Jacques de Vaucanson would have had a well-deserved place in the history of technology for his many innovations, including the design of the first automatic loom and an all-metal-cutting slide lathe, which was pathbreaking for the early machine-tool industry. Yet today he is remembered as a fraudster for his “digesting duck,” which flapped its wings, ate, drank, and defecated. It was all an illusion, with food and water going into one of the many compartments, which then released already digested food as excrement.

Soon after de Vaucanson’s duck came the Hungarian inventor Wolfgang von Kempelen’s Mechanical Turk, an automated chess-playing machine, whose name originated from the life-size model sitting on top, dressed in an Ottoman robe and turban. The Turk defeated many notable chess players, including Napoleon Bonaparte and Benjamin Franklin; solved the well-known chess puzzle, where a knight must move around touching each square of the board once and only once; and even responded to questions using a letter board. Its success, unfortunately, was thanks to an expert chess player concealed inside the structure.

Claims that machines would soon replicate human intelligence generated great hype in the 1950s as well. The defining event, the first step in the current AI approach and the origin of the term *artificial intelligence*, was a 1956 conference at Dartmouth College, funded by the Rockefeller Foundation. Brilliant young scientists working on related topics convened at Dartmouth during the summer. Herbert Simon, a psychologist and an economist who was later awarded the Nobel Prize, captured the optimism when he wrote that “machines will be capable, within twenty years, of doing any work a man can do.”

In 1970 Marvin Minsky, co-organizer of the Dartmouth conference, was still confident when speaking to *Life* magazine:

In from three to eight years we will have a machine with the general intelligence of an average human being. I mean a machine that will be able to read Shakespeare, grease a car, play office politics, tell a joke, have a fight. At that point the machine will begin to educate itself with

fantastic speed. In a few months it will be at genius level and a few months after that its powers will be incalculable.

These hopes of human-level intelligence, sometimes also called “artificial general intelligence” (AGI), were soon dashed. Tellingly, nothing of great value came from the Dartmouth conference. As the spectacular promises made by AI researchers were all unmet, funding for the field dried up, and what came to be called the first “AI winter” set in.

There was renewed enthusiasm in the early 1980s based on advances in computing technology and some limited success of expert systems, which promised to provide expert-like advice and recommendations. A few successful applications were developed in the context of identifying infectious diseases and some unknown molecules. Soon, claims about artificial intelligence reaching human-level expertise were circulating again, and funding resumed. By the end of the 1980s, a second AI winter was upon the field because the promises were again unfulfilled.

The third wave of euphoria started in the early 2000s, focusing on what is sometimes called “narrow AI,” where the objective is to develop mastery in specific tasks, such as identifying an object in pictures, translating text from a different language, or playing a game such as chess or Go. Reaching or surpassing human parity remained the overarching objective.

This time, rather than mathematical and logic-based approaches intended to replicate human cognition, researchers turned various human tasks into prediction or classification problems. For example, recognizing an image can be conceived as predicting which one of a long list of categories the image belongs to. AI programs can then rely on statistical techniques applied to massive data sets to make increasingly accurate classifications. Social media messages that pass among billions of people are an exemplar of this type of data.

Take the problem of recognizing whether there is a cat in a picture. The old approach would have required a machine to model the complete decision-making process used by humans to spot cats. The modern approach bypasses the step of modeling or even understanding how humans make decisions. Instead, it relies on a large data set of humans making correct recognition decisions based on images. It then fits a statistical model to large data sets of image features to predict when humans say that there is a cat in

the frame. It subsequently applies the estimated statistical model to new pictures to predict whether there is a cat there or not.

Progress was made possible by faster computer processor speed, as well as new graphics processing units (GPUs), originally used to generate high-resolution graphics in video games, which proved to be a powerful tool for data crunching. There have also been major advances in data storage, reducing the cost of storing and accessing massive data sets, and improvements in the ability to perform large amounts of computation distributed across many devices, aided by rapid advances in microprocessors and cloud computing.

Equally important has been progress in machine learning, especially “deep learning,” by using multilayer statistical models, such as neural networks. In traditional statistical analysis a researcher typically starts with a theory specifying a causal relationship. A hypothesis linking the valuation of the US stock market to interest rates is a simple example of such a causal relationship, and it naturally lends itself to statistical analysis for investigating whether it fits the data and for forecasting future movements. Theory comes from human reasoning and knowledge, often based on synthesis of past insights and some creative thinking, and specifies the set of possible relationships among several variables. Combining this theory with a relevant data set, researchers fit a line or a curve to a cloud of points in their data set and make inferences and forecasts on the basis of these estimates. Depending on the success of this first approach, there will be additional human input in the form of a revised theory or a complete change of focus.

In contrast, in modern AI applications the inquiry does not start with clear, causal hypotheses. For example, researchers do not specify which characteristics in the digital version of an image are relevant for recognizing it. Multilayer models, applied to vast amounts of data, attempt to compensate for this lack of prior hypotheses. Each different layer may primarily deal with a different level of abstraction; one layer may represent the edges of the picture and identify its broad outlines, whereas another one may focus on other aspects, such as whether an eye or a paw is present in there. These sophisticated tools notwithstanding, without human-machine collaboration, it is difficult to draw the right inferences from data, and this deficiency motivates the need for ever greater amounts of data and computational power to find patterns.

Typical machine-learning algorithms start by fitting a flexible model to a sample data set and then making predictions that are applied to a larger data set. In image recognition, for example, a machine-learning algorithm can be trained on a sample of tagged images that may indicate whether the image contains a cat. This first step leads to a model that can make predictions on a much larger data set, and the performance of these predictions fuels the next round of algorithmic improvements.

This new AI approach has already had three important implications. First, it has intertwined AI with the use of massive quantities of data. In the words of an AI scientist, Alberto Romero, who became disillusioned with the industry and left it in 2021, “If you work in AI you are most likely collecting data, cleaning data, labeling data, splitting data, training with data, evaluating with data. Data, data, data. All for a model to say: *It’s a cat.*” This focus on vast quantities of data is a fundamental consequence of the Turing-inspired emphasis on autonomy.

Second, this approach has made modern AI appear highly scalable and transferable, and of course, in domains much more interesting and important than recognizing cats. Once the problem of recognizing cats in a picture is “solved,” we can move on to doing the same for more complex image-recognition tasks or to seemingly unrelated problems, such as determining the meaning of sentences in a foreign language. The potential, therefore, is for truly pervasive use of AI in the economy and in our lives—for good but often also for bad.

