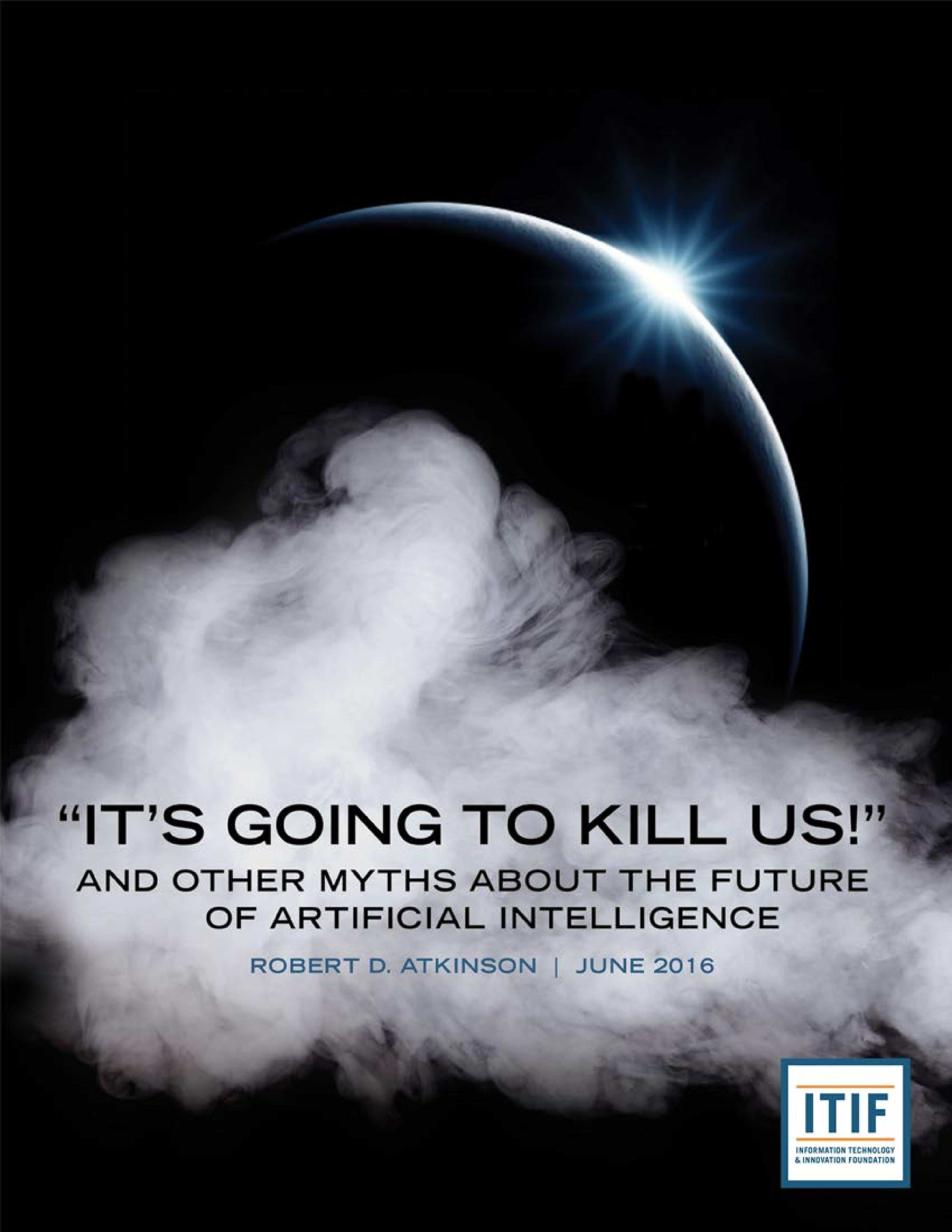


"IT'S GOING TO KILL US!"

AND OTHER MYTHS ABOUT THE FUTURE OF ARTIFICIAL INTELLIGENCE

ROBERT D. ATKINSON | JUNE 2016

ITIF
INFORMATION TECHNOLOGY & INNOVATION FOUNDATION

“It’s Going to Kill Us!” and Other Myths About the Future of Artificial Intelligence

BY ROBERT D. ATKINSON | JUNE 2016

Given the promise that AI holds for economic growth and societal advancement, it is critical that policymakers not only avoid retarding the progress of AI innovation, but also actively support its further development and use.

The past decade has seen many important advancements in computer science that enable software systems to compile and process new information to continually improve the way they function. Indeed, with artificial intelligence (AI), computing is moving from something that “computes” to something that effectively reasons, thinks, and learns. In so doing, it is becoming an ever more powerful and valuable complement to human capabilities: improving medical diagnoses, weather prediction, supply-chain management, transportation, and even personal choices about where to go on vacation or what to buy.

Although artificial intelligence has become commonplace—most smartphones contain some version of AI, such as speech recognition—the public still has a poor understanding of the technology. As a result, a diverse cast of critics, driven by fear of technology, opportunism, or ignorance, has jumped into the intellectual vacuum to warn policymakers that, sooner than we think, AI will produce a parade of horribles: mass unemployment, abuse from “algorithmic bias,” the end of privacy, an atrophying of human agency, and even the destruction of humanity, as “Skynet”-like machines decide the world is better off without us. Indeed, these voices have grown so loud, espousing a message that a click-hungry media eagerly amplifies, that we are very near the point where these narratives may be accepted as truth. Needless to say, when AI is so vociferously demonized (indeed, the engineering magnate Elon Musk has explicitly warned that AI could be “the demon” that threatens our existence, especially if actions are not taken to design systems that can remain under human control), there is a real risk that policymakers will seek to retard its progress.

But AI is like a shovel or a tractor: It is a tool in the service of humans, making our lives vastly better. And given the promise that AI holds for economic growth and societal advancement, it is critical that policymakers actively support its further development and use. The cost of not developing artificial intelligence or developing it more slowly will be enormous: lower growth in per-capita incomes, slower progress in areas such as health and environment, and reduced improvement of quality of a wide array of public and private goods and services. This report explains AI and debunks five prevailing myths that, if left unchecked, could undermine support for it at untold expense to the economy and society:

1. The myth that AI will destroy most jobs,
2. The myth that AI will make humans stupid,
3. The myth that AI will destroy people's privacy,
4. The myth that AI will enable bias and abuse, and
5. The myth that AI will eventually exterminate humanity.