In the extreme, the aim becomes the development of completely autonomous, general intelligence, which can do *everything* that humans can do. In the words of DeepMind cofounder and CEO Demis Hassabis, the objective is “solving intelligence, and then using that to solve everything else.” But is this the best way to develop digital technologies? This question typically remains unasked.

Third and more problematically, this approach has pushed the field even further in the direction of automation. If machines can be autonomous and intelligent, then it is natural for them to take over more tasks from workers. Companies can break down existing jobs into narrower tasks, use AI programs and abundant data to learn from what humans do, and then substitute algorithms for humans in these tasks.

An elitist vision boosts this focus on automation. Most humans, according to proponents of this view, are error-prone and not very good at the tasks they perform. As one AI website states, “Humans are naturally prone to making mistakes.” On the other hand, there are some very talented programmers who could design sophisticated algorithms. As Mark Zuckerberg puts it, “Someone who is exceptional in their role is not just a little better than someone who is pretty good. They are 100 times better.” Or in the words of Netscape cofounder Marc Andreessen, “Five great programmers can completely outperform 1,000 mediocre programmers.” Based on this worldview, it is desirable to use top-down design of technology by exceptional talent to limit human mistakes and their costs in workplaces. Replacing workers with machines and algorithms then becomes acceptable, and collecting massive amounts of data about people comes to be viewed as tolerable. This approach further justifies reaching human parity, rather than complementing humans, as the criterion for progress and comfortably fits with the emphasis of corporations on cutting labor costs.

The Underappreciated Human

Even with displacement and massive data collection, productivity growth from new technologies can sometimes increase demand for workers and boost their earnings. But benefits for workers appear only when new technologies substantially increase productivity. Today, this is a serious concern because AI has so far brought a lot of so-so automation, with limited productivity benefits.

When productivity increases substantially, this can undo some of the negative effects of automation—for example, by increasing demand for labor in nonautomated tasks or stimulating employment in other sectors that expand subsequently. However, if cost reductions and productivity gains are small, these beneficial effects will not take place. So-so automation is particularly troublesome because it displaces workers but fails to deliver in terms of productivity.

In the age of AI, there is a fundamental reason for so-so automation. Humans are good at most of what they do, and AI-based automation is not likely to have impressive results when it simply replaces humans in tasks for which we accumulated relevant skills over centuries. So-so automation is

what we get, for example, when companies rush to install self-checkout kiosks that do not work well and do not improve service quality for customers. Or when skilled customer-service representatives, IT specialists, or financial advisers are sidelined by AI algorithms, which then perform badly.

Many of the productive tasks performed by humans are a mixture of routine and more complex activities that involve social communication, problem solving, flexibility, and creativity. In such activities, humans draw on tacit knowledge and expertise. Moreover, much of this expertise is highly context dependent and difficult to transfer to AI algorithms, thus likely to get lost once the relevant tasks are automated.

To illustrate the importance of accumulated knowledge, take the foraging societies we discussed in Chapter 4. Ethnographic studies show that hunter-gatherers consistently have a remarkable degree of adaptation to local conditions. For instance, cassava (also known as manioc) is a highly nutritious tuber plant originating in the American tropics. It is used for making cassava flour, breads, tapioca, and various alcoholic beverages. The plant is poisonous, however, because it contains two cyanide-producing sugars. If it is eaten raw or cooked without being properly processed, it can cause cyanide intoxication, with severe consequences in extreme cases, including death.

Indigenous peoples in the Yucatán figured out this problem and developed several practices to remove the poison, including peeling the plant and soaking it for a while before cooking it for a long time, and then disposing of the cooking water. Some Europeans at first did not understand these methods and sometimes interpreted them as primitive, nonscientific traditions, only to find out the perilous costs of not following them.

Human adaptability and ingenuity are no less important in the modern economy, though often ignored by technology-minded elites. There is a strong consensus among city planners and engineers that traffic lights are key to the safe and timely flow of cars. In September 2009 the coastal English town of Portishead turned off the traffic lights at one of its busiest intersections. Against the fears of many experts, drivers started using more common sense and responded adaptively to this new organization. At the end of four weeks, traffic flow had improved significantly at the intersection, with no increase in accidents or injuries. Portishead is no outlier. Several other experiments with

such “naked streets” show similar results. There is a debate on the practicality of naked streets in large cities, and a complete lack of traffic lights is unlikely to be workable at the busiest intersections in megacities. Nevertheless, it is hard not to conclude from these experiments that technology, by taking away initiative and judgment from humans, sometimes makes things worse, not better.

The same is true when it comes to production tasks. Human intelligence derives its strength from being situational and social: The ability to understand and successfully respond to one’s environment, enabling individuals to fluidly adapt to changing conditions. For example, people can be more alert when they are in an unfamiliar environment that provides subtle cues of danger, even while resting or in their sleep. In other environments they perceive as predictable, they can perform tasks faster using learned routines. It is also situational intelligence that helps people respond to changing circumstances more broadly and recognize faces and patterns, using inputs from multiple relevant contexts.

Human intelligence is also social in three important ways. First, a lot of the necessary information for successful problem solving and adaptation resides in the community. We acquire it via implicit and explicit communication—for example, by imitating others’ behavior. Interpreting this type of external knowledge is a vital part of human cognition and is the basis of the emphasis on the “theory of the mind” in this area. Theory of the mind is what enables humans to reason about others’ mental state and thus correctly understand their intentions and knowledge.

Second, our reasoning is based on social communication; we develop arguments and counterarguments in favor of different hypotheses and evaluate our understanding in light of this process. Humans would be terrible decision makers without this social dimension of intelligence. Yes, we make mistakes when placed into lab settings that prevent these aspects of intelligence from being activated, but we avoid some of the same mistakes in more natural settings.

Third, humans gain additional skills and capabilities from the empathy they have for others and the sharing of goals and objectives that this enables.

The central role of the situational and social dimensions of intelligence is related to the weak relationship between analytic aspects of human cognition, as measured by IQ tests, and various dimensions of success. Even in

scientific and technical fields, individuals who are the most successful are those who combine moderately high IQ with social skills and other human capabilities.

In most work environments, situational and social intelligence enables not just flexible adaptation to circumstances but also communication with customers and other employees to improve service quality and reduce mistakes. It is therefore not surprising that despite the spread of AI technologies, many companies are increasingly seeking workers with social, rather than mathematical or technical, skills. At the root of this growing demand for social skills is the reality that neither traditional digital technologies nor AI can perform essential tasks that involve social interaction, adaptation, flexibility, and communication.

All the same, ignoring human capabilities can become a self-fulfilling prophecy because automation decisions can gradually reduce the scope for social interaction and human learning. Take customer service again as an example. Well-trained humans can be very effective in dealing with problems, precisely because they form a social bond with the person needing help (for instance, sympathizing with somebody who just had an accident and needs to file a claim). They can quickly understand the nature of their problem, partly because they are communicating with the customer, and come up with solutions that fit the needs on the basis of this communication. These interactions further enable customer-service representatives to get better at their job over time.

Now imagine the situation after the job of customer service is broken into narrower tasks and the front-end ones are assigned to algorithms, which will often fail to fully identify and deal with the complex problems they encounter. Humans are then brought in as troubleshooters, after a long series of menus. At this point, the customer is often frustrated, early opportunities for building a social bond have been lost, and the customer-service representative does not get the same extent of information from communication, limiting their ability to learn from and adapt to the specific circumstances. This makes the customer-service representative less effective and may encourage managers and technologists to seek additional ways of reducing the tasks allocated to them even further.

These lessons about human intelligence and adaptability are often ignored in the AI community, which rushes to automate a range of tasks, regardless of

the role of human skill.

The triumph of AI in radiology is much trumpeted. In 2016 Geoffrey Hinton, cocreator of modern deep-learning methods, Turing Award winner, and Google scientist, suggested that “people should stop training radiologists now. It’s just completely obvious that within five years deep learning is going to do better than radiologists.”

Nothing of the sort has yet happened, and demand for radiologists has increased since 2016, for a very simple reason. Full radiological diagnosis requires even more situational and social intelligence than, for example, customer service, and it is currently beyond the capabilities of machines. In fact, recent research shows that combining human expertise with new technology tends to be much more effective. For example, state-of-the-art machine-learning algorithms can improve the diagnosis of diabetic retinopathy, which results from damage to blood vessels on the retina among diabetic patients. Nevertheless, accuracy increases significantly more when algorithms are used to identify difficult cases, which are then assigned to ophthalmologists for better diagnosis.

The chief technology officer of Google’s self-driving car division confidently expected in 2015 that his then-eleven-year-old son would not need to get a driver’s license by the time he turned sixteen. In 2019 Elon Musk predicted that Tesla would have one million fully automated, driverless taxicabs on the streets by the end of 2020. These predictions have not come to pass for the same reason. As the naked-streets experiment emphasized, driving in busy cities requires a tremendous amount of situational intelligence to adapt to changing circumstances, and even more social intelligence to respond to cues from other drivers and pedestrians.

General AI Illusion

The apogee of the current AI approach inspired by Turing’s ideas is the quest for general, human-level intelligence.

Despite tremendous advances such as GPT-3 and recommendation systems, the current approach to AI is unlikely to soon crack human intelligence or even achieve very high levels of productivity in many of the decision-making tasks humans engage in. Tasks that involve social and situational aspects of human cognition will continue to pose formidable

challenges for machine intelligence. Once we look at the details of what **has** been achieved, the difficulty of translating existing successes to most **human** tasks becomes clear.

Take the most vaunted successes of AI, such as the AlphaZero **chess** program, discussed in Chapter 1. AlphaZero is even argued to be “creative” because it has come up with moves that human chess masters had not considered or seen. Nevertheless, this is not true intelligence. To start with, AlphaZero is an extremely specialized program and can play only chess and other similar games. Even the simplest tasks beyond chess, such as simple arithmetic or playing games with more social interaction, are beyond AlphaZero’s capabilities. Worse, there is no obvious way in which AlphaZero’s architecture can be adapted to do many of the simple **things** humans do, such as drawing analogies, playing games that have less-strict rules, or learning a language, which is done masterfully by hundreds of millions of one-year-olds every year.

AlphaZero’s intelligence within chess is also very specific. Although AlphaZero’s chess moves within the rules of the game are impressive, they do not involve the type of creativity that humans regularly engage in—such as drawing analogies across unstructured, disparate environments and coming up with solutions to new and varied problems.

Even GPT-3, though more versatile and impressive than AlphaZero, shows the same limitations. It cannot perform tasks beyond those for which it has been pretrained and shows no judgment, so conflicting or unusual instructions can stump it. Worse, this technology has no element of the social or situational intelligence of humans. GPT-3 cannot reason about the context in which the tasks it is performing are situated and draw on causal relationships that exist between actions and effects. As a result, it sometimes misunderstands even simple instructions and has little hope of responding adequately to changing or completely new environments.

Indeed, this discussion illustrates a broader problem. Statistical approaches used for pattern recognition and prediction are ill-suited to capturing the essence of many human skills. To start with, these approaches will have difficulty with the situational nature of intelligence because the exact situation is difficult to define and codify.

Another perennial challenge for statistical approaches is “overfitting,” which is typically defined as using more parameters than justified for fitting

some empirical relationship. The concern is that overfitting will **make** a statistical model account for irrelevant aspects of the data and then **lead** to inaccurate predictions and conclusions. Statisticians have devised **many** methods to prevent overfitting—for example, developing algorithms **on** a different sample than the one in which they are deployed. Nevertheless, overfitting remains a thorn in the side of statistical approaches because **it** is fundamentally linked to the shortcomings of the current approach to AI: **lack** of a theory of the phenomena being modeled.

To explain this problem, it is useful to have a broader understanding of the overfitting problem, based on using irrelevant or **nonpermanent features** of an application. Consider the task of distinguishing wolves from huskies. Although humans are excellent at this task, it turns out to be a difficult one for AI. When some algorithms managed to achieve good performance, it was later understood that this was thanks to overfitting: huskies were recognized from urban backgrounds, such as nice lawns and fire hydrants, and wolves from natural backgrounds, such as snowy mountains. These are irrelevant characteristics in two fundamental senses. First, humans do not rely on these backgrounds for defining or distinguishing the animals. Second, and more troublingly, as the climate warms, wolves' habitats may change, or wolves may need to be identified in different settings. In other words, because the background is not a defining characteristic of wolves, any approach that relies on it will lead to mistaken predictions as the world evolves or the context changes.