WHAT IS ARTIFICIAL INTELLIGENCE?

Artificial intelligence is a part of computer science devoted to creating computing machines and systems that perform operations analogous to human learning and decisionmaking. Also referred to as deep learning, cognitive computing, machine learning, and machine intelligence, the Association for the Advancement of Artificial Intelligence describes artificial intelligence as "the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines."[1] To be sure, as noted below, "intelligence" can vary greatly, and the term does not automatically imply human-level intelligence. Indeed, your dryer is intelligent if it shuts off when it senses your clothes are dry.

Machine intelligence involves many functionalities, including but not limited to: a) learning, which includes several approaches such as deep learning (for perceptual tasks), transfer learning, reinforcement learning, and combinations thereof; b) understanding, or deep knowledge representation required for domain-specific tasks such as cardiology, accounting, or law; c) reasoning, which comes in several varieties such as deductive, inductive, temporal, probabilistic, and quantitative; and d) interaction, with people or other machines to collaboratively perform tasks, and for learning from the environment. With these capabilities, Guruduth S. Banavar, IBM's vice president for cognitive computing, predicts, "experts of the future will routinely work with learning and reasoning machines to do their day-to-day tasks in a very deep collaborative relationship between people and machines. This is nothing to be fearful of; it is an evolution, and I think it is going to be much better for the world."[2]

The promise of artificial intelligence has been around since the beginning of electromechanical computing, after WWII. But hope has long raced far ahead of reality. At the first AI conference, held at Dartmouth College in 1956, the view was that AI could almost be achieved in one summer's worth of work. In the 1960s and 1970s, some computer scientists predicted that within a decade we would see machines that could think like humans. In 1965, the Nobel laureate Herbert Simon predicted that "machines will be capable, within 20 years, of doing any work a man can do."[3] Two years later, AI pioneer Marvin Minsky predicted, "In from three to eight years we will have a machine with the general intelligence of an average human being."[4] Around the same time, legendary computer scientist John McCarthy told the Defense Department that it would be possible to build "a fully intelligent machine" in the space of a decade.[5]

Needless to say, the timeline of that promise's emergence was vastly underestimated (as likely will be the case with current predictions that human-level AI will emerge in our lifetimes). Even the minimal progress that was achieved came slowly, leading to two distinct periods those in the field call "AI winters," when the shine was off the apple and funding for AI shrunk. Fortunately, we finally appear to have entered a period of "AI spring," with regular announcements of AI breakthroughs: IBM's Watson wins against *Jeopardy* champions; a Facebook algorithm recognizes digital pictures and communicates the information in audio form; Google's AI entry beats the world champion Go player; Siri's creators develop an AI voice assistant that will order a pizza.[6]

The reason for these advancements is the emergence of what computer scientists refer to as "machine-learning" systems. Before these learning algorithms, computer scientists had to program a wide array of functions into a system for it to mimic intelligence. As computer scientist Pedro Domingos writes in his book *The Master Algorithm,* "learning algorithms are algorithms that make other algorithms."[7] This capability has emerged for a number of reasons, including better hardware (faster processers, more abundant storage, etc.) and better algorithms. Domingos writes about how machine learning already surrounds us in our daily lives, from a Nest thermostat that learns from a user's preferences, to Pandora's music service that learns to recommend music, Bing's airfare predictions, Google's automatic language translator, Yelp's system that helps us find a place to eat, or Match.com's method of finding people ideal mates. As the technology progresses, there is no doubt that machine-learning capabilities will continue to broaden and improve, bringing significant benefits to individuals and societies along the way.

Coupled with the widespread job destruction following the great recession, AI has become a fixture in the news (and the movies), looking alternatively promising and threatening.

An important development in machine learning is deep learning. As computer scientist Amit Karp writes, "deep learning relies on simulating large, multilayered webs of virtual neurons, which enable a computer to learn to recognize abstract patterns."[8] It is called "deep" because there are multiple layers of processing. The new versions work with more layers, making the network deeper, hence the name deep learning. These programs can try out a wide range of options, and on the basis of statistical algorithms, automatically identify patterns that previously were hidden in large data sets. In other words, deep-learning systems are able to extract insights out of large data sets and then apply them to do a particular task better the next time.

THE CURRENT ENVIRONMENT FOR AI POLICY

For more than half a century after the first theoretical discussions about AI, computer scientists involved in the field labored in relative obscurity, because there was little popular interest in the field. Occasionally, a particular AI breakthrough would make the news, as when IBM's Deep Blue beat chess champion Garry Kasparov, but by and large only those directly involved in the discipline thought much about it. The U.S. government was interested and supportive of AI, but principally through its support for scientific research, especially through the Defense Department.

That has all changed in the last decade, in large part because of the new advances in machine leaning that have finally shown broader promise for AI. Coupled with the widespread job destruction following the great recession, AI has become a fixture in the news (and the movies), looking alternatively promising and threatening.

And into this milieu have jumped both fanatical true believers in AI—people who are enthusiastically awaiting the "singularity" when they will be able to upload their minds into silicon—and fearful machine-learning Luddites, who are convinced that AI sci-fi movies are close to the mark and that AI is poised to take over humanity. Against this backdrop, it is easy to see how serious policy debates about AI are close to going off the rails. For both the Singularitarians and the Luddites, strong AI is near at hand, so we are poised for either Utopian transformation or apocalyptic destruction: Take your pick. Because human-level AI could emerge any year, according to the prevailing narrative, it is high time to put in place safeguards and restrictions.

This has turned into a true techno-panic.[9] Indeed, rather than measured, rational discussions about AI, most of the social and political commentary has been hype, bordering on urban myth, and even apocalyptic. *New York Times* reporter John Markoff writes that some say it is possible that "these powerful and productive technologies, rather than freeing humanity, will instead facilitate a further concentration of wealth, fomenting vast new waves of technological unemployment, casting an inescapable surveillance net around the globe, while unleashing a new generation of autonomous superweapons."[10] Nick Bostrom, who has been called "a philosopher of remarkable influence," writes that a world with advanced AI would produce "economic miracles and technological awesomeness, with nobody there to benefit," like "a Disneyland without children," because the AI would first kill us all.[11] Thankfully, reality is much more prosaic: AI will likely continue improving gradually, under human control, providing important economic and societal benefits in the process. Policymakers, take note.