Overfitting is particularly troublesome for machine intelligence because it creates a false sense of success, when the machine is in reality performing badly. For instance, a statistical association between two variables, say temperature and GDP per capita across countries, does not necessarily indicate that climate has a sizable impact on economic development. It may simply result from how European colonialism impacted areas with different climatic conditions and in different parts of the globe during a specific historical process. But without the right theory, it is easy to confuse causation and correlation, and machine learning often does this.

The overfitting problem becomes much worse when algorithms are dealing with an inherently social situation where humans react to new information. Human responses will mean that the relevant context evolves frequently, or it may change *because* of the actions they take on the basis of

the information that algorithms provide. Let us give an economic illustration. An algorithm may observe the mistakes a person makes when looking for a job—for example, seeking occupations that have few job postings relative to the number of people applying to them—and may try to correct them. Procedures developed against overfitting, such as separating the training and the testing samples, do not remove the relevant overfitting problem: both samples may be adapted to a particular environment in which there are many unfilled vacancies in the retail sector. But this might change over time precisely because we are dealing with a social situation where humans respond to the available evidence. For instance, as individuals are encouraged by algorithms to apply to them, retail jobs may become oversubscribed and no longer as attractive. Without fully understanding this situational and social aspect of human cognition and how behavior changes dynamically, overfitting will continue to bedevil machine intelligence.

There are other troubling implications of AI's lack of social intelligence. Although it uses data from a large community of users and thus can embed the social dimension of data, with existing approaches it does not leverage the fact that human understanding is founded on selective imitation, communication, and argumentation between people. As a result, many automation attempts appear to reduce, rather than increase, flexibility, which well-trained workers can achieve by rapidly and fluidly responding to changing circumstances, often leveraging skills and perspectives they learn from their coworkers.

Of course, these arguments do not rule out the possibility that a completely new approach can crack the problem of AGI in the near future. Yet there is so far no indication that we are close to coming up with such an approach. Nor is this the main area in which AI dollars are being invested. Industry's focus continues to be on extensive data collection and the automation of narrow tasks based on machine-learning techniques.

The economic problem from this business strategy is clear: when humans are not as useless as sometimes presumed and intelligent machines not as intelligent as typically assumed, we get so-so automation—all the displacement and little of the promised productivity gains. In fact, even the companies themselves do not benefit much from this automation, and some of the AI adoption may be because of hype, as the former AI scientist Alberto Romero, whom we quoted earlier, noted: “The marketing power of AI is

such that many companies use it without knowing why. Everyone wanted to get on the AI bandwagon.”

The Modern Panopticon

Another popular use of modern AI illustrates how enthusiasm for autonomous technology, together with massive data collection, has forged a very specific direction for digital technologies and how it has again caused modest gains for corporations and significant losses for society and workers.

The use of digital tools for worker monitoring is nothing new. When the social psychologist and business scholar Shoshana Zuboff interviewed workers experiencing the introduction of digital technologies in the early 1980s, a common refrain was about the intensification of monitoring by management. As one office worker put it, “The ETS [digital expense tracking system] has become a vehicle for management to check up on us. They can pick up any changes on a minute-by-minute basis if they want to.”

But earlier efforts pale in comparison to what we see today. Amazon, for example, collects a huge amount of data about its delivery workers and warehouse employees, which are then combined with algorithms for restructuring work in a way that increases throughput and minimizes disruptions.

The company, which is the second-largest private-sector employer in the United States, pays higher minimum wages than several other retailers, such as Walmart. But there is a fundamental sense in which Amazon jobs are not good jobs. Workers must abide by strict, fast-paced work routines and are continuously monitored to make sure that they are not taking longer or more frequent breaks and are exerting the required effort at all times. Recent news reports reveal that a sizable proportion of workers from many facilities are fired for not meeting these work expectations, and some of these terminations are automatic, based on the data collected (although Amazon disputes that there are automatic terminations). In the words of a labor advocate, “One of the things we hear consistently from workers is that they are treated like robots in effect because they’re monitored and supervised by these automated systems.”

Jeremy Bentham’s panopticon was meant to be a model not only for prisons but also for early British factories. But eighteenth- and nineteenth-

century bosses did not have the technology for constant surveillance. Amazon does. In the words of one New Jersey employee, “They basically can see everything you do, and it’s all to their benefit. They don’t value you as a human being. It’s demeaning.”

These high-monitoring environments are not just demeaning but also dangerous. A recent OSHA report finds that in 2020, Amazon warehouse workers suffered about 6 serious injuries per 200,000 hours worked, nearly twice as high as the average in the warehousing industry, and other studies find even higher injury rates, especially in peak business times such as the Christmas season, periods during which monitoring of workers intensifies. Amazon additionally requires its delivery employees and contractors to download and continuously run a data tracking app called “Mentor,” which enables closer monitoring. The company recently announced additional AI tools for tracking delivery workers. FedEx and other delivery services also collect a lot of data from their employees and use these for imposing strict scheduling constraints, which explains why many delivery workers are perpetually in a race against time.

Extensive data collection is now spreading to white-collar occupations, with employers tracking how employees use their time on computers and various communication devices.

Some amount of monitoring is part of the prerogative of an employer, who needs to ensure that workers perform the tasks assigned to them and do not damage or misuse machinery. Traditionally, however, workers used to be motivated not just by monitoring but also by the goodwill that developed between them and their employer because of high wages and the general amenities of the workplace. For instance, an employer or a supervisor could recognize that the worker might not be feeling fully well on a given day and cut them some slack, or, conversely, employees could be willing to work harder than usual when the need arose on an occasional basis. Monitoring enables employers to cut wages and get more work out of the workers. In this way, monitoring is a “rent-shifting activity,” meaning that it can be used to prevent sharing of productivity gains and to shift rents away from workers, without improving their productivity much or at all.

Another domain in which AI methods have been deployed for rent shifting is work scheduling. A key source of autonomy for workers is a clear separation between work and leisure time, and predictable scheduling. Take

employees at fast-food restaurants. If they know that they must come to work at 8 a.m. and leave at 4 p.m., this gives them a high degree of predictability and some amount of autonomy beyond this eight-hour window. But what happens if the manager suddenly finds out that there will be many more customers coming after 4 p.m.? She may have an incentive to reduce this autonomy and order employees to stay past 4 p.m. Can she do that?