EXCITEMENT VS. PANIC ABOUT AI

There are at least seven key reasons why people make extreme claims about AI and why the claims garner such widespread attention. First, humans have had a longstanding attraction-repulsion relationship with the idea that nonhumans might challenge us. From the Hebrew myth of golem, to Mary Shelley's *Frankenstein*, to movies such as Stanley Kubrick's *2001: A Space Odyssey*, *The Terminator,* and most recently *Ex Machina*, to video commentary pieces with titles such as "We should be more afraid of computers than we are," it has been common to view technology as a threat.[12] As computer scientist and philosopher George Zarkadakis writes, "The influence of Mary Shelley's *Frankenstein* in literature cannot be overstated. It encapsulates the core narrative of fear about science and technology, and about Artificial Intelligence in particular."[13] He goes on to note, "most narratives only add to the confusion by devising fictional characters that are machines with human characteristics, or souls. The conflicting literary narratives of love and fear condition the ways we discuss robots, androids, and intelligent machines."[14] This is why it is common to read statements such as "with algorithms, we don't have an engineering breakthrough that's making life more precise, but billions of semi-savant mini-Frankensteins."[15]

Second, AI systems are complex and specialized, so it is easy to make claims about them, because even outlandish assertions are unlikely to be challenged. Just as many privacy advocates regularly claim that particular new, unused technologies will bring "the end of privacy," it is easy for AI fearmongers to make all sorts of negative claims.[16] In particular, critics like to suggest that because computers are assisting with ever-more complicated choices, they have become their own living, thinking beings. But this is simply wrong. As philosophy professor John Searle wrote about IBM's Watson, "IBM invented an ingenious program—not a computer that can think. Watson did not understand the questions, nor its answers, not that some of its answers were right and some wrong, not that it was playing a game, not that it won—because it doesn't understand anything."[17] To this, James Barrat, a documentarian and author who wrote the anti-AI book *Artificial Intelligence and the End of the Human Era*, blithely responds, "As for whether or not Watson thinks, I vote that we trust our perceptions."[18] By this logic, we should believe the earth is flat. Given that computer scientists often fail to see through the absurdity of these statements, how can nonexperts be expected to recognize the fallacies in these claims? Because of this dynamic, it is easy for alarmists to earn attention and plaudits. In short this all harkens back to Arthur C. Clarke's third law: "any sufficiently advanced technology is indistinguishable from magic." And when it's magic, anything is possible, at least in our imagination.

Rather than measured, rational discussions about AI, most of the social and political commentary has been hype, bordering on urban myth, and even apocalyptic.

Third, the language that alarmists often use to describe advanced computing programs makes it sound as if the programs are like brains, and therefore it is only a matter of time before they exceed human capabilities. Yet, as Zarkadakis writes:

> Unfortunately, this huge difference between computers and humans has become a footnote in the contemporary debate about AI. Again, meaning and language confuse us. We are trapped in metaphor, because there is no other way to communicate amongst ourselves. … We therefore think computers are intelligent not because they are, but because this is how natural language compels us to think. … A computer that behaves intelligently will be considered intelligent, even if it is a philosophical zombie.[19]

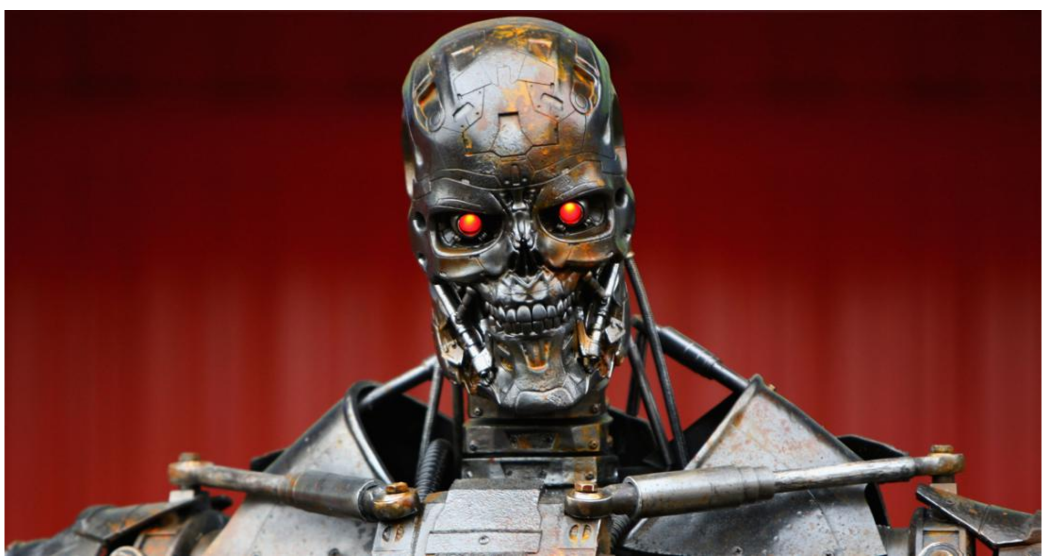
Zarkadakis goes on to ask, "Do computers really 'think'? Is 'intelligence' the same thing as 'consciousness'? Is the brain a 'computer'? Unfortunately, we do not seem to care enough about answering these sorts of questions properly nowadays. In our modern world of mass media and short attention spans, words are increasingly used as flashing slogans."[20] Yann LeCun, founder of the NYU Center for Data Science and now Director of AI Research at Facebook, agrees, writing, "describing AI like the brain gives a bit of the aura of magic to it, which is dangerous. It leads to hype; people claim things that are not true. AI has gone through a number of AI winters because people claimed things they couldn't deliver."[21]

One reason we use humanlike terms to describe AI is that humans appear hardwired to anthropomorphize AI systems, especially robots. Everywhere one looks these days, one sees pictures of robots with smiling faces reaching out to offer a helping hand, or leering menacingly. When we hear the soothing and somewhat attractive voice of Apple's Siri asking what we want, it is easy to think of Siri as a person rather than a bunch of ones and zeros embedded in silicon in a way that stimulates audio waves through a speaker. As futurist Paul Sappho writes, "It's in our nature to infer sentience at the slightest hint that life might be present."[22] Computer scientist Jerry Kaplan agrees, writing:

> Following the tendency for AI researchers to anthropomorphize, they called these programs 'neural networks.' The field has a long history of exploiting our natural tendency to anthropomorphize objects. Without a deep understanding of how these systems work, and with humans as the only available exemplars with which to interpret the results, the temptation to view them as humanlike is inevitable. But they aren't.[23]