The answer depends on countervailing powers—for example, collective agreements preventing such impositions; on goodwill and norms of what is acceptable in workplaces; and on technology, which determines whether companies can predict demand in advance and arrange real-time scheduling.

Countervailing powers were already absent, especially in the service industry, and goodwill and norms of respect for worker autonomy had long subsided. The remaining barrier, technology, has now been overcome with AI and massive data collection, paving the way to “flexible scheduling.”

Many customer-facing industries have abandoned predictable schedules, such as 8 a.m.–4 p.m. work hours, adopting instead a combination of “zero-hour contracts” and real-time schedule changes. Zero-hour contracts mean that the company rescinds the commitment that it will employ and pay the worker for regular hours every week. Real-time scheduling, on the other hand, allows companies to call employees on their cell phone the night before asking them to be at work early the following morning or extend their regular hours into longer workdays. It also includes canceling shifts at short notice, which reduces worker income.

Both are rooted in data-crunching and AI technologies—for example, scheduling software offered by tech companies such as Kronos—that enable employers to predict the demand they are going to face and then compel the workers to adapt to it. An extreme version of these practices is “clopening,” the name given to the practice of the same employee closing late one evening and then opening the store early the following morning. This is once again imposed on workers, often at the last moment, as the managers, empowered by AI tools, see it fitting their needs.

There are many parallels between flexible scheduling practices and worker monitoring. The most important is that they are both examples of so-so technologies: they create little productivity gains, despite substantial costs for workers. With additional monitoring, companies can abandon efforts to build goodwill and cut wages. But this does not increase productivity by

much: workers do not become better at their job because they are being paid less, and in fact may lose motivation and become less productive. With flexible scheduling, companies can increase their revenues a little by having more employees when the demand is high and fewer when the store is less busy. In both cases, the burden on workers is more substantial than the productivity benefits. In the words of a British worker employed with a zero-hour contract, “There is no career progression... [I’ve] been in the job for six and a half years. Since then the role hasn’t changed, no promotion. I’ve got no promotional prospects at all. I asked if I could perhaps go on a course, and I got an absolute no for that one.” No matter the costs on workers and the small, ephemeral productivity gains, companies intent on cost cutting and increasing control over workers are continuing to demand AI technologies, and in response researchers beholden to the AI illusion are supplying them.

But is there another way than using digital technologies in the service of ceaseless automation and worker monitoring? The answer is yes. When digital technologies are steered toward helping and complementing humans, the results can be, and have been, much better.

A Road Not Taken

When interpreting both recent and distant history, there is often a deterministic fallacy: what happened had to happen. Often, this is not accurate. There are many possible paths along which history could have evolved. The same is true for technology. The current approach that dominates the third wave of AI based on massive data harvesting and ceaseless automation is a choice. It is in fact a costly choice, not just because it is following the bias of elites toward automation and surveillance, and damaging the economic livelihood of workers. It is also diverting energy and research away from other, socially more beneficial directions for general-purpose digital technologies. We will next see that paradigms prioritizing machine usefulness have had some remarkable successes in the past when tried and offer many fruitful opportunities for the future.

Even before the Dartmouth conference, MIT polymath Norbert Wiener had articulated a different vision, one that positioned machines as complements to humans. Although Wiener did not use the term, MU (machine usefulness) is inspired by his ideas. What we want from machines is not

some amorphous notion of intelligence or “high-level capabilities” but **their** use for human objectives. Focusing on MU, rather than AI, is more likely to get us there.

Wiener identified three critical issues that have stymied dreams of autonomous machine intelligence since Turing. First, surpassing and replacing humans is difficult because machines are always imperfect at imitating living organisms. As Wiener put it in a slightly different context, “The best material model for a cat is another, or preferably the same cat.”

Second, automation had an immediate negative effect on working people: “Let us remember that the automatic machine, whatever we think of any feelings it may have or may not have, is the precise economic equivalent of slave labor. Any labor which competes with slave labor must accept the economic consequences of slave labor.”

And finally, the drive for automation also meant that scientists and technologists could lose control over the path of technology. “It is necessary to realize that human action is a feedback action” means that we adjust what we do based on information about what is happening around us. But “when a machine constructed by us is capable of operating on its incoming data at a pace which we cannot keep, we may not know, until it is too late, when to turn it off.” None of this was inevitable, however: machines could be harnessed to the service of humans as a complement to our skills. As Wiener wrote in an article drafted in 1949 for the *New York Times* (parts of it were published posthumously in 2013), “We can be humble and live a good life with the aid of the machines, or we can be arrogant and die.”

Two visionaries picked up Wiener’s torch. The first was J. C. R. Licklider, who focused on encouraging others to adopt and develop this approach in productive ways. Originally trained as a psychologist, Licklider subsequently moved into information technology, and he proposed ideas that would become critical for networked computers and interactive computing systems. A clear articulation of this vision is contained in his 1960 pathbreaking article, “Man-Computer Symbiosis.” Licklider’s analysis is still relevant today, more than sixty years after its publication, especially in his emphasis that “relative to men, computing machines are very fast and very accurate, but they are constrained to perform only one or a few elementary operations at a time. Men are flexible, capable of ‘programming themselves contingently’ on the basis of newly received information.”

The second proponent of this alternative vision, Douglas Engelbart, also articulated ideas that are precursors to our notion of machine usefulness. Engelbart strove to make computers more usable and easier to operate for nonprogrammers, based on his belief that they would be most transformative when they were “boosting mankind’s capability for coping with complex, urgent problems.”

Engelbart’s most important innovations were revealed in spectacular fashion in a show that was later christened as the “Mother of All Demos.” At a conference organized by the Association for Computer Machinery, jointly with the Institute for Electrical and Electronics Engineers on December 9, 1968, Engelbart introduced the prototypical computer mouse. This contraption, consisting of a big roller, a wooden-carved frame, and a single button, looked nothing like the computer mouse we are used to today, but with wires sticking from its back, it did look enough like a rodent to get the name. It transformed what most users could do with computers at one fell swoop. It was also the innovation that propelled Steve Jobs and Steve Wozniak’s Macintosh computers ahead of PCs and operating systems based on Microsoft. Other Engelbart innovations, some of them also showcased at the Mother of All Demos, include hypertext (which now powers the internet), bitmapped screens (which made various other interfaces feasible), and early forms of the graphical user interface. Engelbart’s ideas continued to generate several other advances, especially under the auspices of the Xerox company (and many of these ideas were again critical for Macintosh and other computers).