As Carnegie Mellon researchers found, "as people interact more with robots, their abstract conceptions of them will become more anthropomorphic."[24] An MIT study asked participants to strike a small robotic "bug" with a hammer, and people were significantly less likely to do so when told that that the robot had a name and "liked to play," among other things.[25] While understandable, this is about as rational as not wanting to kick the vending machine that fails to deliver a candy bar, just because it has voice prompts and someone has drawn a happy face on it and called it Vendy. In other words, by ascribing to robots and AI systems human characteristics, it is easy to believe that they also have autonomous will that could, with the wrong programming or with greater self-awareness, become malevolent. Likewise, referring to artificial intelligence as "superhuman," as Bostrom does, distorts what is really going on and stokes fears. To be sure, in some narrow ways (e.g., beating a person at chess), AI systems are better than humans and thus, in a sense, superhuman. But we do not refer to bulldozers or tractors as superhuman because they can lift 100 times more weight than a human. They are tools that serve our needs, and the same is true of AI.

Figure 1: An increasingly popular image of the future of AI



A fourth reason extreme claims about AI have become commonplace is that an intensely competitive environment for media attention rewards sensational material with more social media likes, retweets, and other forms of viral amplification. This is why when *Time* magazine publishes an article on "What Seven of the World's Smartest People Think

About AI," it features a picture of the Terminator robot. (See Figure 1.) Moreover, while most academics were once content teaching and publishing in scholarly journals, many now define success as giving a popular TED talk or attracting 100,000 followers on Twitter. The best way to get both is to make extreme claims—the more extreme the better.

Fifth, it is easier to succumb to fear of AI when most of the benefits it promises are in the future. As AI develops and more people experience the ways it can improve their lives, it is very likely that attitudes toward AI will become much more positive. As Eric Topol, a medical researcher at Scripps, writes:

As AI develops and more people experience the ways it can improve their lives, it is very likely that attitudes toward AI will become much more positive.

> Almost any medical condition with an acute episode—like an asthma attack, seizure, autoimmune attack, stroke, or heart attack—will be potentially predictable with artificial intelligence and the Internet of all medical things. ... When the time comes, those who fear AI may suddenly embrace it.[26]

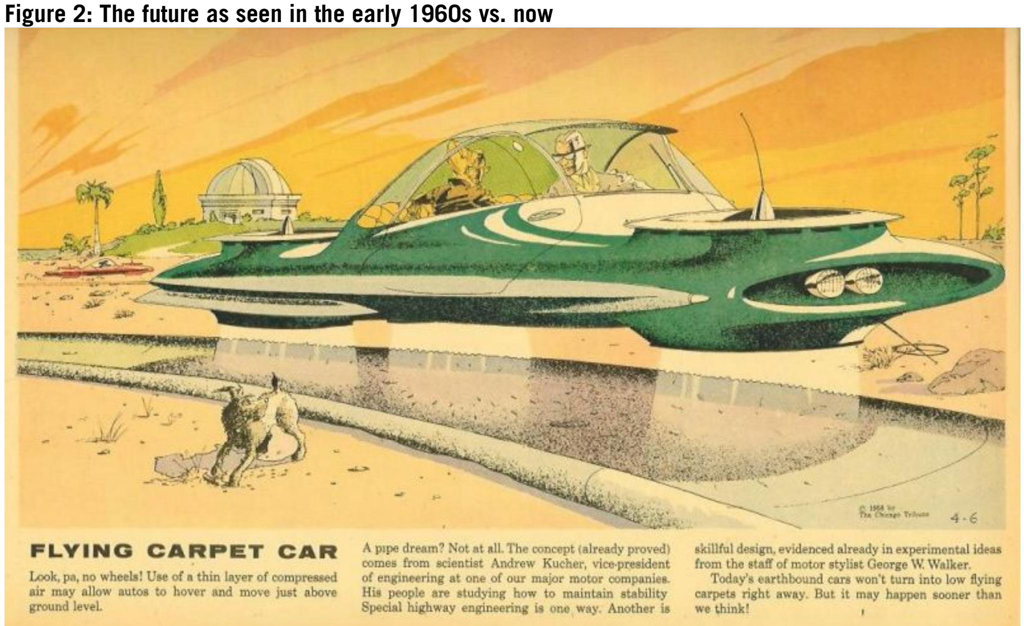
Sixth, it is easy to overestimate the pace of change, and thus scare people into wanting stasis. Indeed, tech forecasters have a long track record of being wildly overoptimistic about the magnitude and pace of future innovations. In 1956, RCA CEO David Sarnoff predicted that by 1980 we would have atomic-powered cars, missile transportation of mail and other freight, and great fleets of personal helicopters.[27] In 1967, 11 years later, futurist Herman Kahn's book *The Year 2000* was published, in which he and a team of leading futurists relied on the new "science" of forecasting to predict which technologies would emerge by the year 2000. But Kahn and his team performed dismally, getting only about 20 percent of their predictions right, and foretelling a host of technologies that don't look likely anytime soon, such as airborne cars, weather control, and lunar settlements. Much like today's believers in the emergence of an artificial general intelligence (AGI) that has human capabilities, Kahn and his team thought this amazing future would be on our doorstep shortly. They wrote:

> This seems to be one of those quite common situations in which early in the innovation period many exaggerated claims are made, then there is disillusionment and swing to over conservative prediction and a general pessimism and skepticism, and then when a reasonable degree of development has been obtained and a learning period navigated, many—if not all—of the early 'ridiculous' exaggerations are greatly exceeded. It is particularly clear that if computers improve by five, ten or more orders of magnitude over the next 33 years, this is almost certain to happen.[28]

In fact, computers improved by many more than 10 orders of magnitude, and yet few of their predictions came to pass. So when today's futurists say things like "Cognitive architectures, whose makers hope will attain human-level intelligence, some believe within a little more than a decade," you can rest assured that they are making the same mistake Sarnoff, Kahn, and so many others did.[29]

Seventh, societal attitudes toward technological change are more pessimistic than they were even a generation ago. Prior to the 1970s, virtually all Americans believed that technological innovation was an unalloyed good and that any efforts to try to stop progress were folly. But a combination of events (e.g., Three Mile Island, Bhopal, global warming), coupled with a more negative and skeptical view toward the benefits of technology, has made techno-skepticism, techno-pessimism, and even techno-panic acceptable, if not fashionable.[30] University social science departments and a media that now focus more on the downsides and risks of new technology stoke this skepticism. To get a sense of this evolution, we have only to compare Roger Radebaugh's hopeful "Closer Than We Think" cartoon series from the late 1950s to today's movies about future technology, like the dark and fearful *Ex Machina* in which a major tech company creates AGI and the results are not pretty. (See figure 2.)