Wiener, Licklider, and Engelbart’s alternative vision laid the foundations of some of the most fruitful developments in digital technology, even if today this vision is overshadowed by the AI illusion. To understand these achievements, and why they have not received as much attention as the successes of the dominant paradigm, we first need to discuss how MU works in practice.

Machine Usefulness in Action

We can distinguish four related but distinct ways in which digital technologies can be steered in the direction of MU, helping and empowering humans.

First, machines and algorithms can increase worker productivity in tasks they are already performing. When a skilled artisan is given a better chisel or an architect has access to computer-aided design software, their productivity can increase significantly. Such productivity increases need not just come from new tools and can also be accomplished by improving machine design. This is the aspiration of the fields known as human-computer interaction and human-centered design. These approaches recognize that all machines, and in particular computers, need to have certain features to be most productively used by people, and they prioritize designing new technologies that increase human convenience and usability. When successful, as was Engelbart's mouse and graphic user interface, new digital technologies can be what Steve Jobs referred to as "a bicycle for our minds" and expand human skills. Because this approach puts machine capabilities at the service of people, it tends to complement human intelligence.

Although this approach has already generated notable benefits, much more can be done. Virtual- and augmented-reality tools hold tremendous promise to increase human capabilities in tasks such as planning, design, inspection, and training. But applications can go beyond technical and engineering jobs.

The current consensus in the technology and engineering community is summarized by Kai-Fu Lee: "Robots and AI will take over the manufacturing, delivery, design and marketing of most goods." Such claims notwithstanding, as we saw in Chapter 8, efforts to deploy new software tools have been an important source of productivity growth in the context of the German Industry 4.0 program, which has enabled greater flexibility in the face of changing circumstances or demands.

This potential is even better illustrated by Japanese manufacturing, where many companies have prioritized flexibility and worker participation in decision making, even as advanced and sometimes automated machinery has been introduced. This approach was pioneered by W. Edwards Deming, another engineer following the same vision as Wiener, Licklider, and Engelbart. Deming was instrumental in setting up a quality-centered, flexible production approach in Japanese manufacturing. In return, he received the highest honors in Japan, and the Deming Prize has been established in his name. Augmented and virtual reality currently provides many new avenues for this type of human-machine collaboration, including improved

capabilities for precision work by humans, more adaptive designs, and greater flexibility in responding to changing circumstances.

The second type of MU is even more important and was our focus in chapters 7 and 8: the creation of new tasks for workers. These tasks were critical for expanding the demand for both skilled and unskilled workers even as manufacturers such as Ford automated parts of the production process, reorganized work, and transitioned to mass production. Digital technologies have also created various new technical and design tasks over the last half century (even if most companies have prioritized digital automation). Augmented and virtual reality can also generate more new tasks in the future. Education and health care provide a vivid illustration of how algorithmic advances can introduce new tasks. More than four decades ago, Isaac Asimov noted the problem of our current system of education: “Today, what people call learning is forced on you. Everyone is forced to learn the same thing on the same day at the same speed in class. But everyone is different. For some, class goes too fast, for some too slow, for some in the wrong direction.” When Asimov wrote these words, his proposal for personalized teaching was purely aspirational. Short of one-on-one teaching for all students, there was little possibility for such personalization. Today, we have the tools for making personalization a reality in many classrooms. Indeed, it should be possible to reconfigure existing digital technologies for this purpose. The same statistical techniques used for task automation can also be used for identifying in real time groups of students who have difficulties with similar problems, as well as students who can be exposed to more advanced material. The relevant content can then be adjusted for small groups of students. Evidence from the field of education research indicates that such personalization has considerable return and is most useful where exactly society has the greatest need: improving the cognitive and social skills of students from low socioeconomic backgrounds.

The situation in health care is similar: the right type of MU can significantly empower nurses and other health care professionals, and this would be most useful in primary health, prevention, and low-tech medical applications.

The third contribution of machines to human capabilities may be even more relevant in the near future. Decision making is almost always constrained by accurate information, and even human creativity relies on

accessing accurate information in a timely fashion. Most creative tasks require drawing analogies, finding new combinations of existing methods and designs. People doing this work then come up with previously untried schemes that are confronted with evidence and reasoning, and are subsequently further refined. All these human tasks can be helped by accurate filtering and the provision of useful information.

The World Wide Web, often associated with the British computer scientist Tim Berners-Lee, is a quintessential example of this type of aid to human cognition. By the late 1980s, the internet, the global network of computers communicating with one another, had been around for about two decades, but there was no easy way of accessing the trove of information that existed in this network. Berners-Lee, together with Belgian computer scientist Robert Cailliau, extended Engelbart's hypertext idea and introduced hyperlinks to allow information on one site to be linked to the relevant information on other parts of the internet. The two scientists wrote the first web browser to retrieve this information, and named it the World Wide Web or simply the Web. The Web is a milestone in human-machine complementarity: it enables people to access information and wisdom that other humans have produced to a degree essentially unparalleled in the past.

MU can enable many more applications that provide better information to people in their capacities as workers, consumers, and citizens. Recommendation systems, at their best, have this ability: they can aggregate masses of information from others and present relevant aspects to users to aid in their decision making.

The fourth category, based on the use of digital technologies to create new platforms and markets, may turn out to be the most important application of the Wiener-Licklider-Engelbart vision. Economic productivity is inseparable from cooperation and trading. Bringing together people with different skills and endowments has always been a major aspect of economic dynamism and can be powerfully expanded by digital technologies.

A brilliant illustration of this phenomenon is provided by the fishing industry in the southern Indian state of Kerala, which was revolutionized by the use of mobile phones. In some local beach markets in Kerala, fishermen would come in with a good catch of fish but would encounter insufficient demand, driving the price to zero and leaving a lot of the fish to rot. A few kilometers away, a different beach market would have few fish for sale and

many buyers, leading to high prices, unmet demand, and widespread inefficiencies. Beginning in 1997, mobile phone service was introduced throughout Kerala. Fishermen and wholesalers started using mobile phones to acquire information about the distribution of supply and demand across beach markets. Subsequently, price dispersion and fish wastage dropped sharply. The basic economics of this story is clear: communication technology enabled the creation of a unified fish market, and a careful study of this episode documents that both fishermen and consumers benefited significantly.