Figure 2: The future as seen in the early 1960s vs. now



Put this all together, and it is not surprising that much of what has been written about the social and economic impacts of AI is so ludicrous that it defies logic. Many claims are so comical in nature that it is surprising that people take them seriously. As Daniel Dennet, co-director of the Tufts University Center for Cognitive Studies, writes:

> The Singularity—the fateful moment when AI surpasses its creators in intelligence and takes over the world—is a meme worth pondering. It has the earmarks of an urban legend: a certain scientific plausibility ('Well, in principle I guess it's possible!') coupled with a deliciously shudder-inducing punch line ('We'd be ruled by robots!') … Wow! Following in the wake of decades of AI hype, you might think the Singularity would be regarded as a parody, a joke, but it has proved to be a remarkably persuasive escalation.[31]

It is not surprising that much of what has been written about the social and economic impacts of AI is so ludicrous that it defies logic.

Former Stanford computer science professor Roger Schank sums it up well: "'The development of full artificial intelligence could spell the end of the human race,' Hawking told the BBC. Wow! Really? So, a well-known scientist can say anything he wants about anything without having any actual information about what he is talking about and get worldwide recognition for his views. We live in an amazing time."[32] In short, believers of artificial super intelligence rely on magical thinking. They have invented in their minds a really advanced form of technology, "super AI." It is so advanced that it is magic. So it can do anything and have any set of behaviors, even truly destructive ones, worse than we can imagine. Indeed, the worse the better, as it triggers the sensation we get in horror movies and propels us to act. Any argument against the prognostications of what it will do and when are countered with, essentially, "You just don't understand how magic it is going to be!"

It is not as if fear of technological progress is new; what is new is that fearful voices now drown out the optimistic ones. Since the dawn of the industrial revolution, Luddites, Marxists, Romantics, and a host of others have decried technological innovation, or at least the forms it was taking in their times. In 1927, the English bishop E. A. Burroughs asked, "Is Scientific Advance Impeding Human Welfare?" and advocated a freeze on scientific research.[33] British philosopher Bertrand Russell, in a 1951 essay, asked, "Are Human Beings Necessary?" while the next year Kurt Vonnegut wrote his first novel, *Player Piano*, in which automation destroys most jobs. Four years later, Robert MacMillan, a professor at Cambridge University, joined a long list of 1950s commentators decrying the new automation technologies (indeed, the term "automation" was coined shortly before this), writing a book called *Automation: Friend or Foe*?[34] In 1961, Norbert Weiner, the famous MIT mathematician, wrote in *Cybernetics,* "Let us remember that the automatic machine is the precise economic equivalent of slave labor."[35]

But while these voices occasionally popped up, they were episodic and quixotic. Society at large had an overwhelming positive view toward technological innovation. As science historian Robert Friedel writes, "at every step along the way in … history there have been debates, sometimes quiet, often violent, about improvement. Who should define it? Who should benefit from it? Who must pay the inevitable costs?" But as Friedel notes, until recently "technology and technological solutions remained for most people in all sectors of society, a source of expected change, and typically, of improvement in everything."[36] This is why when, the father of computer science John von Neumann wrote in 1956 "for progress there is no cure," he was speaking for the vast majority of the population.[37] What has changed is that today the voices of fear and opposition are much more numerous, vocal, and influential.

If we want progress—an increase in economic growth, improved health, a better environment, etc.—then it is time to regain our sense of optimism about the promise of technological innovation. In particular, when it comes to AI, we should be enthusiastic and excited, not fearful and cautious. University of Maryland professor of neurobiology Robert Provine sums it up: "Fear not the malevolent toaster, weaponized Roomba, or larcenous ATM. Breakthroughs in the competence of machines, intelligent or otherwise, should not inspire paranoia about a future clash between humanity and its mechanical creations."[38] Or as Roger Schank writes, "Everyone should stop worrying and start rooting for some nice AI stuff we can all enjoy."[39]

If we want progress, then it is time to regain our sense of optimism about the promise of technological innovation.

While the risks from AI are vastly overstated, it is important to note that, as with any technology, there are risks—but, as with any technology, the developers and managers have strong incentives to reduce and manage those risks. Indeed, the AI community, like all scientific and engineering communities, is not blind to risks. They are actively working to reduce them as they develop the technology further. AI scientists Thomas Dietterich and Eric Horvitz write in *The Communications of the ACM* that there are five kinds of risks from AI: "bugs, cybersecurity, the 'Sorcerer's Apprentice,' shared autonomy, and socioeconomic impacts."[40] They go on to explain that bugs refer to programming errors in the AI software. The second is the risk from cyberattacks. These two risks are no different in nature than current risks of all computer systems and while there will likely continue to be problems, over time scientists and programmers will work to minimize risk, just as engineers work to minimize risk in other technologies like cars, pipelines, and chemical factories. Sorcerer's Apprentice risks refer to the AI system doing the wrong thing—such as an autonomous vehicle driving to the airport at 100 miles an hour when the driver tells it to "get us to the airport as quickly as possible." But this a technical problem; systems can be programmed to ask for clarification ("do you want me to drive 100 miles an hour?") or to drive within certain limits (such as never exceeding the speed limit by more than a set amount over a given distance, regardless of what the driver says). It is also a human factors problem, but one that should diminish as people get used to interacting with AI machines and know how to ask for what they want. The shared autonomy problem refers to AI systems where the machine and the human jointly control things, sometimes sequentially. The problem can arise if the handoff to a human is too sudden and they are not prepared to control it.[41] But as the authors note, "AI methods can help solve these problems by anticipating when human control will be required and providing people with the critical information that they need."[42] Their last point on socioeconomic impacts is discussed below.

It is important to note that risk potential does not provide a reason for slowing AI development. If anything, the fact that there are risks, as there are with every technology, points to why societies should be investing more resources in AI research. Cutting funding for AI research will mean less research on reducing risks. We need funding agencies to devote even more attention and support to these four kinds of risks as AI continues to grow.

In summary, no technology is an unalloyed good, but virtually all technologies that emerge in the marketplace produce benefits vastly in excess of their costs. AI is and will continue to be no different. But it will emerge much more slowly if we do not think rationally about it. And this requires confronting commonly held myths about the impacts of AI.