Opportunities for new connections and market creation are potentially greater with digital technologies, and some platforms have already made use of them. An inspiring example is the mobile currency and money-transfer system M-Pesa, which was introduced in Kenya in 2007 and provides cheap and fast banking services using mobile phones. This system spread to 65 percent of the Kenyan population two years after its introduction and has since been adopted by several other developing countries. It is estimated to have generated broad-based benefits to these economies. As another example, Airbnb has created a new market where people can rent accommodations, expanding choice for consumers and generating competition with hotel chains.

Even in areas such as translation where AI-based automation has been quite successful, there are complementary alternatives based on the creation of new platforms. For example, rather than just relying on fully automated and often low-quality translation, one could also build platforms that bring together people needing higher-quality language service and qualified multilingual people around the world.

New platforms need not be confined to those specializing in monetary transactions. Decentralized digital structures can be used to build platforms for broader forms of collaboration, sharing of expertise, and collective action, as we will discuss in [Chapter 11](#).

The successes of MU we have mentioned are among the most productive applications of digital technologies and have paved the way to myriad other innovations. Nevertheless, they are, overall, marginal to the current direction of AI. For 2016, McKinsey Global Institute estimated that \$20–\$30 billion out of the total global AI spending of \$26–\$39 billion comes from a handful of big tech companies in the United States and China. Unfortunately, as far as

we can tell, most of this spending appears to go toward massive data collection that is targeted at automation and surveillance.

So why are tech companies not developing tools that help humans and at the same time boost productivity? There are several reasons for this, all of them informative about the broader forces we are confronted with. Consider the teaching example, and recall that new tasks, as in this example, are useful in part because they increase productivity by generating meaningful and high-paying jobs for humans—in this instance, for teachers. Yet new teaching tasks imply greater costs for schools already strapped for cash. Most public schools, like other modern organizations, have to focus on containing labor costs and may struggle to hire additional teachers. Consequently, new algorithms for automated grading or automated teaching could appear more attractive to them.

The same is true in health care. Despite the \$4 trillion that the United States spends on health care, hospitals also face budget pressure, and a shortage of nurses became painfully evident during the COVID-19 pandemic. New technologies that increase nurses' capabilities and responsibilities would mean hiring more nurses for higher-quality health care. This observation reiterates a key point: human-complementary machines are not attractive to organizations when they are intent on cost cutting.

Another challenge is that new platforms and methods of aggregating and providing information to users also open up possibilities for novel exploitative uses. The World Wide Web, for instance, has become as much a platform for digital advertisement and propagation of misinformation as a source of useful information for people. Recommendation systems are often used for steering customers to specific products, depending on the platform's financial incentives. Digital tools can provide information to managers not just for better decision making but also for the better monitoring of workers. Some of the AI-powered recommendation systems have incorporated and reintensified existing biases—for example, toward race in hiring or toward race in the justice system. Platforms for ride sharing and delivery have imposed exploitative arrangements on workers lacking protection or job security. Hence, the way in which even the most promising applications of human-machine complementarity are used is still dependent on market incentives, the vision and priorities of tech leaders, and countervailing powers.

Besides, there is an equally insurmountable barrier to human-machine complementarity. Under the shadow of the Turing test and the AI illusion, top researchers in the field are motivated to reach human parity, and the field tends to value and respect such achievements ahead of MU. This then biases innovation toward finding ways of taking tasks away from workers and allocating them to AI programs. This problem is, of course, amplified by financial incentives coming from large organizations intent on cost cutting by using algorithms.

Mother of All Inappropriate Technologies

It is not only workers and citizens in the industrialized world who will pay the price for the AI illusion.

Despite economic growth in many poorer nations over the last five decades, more than three billion people in the developing world still live on less than \$6 per day, making it difficult for them to achieve three square meals each day, together with money for housing, clothing, and health care. Many pin their hopes on technology to alleviate this poverty. New technologies, introduced and perfected in Europe, the United States, or China, can be transferred to and adopted by developing nations and power their economic growth. International trade and globalization are also argued to be critical ingredients in this process, for low-income nations can export the products they produce with advanced technologies.

Success stories of very rapid economic growth, including South Korea, Taiwan, Mauritius, and more recently China, seem to bear this out. Each country achieved per-capita average growth rates of over 5 percent a year for periods of more than thirty years. In all these cases, industrial technologies played a major role in growth, as did exports to world markets.

But how and whether developing countries benefited from technology imports is more nuanced than typically presumed. A few economists, such as Frances Stewart, realized in the 1970s that technology imports may not work, and in fact may make things worse in terms of inequality and poverty, because the West's technologies are often "inappropriate" for the needs of developing nations. African agriculture illustrates the problem. High- and middle-income countries account for almost all the research spending on agricultural technologies, and a significant fraction has been targeted at the

most perennial problem of agriculture: crop pests and pathogens, which are estimated to destroy perhaps as much as 40 percent of the world's agricultural output. For example, the European maize borer, which affects corn in Western Europe and North America, has received a lot of attention, and resistant strains of crops have been developed (including more than five thousand biotech patents and various genetically modified varieties). The same is true for the western corn rootworm, also affecting corn in the United States and parts of Western Europe, and the cotton bollworm, once a key threat against US cotton.

But these crops and chemicals are not very useful for African and South Asian agriculture, which faces different pests and pathogens. The African maize stalk borer, which afflicts the same crops in Africa, and the desert locust, which ravages almost all crops in Africa and much of South Asia, are phenomenal barriers to agricultural productivity in these regions. But these have received much less attention (very few patents and no genetically modified varieties). The overall amount of research dollars and new innovations targeted at the problems of the low-income developing world have been pitiful in general. Estimates suggest that global agricultural productivity could be increased by as much as 42 percent if biotech research effort was redirected away from Western pests and pathogens toward those afflicting the developing world.

New crops and agricultural chemicals targeted predominantly to Western agriculture are an example of inappropriate technology. Stewart's emphasis was not so much on pests and pathogens, but on how capital-intensive new production methods were. For instance, complex industrial machinery in manufacturing and combine harvesters in agriculture may be mismatched to the needs of the developing world, where capital is scarce and creating jobs—good jobs—for their population during the growth process is a major imperative.

Such mismatches are costly for economic development. Developing nations may end up not using new technologies because they are ill-suited to their needs or are too capital-intensive. Indeed, crop varieties developed in the United States are rarely exported to poorer nations, unless they happen to have a very similar climate and pathogens. Even when new technologies developed in advanced economies are introduced in the developing world, the benefits are often limited because the receiving countries may lack the

highly skilled labor required to maintain and operate the latest machines. Additionally, technologies imported from the rich world tend to create a dual structure, with a highly capital- and skill-intensive sector paying decent wages alongside a much larger sector with few good jobs. In sum, inappropriate technologies fail to reduce world poverty and may instead increase inequality both between the West and the rest, and within developing nations.

Many in the developing world were already aware of these imperatives. Some of the most transformative innovations of the twentieth century were developed in what is now referred to as the “Green Revolution,” which was spearheaded by researchers in Mexico, the Philippines, and India. New rice varieties invented in the West were not suitable to the soil and climatic conditions in these countries. A breakthrough came in 1966 with the breeding of a new hybrid rice variety, IR8 rice, which rapidly doubled rice production in the Philippines. IR8 and related cultivars developed in collaboration with Indian research institutes were soon being adopted in India as well and revolutionized that country’s agriculture, in some places increasing yields by as much as tenfold. International funding from the Rockefeller Foundation and the leadership of scientists, especially the agronomist Norman Borlaug, who was later awarded the Nobel Peace Prize for saving more than a billion people from starvation, were instrumental as well.

Today, we are confronted with the mother of all inappropriate technologies, in the form of AI, but there are no efforts analogous to the Green Revolution (nor are many AI researchers attempting to fill Borlaug’s shoes).

Poverty reduction and rapid economic growth in cases such as South Korea, Taiwan, and China did not just come from the import of Western production methods. Economic success resulted from new technologies enabling the human resources of these countries to be used more effectively. In all these cases, the technologies created new employment opportunities for most of the workforce, and the countries themselves also increased investment in education in order to improve the match between the technologies and their population’s skills.

The current trajectory of AI is precluding this pathway. Digital technologies, robotics, and other automation equipment have already increased the skill requirements of global production and started remaking

the international division of labor—for example, contributing to a process of deindustrialization in many developing nations that have workforces primarily consisting of people with low education.

AI is again the next act in this process. Rather than creating jobs and opportunities for the majority of the population in poor and middle-income countries, the current path of AI is raising the demand for capital, highly skilled production workers, and even higher-skilled services, such as from management-consulting and tech companies. These are exactly the resources that are most lacking in the developing world. As in the examples of export-led growth and the Green Revolution, many of these economies have abundant resources that can be used for spearheading economic growth and reducing poverty. But these are the resources that will remain unused if the future of technology moves in the pathways that the AI illusion dictates.

Rebirth of the Two-Tiered Society

The Industrial Revolution started in eighteenth-century Britain, where most of the population had little political or social power. Predictably, the direction of progress and productivity growth in such a two-tiered society initially worsened the living conditions of millions. This began changing only when the distribution of social power shifted, altering technology's course so that it raised the marginal productivity of workers. Also critical were institutions and norms for robust rent sharing in workplaces, ensuring that higher productivity translated into wage growth. This struggle over technology and worker power started to transform the highly hierarchical nature of British society in the second half of the nineteenth century.

In chapters 6 and 7, we followed this process from Britain to the United States, as technological leadership shifted. Twentieth-century US technology moved even more decisively in the direction of raising worker marginal productivity. In this way it laid the foundations of shared prosperity, not just domestically but also in much of the world, as American techniques and innovations spread globally and enabled mass production and the rise of a middle class in scores of countries.

The United States has remained at the forefront of technology over the last fifty years, and its production methods and practices, especially its digital innovations, are still spreading throughout the world, but now with very

different consequences. The US model of shared prosperity broke down as power became concentrated in the hands of big corporations, the institutions and norms of rent sharing unwound, and technology went in a predominantly automating direction starting around 1980.

All of this was underway, and the vision that animates the use of technology for automating work, monitoring, and squeezing out workers was firmly in place, before the latest wave of AI. We were already on our way back to a two-tiered society long before the 2010s. With a heightened AI illusion, we are seeing this process accelerate.

Modern AI amplifies the tools in the hands of tech elites, enabling them to create more ways of automating work, sidelining humans, and supposedly doing all sorts of good deeds such as increasing productivity and solving major problems facing humanity (they claim). Empowered by AI, these leaders feel even less need to consult the rest of the population. In fact, many of them think that most humans are not that wise and may not even understand what is good for them.

The marriage of digital technologies and big business had created a growing number of billionaires by the mid-2000s. Such fortunes multiplied once AI tools started spreading in the 2010s. But this was not because AI turned out to be anything as productive or amazing as its boosters have maintained. On the contrary, AI-based automation often fails to increase productivity by that much. Worse, it is no way to build shared prosperity. It nevertheless enthralls and enriches tycoons and top managers as it disempowers workers and opens up new ways of monetizing information about people, which we will discuss in Chapter 10.

That all of this gets ignored in a mad rush to use digital technologies to automate work and monitor humans is the reason why we have dubbed this new phase of the vision the AI illusion. This illusion is set to intensify in the next decade, as more powerful algorithms are developed, global online connectivity grows, and household appliances and other machines become permanently connected to the cloud, allowing more extensive data collection.

Today we are moving closer to H. G. Wells's *Time Machine* future dystopia. Our society has already become two-tiered. On top there are the big tycoons, who firmly believe they have earned their wealth because of their amazing genius. At the bottom we have regular people whom tech leaders view as error-prone and ripe for replacement. As AI penetrates more

and more aspects of modern economies, it looks increasingly likely that the two tiers will grow further apart.

None of this had to be the case. Digital technologies did not have to be used for just automating work, and AI technologies did not have to be applied indiscriminately to amplify the same trend. The tech community **did** not have to be mesmerized by machine intelligence instead of working on machine usefulness. There is nothing foreordained about this path of technology, nor is there anything inevitable about the two-tiered society **that** our leaders are creating.

There are ways out of our current conundrum by reconfiguring the distribution of power in society and redirecting technological change. Such change will have to work through bottom-up, democratic processes. Ominously, however, AI is also breaking democracy.