MYTH 1: AI WILL DESTROY MOST JOBS

Reality: AI will be like past technologies, modestly boosting productivity growth and having no effect on the overall number of jobs and unemployment rates.

Even if AI does not turn into superintelligence, many argue that even modest advances in AI will destroy massive numbers of jobs. We will be alive, but most of will be on the dole. In this increasingly widely held view, AI will power a productivity explosion that will be so great it will destroy jobs faster than the economy can keep up, creating a permanent unemployed underclass dominated by a class of elite "machine owners." Jerry Kaplan writes. "We're about to discover that Karl Marx was right: the inevitable struggle between capital (whose interests are promoted by management) and labor is a losing proposition for workers. … They will offer us the minimum required to keep us satisfied while pocketing the excess profits, just as any smart businessperson does.[43]" Moshe Vardi, a professor at Rice University, predicts that with artificial intelligence, global unemployment will reach 50 percent.[44] Stuart Elliott, in a paper for the National Research Council, extrapolates Moore's Law and argues that in 23 years computers are likely to displace 60 percent of jobs.[45]

Perhaps no one has done more to popularize the idea that AI (and robotics) will destroy jobs than MIT professors Erik Brynjolfsson and Andrew McAfee. In *The Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*, they write that workers are, "losing the race against the machine, a fact reflected in today's employment statistics."[46]

These are not new predictions. For almost half a century, people have been warning of the coming jobs collapse from AI. Economist Gail Garfield Schwartz wrote in 1982 that, with robots and AI, "perhaps as much as 20 percent of the work force will be out of work in a generation." Wasily Leontief wrote in 1983, "We are beginning a gradual process whereby over the next 30-40 years many people will be displaced, creating massive problems of unemployment and dislocation."[47] And the next year Nil Nilson, a computer science professor at Stanford and former head of the American Association for Artificial Intelligence, wrote "We must convince our leaders that they should give up the notion of full employment. The pace of technical change is accelerating." It was no coincidence that they were writing in the midst of one of the worst recessions since the Great Depression when unemployment was high. In 1993, in the wake of the recession of 1990 and 1991, Daniel Crevier wrote in *AI: The Tumultuous History of the Search for Artificial Intelligence*, "The management and service sectors of the economy will be in no position to absorb displaced plant employees, because automation will wreak havoc in these activities."[48]

Today, these claims have resonance for two reasons. The first is that, like the early 1980s and 1990s, the U.S. is recovering from a recession, and many people are conflating job loss from a recession with job loss from technology. The second is because while automation has had a larger effect on some kinds of physical labor in the past, today it is also affecting mental labor. That is scaring many white collar workers who thought their jobs were immune.

This narrative is powered by a few constantly repeated anecdotes that are intended to convince people that "AI is coming for your job next." Whether it is AI programs that can write sports articles or smart document-scanning systems that can perform legal document review, the narrative is that these job-taking technologies are running rampant. What the proponents of this story do not say is that while these technologies can do these narrow tasks reasonably well, they are confined to narrow tasks. An AI system cannot write an article about why the Golden State Warriors performed so well this year in the NBA or prepare a court brief.

The pessimistic, if not apocalyptic, views that AI will kill jobs suffers from two major errors. The first is that the capabilities of AI to replace humans is limited. The second is that even if AI were more capable, there still would be ample jobs.

Before discussing why these fearful prognostications are wrong, it is worth remembering why AI-based productivity is so important to our future. Because if these systems are so powerful, why has productivity in developed economies slowed dramatically since 2009? The simple answer is that without increased productivity, it will be impossible to raise living standards in a sustainable way. As Vice Chairman of the Federal Reserve Bank Stanley Fischer states, "There are few economic issues more important to our economy and others than productivity growth."[49] And going forward AI will be an important tool to boost productivity.

The pessimistic, if not apocalyptic, views that AI will kill jobs suffers from two major errors. The first is that the capabilities of AI to replace humans is limited. The second is that even if AI were more capable, there still would be ample jobs.

AI is limited in its ability to replace workers

It is actually hard, not easy, for technology, AI or otherwise, to eliminate jobs, as evidenced by the fact that if we are lucky, productivity in advanced economies will grow by more than 2 percent per year. A key reason why it is hard to automate jobs with AI is that virtually all AI is "narrow AI," focused on doing one thing really well. Within particular occupations, the introduction of AI may not lead to job loss. Take AI used for radiological analysis. The AI system may in fact become better than a human radiologist at reading images. But radiologists do much more than read images: They communicate with patients and other doctors; they prescribe scans; they interpret results; etc.

In fact, there are thousands of different occupations and, within each occupation, tens if not hundreds of different tasks that workers do. In other words, there would need to be millions of AI systems built to come close to covering even half the jobs in a developed economy, and it would take decades to build systems that could apply AI to all of these. Moreover, for many tasks, AI, no matter how advanced, can't replace human workers.

One reason is that, for many occupations, AI tools don't affect the occupation so much as they affect some tasks performed in an occupation. As the McKinsey Global Institute concludes, "Very few occupations will be automated in their entirety in the near or medium term. Rather, certain activities are more likely to be automated, requiring entire business processes to be transformed, and jobs performed by people to be redefined."[50] In other words, in many cases AI will lead more to job redefinitions and opportunities to add more value, not to outright job destruction. If 20 percent of an administrative assistant's time is spent on tasks that can be automated with an intelligent assistant, that doesn't necessarily mean we lose 20 percent of administrative assistants—it means they can spend that time doing more meaningful things instead of routine tasks such as weekly scheduling.

Even where AI is applicable, a lot of AI tools will augment skills, not replace workers, no matter how it is designed. But many AI dystopians assume that all AI technology will be a substitute for human workers, rather than a complement. Technology author Don Tapscott writes, "Soon the work will stay here but be done by computers. For example, IBM's Watson computer diagnoses cancers with much higher levels of speed and accuracy than skilled physicians do."[51] It may very well do that, but when people get cancer they are not going to have Watson diagnose and treat it using their smartphones. They will go to their doctors, as they always have, and AI technologies will help the physicians make better decisions. Much of AI will improve quality, as systems like Watson do for health care, and will for many other areas such as weather forecasting and education. For these kinds of applications, there is little risk of job loss.

In fact, there was a negative correlation between the risk of automation in an industry as defined by Osborne and Frey and the industry productivity growth of 0.26.

AI will also expand output. As Domingos writes, "Most of all, automation makes all sorts of things possible that would be way too expensive if done by humans. ATMs replaced some bank tellers, but mainly they let us withdraw money anytime, anywhere. If pixels had to be colored one at a time by human animators, there would be no *Toy Story* and no video games."[52] AI will be no different, enabling new things to be done and produced.

This is why predictions of massive job loss from AI will most surely not be fulfilled. One of the most widely cited studies on this matter, from Osborne and Frey, found that 47 percent of U.S. jobs *could* be eliminated by technology over the next 20 years.[53] But they appear to overstate this number by including occupations that have little chance of automation, such as fashion models. Osborne and Frey rank industries by the risk that their workers would be automated. They find that, in accommodation and food services, "as many as 87 percent of workers are at risk of automation, while only 10 percent of workers in information are at risk."[54] While this is a speculation about the future, one would expect that there would be some positive correlation between recent productivity growth and risk of automation. In other words, industries they expect to be most at risk of being automated (by definition, through productivity growth) should have enjoyed higher productivity growth in the last few years, since many of the technologies Osborne and Frey expect to drive automation are already here, albeit not at the same levels of development or deployment. But in fact, there was a negative correlation between the risk of automation in an industry as defined by Osborne and Frey and the industry productivity growth of 0.26.

Moreover, many of the forecasts of job displacement may look daunting, but in fact they are not. Even if Osborne and Frey are right and 47 percent of jobs are eliminated by technology over the next 20 years, this would be equivalent to an annual labor productivity rate of 3 percent a year, just 7 percent higher than the productivity rate of the U.S. economy from 1947 to 1973, when unemployment was at very low levels and job creation was high. Similarly, a Citibank report on the future of work predicted that new developments in computer "algorithms could displace around 140 million knowledge workers globally."[55] This might sound ominous indeed until one realizes that this accounts for just 4.6 percent of global employment and any process is likely to take at least a decade or two to work its way through the labor market. Likewise, a World Economic Forum study proclaims that 5 million jobs could be eliminated by machine learning between now and 2020 in 15 nations. But this accounts for a miniscule 0.05 percent of jobs being lost per year.

Finally, some AI dystopians claim while these limitations exist with current AI, that human-level artificial general intelligence (AGI) will be different, and able to do all jobs and tasks.[56] For these pessimists, AGI will eclipse the full range of human ability—not only in routine manual or cognitive tasks, but also in more complex actions or decisionmaking. The logic is as follows: In order for there to be labor demand, there must be things that humans can do better or more cheaply than machines, but AI is becoming more useful than workers in almost every conceivable way. The gloomy conclusion is we will all be living in George Jetson-land (from the U.S. TV show from the 1960s, *The Jetsons*), but unlike George, we won't be working at Spacely Sprockets; we will be at home on welfare, with only Mr. Spacely employed, because he is the one who owns the robots. But as discussed below, the likelihood of AGI in our lifetime is extremely low. Moreover, as discussed next, unless AI could do every single job in the world, which it will not be able to do (certainly not in the next century or two), even high productivity growth will not lead to fewer jobs because demand for goods and services will grow apace and those jobs that AI can't do will need to be done by humans.

High productivity growth does not increase unemployment rates or inequality

The second mistake AI dystopians make is subscribing to what economists call the "lump of labor fallacy," the view that once a job is gone there are no others created. History, logic, and scholarly evidence are all clear that higher productivity growth does not lead to fewer jobs.

Historically, there has been a negative relationship between productivity growth and unemployment rates. In other words, higher productivity meant lower unemployment. This correlation is shown in the 2011 McKinsey Global Institute report, "Growth and Renewal in the United States: Retooling America's Economic Engine."[57] MGI looked at annual employment and productivity change from 1929 to 2009 and found that increases in productivity are correlated with increases in subsequent employment growth, and that the majority of years since 1929 feature concurrent employment and productivity gains. In looking at 71 10-year slices, only 1 percent had declining employment and increasing productivity. The rest showed increasing productivity and employment. In looking at 76 five-year periods, just 8 percent had declining employment and increasing productivity.

In the 1960s, U.S. productivity grew 3.1 percent per year while unemployment averaged 4.9 percent. However, during the 1980s, productivity grew just 1.5 percent while unemployment rates averaged 7.3 percent. And in the 2000-2007 period, productivity was growing at a healthy 2.7 percent per year while the unemployment rate was under 5 percent. But from 2008 to 2015, productivity growth was only 1.2 percent, yet the unemployment rate averaged over 7.5 percent. Internationally, we see similar patterns. A cross-national sample of productivity growth and average unemployment rates over the period from 1990 to 2011 shows essentially no relationship.[58]

The pessimistic, if not apocalyptic, views that productivity kills jobs suffer not only from a lack of historical perspective, but also from a fundamental flaw in logic. That flaw is not that people who lose their jobs will get jobs making the new machines. No rational organization spends money to increase productivity unless the savings are greater than the costs. If there are the same number of jobs in the company making the machines as there are lost in the companies using the machines, then costs cannot fall.

History, logic, and scholarly evidence are all clear that higher productivity growth does not lead to fewer jobs.

The flaw is that AI dystopians ignore that productivity creates new demand, which in turn creates jobs. If jobs in one firm or industry are reduced or eliminated through higher productivity, then by definition production costs go down. These savings are not put under the proverbial mattress; they are recycled into the economy, in most cases though lower prices or higher wages. This money is then spent, which creates jobs in whatever industries supply the goods and services that people spend their increased savings or earnings on. As a side note, the same logic is true for profits as well. Even if all the savings went to profits, these are distributed to shareholders, who in turn spend at least some of this money, creating demand that is met by new jobs. Even if the shareholders don't spend all of it, the savings reduce interest rates, which leads to new capitalized spending (e.g., car loans and mortgages) and investment, which in turn creates jobs in the firms producing this additional output. Moreover, because of competitive pressures in industries, firms don't have unlimited pricing power. If they did, then firms could just raise prices with impunity. Competitive labor and product markets force firms to pass savings along in the form of lower prices (or higher wages).

Some will argue that people won't spend the money from lower prices or higher wages, and therefore jobs won't be created. But most workers would have little problem finding ways to spend their added income if their take-home pay increased from a doubling or even tripling of productivity. In fact, the first thing most would likely do is break out their shopping lists. To see where the new jobs from higher productivity would likely be created, we only have to look at how those in the top-income quintile spend their money versus those in the middle. According to the Bureau of Labor Statistics, top-income households spend a larger share of their income on things such as education, personal services, hotels and other lodging, entertainment, insurance, air travel, new cars and trucks, furniture, and major appliances. So if U.S. productivity doubles, people would spend at least double on these kinds of goods and services, and employment would grow in these industries.

Even if productivity were miraculously to increase by a factor of 5 or even 10, the vast majority of U.S. households would have no problem spending all this added income.

This would be either as personal consumption (e.g., many people might chose to send their children to better private schools; add an addition to their home; buy a vacation home; eat out more; go to more concerts, plays, and sporting events; hire a cleaning person or personal coach; go on more and better vacations, including flying first class; or buy more expensive foods and wines) or through higher tax revenues spent on public goods (such as a cleaner environment, cities with better urban amenities and design, more and better infrastructure, more police and crime prevention, more spending to tear down and rebuild dilapidated buildings, more social services to help disadvantaged families, etc.). All this is even more likely in developing nations, where median per-capita income is just $6,000. Productivity in these nations could increase by a factor of 50 and still come nowhere near exhausting people's desires for goods and services.

As a recent study by Deloitte notes, technological innovation crates jobs in four different ways.[59] First, in some sectors where demand is responsive to price changes, automation reduces prices but also spurs more demand, leading to at least compensating job creation. For example, as TV prices have fallen and quality increased, people have bought many more TVs. Second, jobs are created making the automation equipment to make the TVs. Third, in some industries, technology serves as a complement to workers, making output more valuable, leading to increased demand. For example, as doctors have gained better technology, the demand for health care has increased. Finally, as discussed above, reduced prices from automation increase consumers' purchasing power, which creates jobs in the industries they spend their new additional income on.

Not only is the notion that productivity kills jobs rebutted by logic and history, virtually all academic studies on the topic have found that productivity increases do not decrease the number of people working nor raise the unemployment rate. If anything, the opposite is true. Trehan found that "The empirical evidence shows that a positive technology shock leads to a reduction in the unemployment rate that persists for several years."[60] The Organisation for Economic Cooperation and Development (OECD) finds that, "Historically, the income generating effects of new technologies have proved more powerful than the labor-displacing effects: technological progress has been accompanied not only by higher output and productivity, but also by higher overall employment."[61] In its 2004 *World Employment Report,* the International Labour Organization found strong support for simultaneous growth in productivity and employment in the medium term.[62] In a paper for the International Labour Organization's *2004 World Employment Report*, Van Ark, Frankema, and Duteweerd found strong support for simultaneous growth in per-capita income, productivity, and employment in the medium term.[63] A study by Industry Canada's Jianmin Tang found that for 24 OECD nations, "at the aggregate level there is no evidence of a negative relationship between employment growth and labour productivity growth… .This finding was robust for rich or poor countries, small or large, and over the pre- or post-1995 period."[64] The United National Industrial Development Organization finds that in fact, "productivity is the key to employment growth."[65]

But even this logic and evidence does little to assuage the fears that AI will take our jobs, because opponents claim it will take all jobs, even if there is increased demand. The narrative is as follows: As automation reduced agricultural jobs, people moved to manufacturing jobs. After manufacturing jobs were automated, they moved to service-sector jobs. But as robots and machine automate these jobs, too, there will be no new sectors to move people into.

A decade ago Brian Arthur wrote, "when farm jobs disappeared, we still had manufacturing jobs, and when these disappeared we migrated to service jobs. With this digital transformation, this last repository of jobs is shrinking—fewer of us in the future."[66] Ray Kurzweil argued that because of Moore's Law, IT will remain on a path of rapidly declining prices and rapidly increasing processing power, leading to developments we can only barely imagine, such as smart robots and bio-IT interfaces.[67] Kurzweil claimed, "gains in productivity are actually approaching the steep part of the exponential curve."[68] (In fact, productivity growth rates fell by half after he wrote this.)

The techno-Utopians' "nowhere left to run" argument is absurd on its face because global productivity could increase by a factor of 50 without people running out of things to spend their increased incomes on.

The techno-Utopians' "nowhere left to run" argument is absurd on its face because global productivity could increase by a factor of 50 without people running out of things to spend their increased incomes on. Moreover, if we ever get that rich, there would be a natural evolution toward working fewer hours.

It is worth noting that the majority of those arguing that AI will eliminate jobs are technologists, not economists. Case in point is computer scientist Jeremy Howard who, in speaking about improvements in machine learning, rightly points out that in developed nations over 80 percent of jobs are services. He then says that AI-enabled computers have learned how to do services (e.g., to recognize images, to speak, etc.). Therefore, ipso facto 80 percent of jobs in advanced economies will be eliminated.[69] But to jump from the fact that machines can "learn" to speak Chinese, recognize patterns in X-rays, and write short descriptions of pictures to the statement that they will eliminate jobs such as barbers, trial lawyers, social workers, gardeners, policemen, etc., is sloppy at best.

There is one other problem with the way AI dystopians frame the issue. They focus largely on the lost jobs, not on gained benefit. We see this when John Markoff discusses the benefits and costs of AI-enabled autonomous vehicles: "More than 34,000 people died in 2013 in the United States in automobile accidents, and 2.36 million were injured. Balance that against the 3.8 million people who earned a living by driving commercially in the United States in 2012."[70] There is frankly not very much to balance here. Autonomous vehicles will save over $1 trillion a year, much of that due to significantly reduced traffic accidents. The benefits from saving so many lives is simply not comparable to the costs to truck drivers, some of whom would, over time, be forced take weeks or even a few months to find a new job. To be sure this is not to dismiss the costs to dislocated workers or the need to put in place better policies to ease transitions to new employment. It is to say that we have to focus on overall societal benefits, and not work to prevent any and all dislocations.

One of many problems with these claims of AI-induced job loss is that they are likely to lead policymakers down the wrong path, supporting policies that would limit rather than accelerate and expand the use of AI. In his book *Machines of Loving Grace,* John Markoff mostly rejects the idea that we should use technology to substitute for labor, instead arguing it should be a complement. He writes, "One approach supplants humans with an increasingly powerful blend of computer hardware and software. The other extends our reach intellectually, economically, and socially using the same ingredients."[71] He goes on to write, "The separation of the fields of AI and human-computer interaction, or HCI, is partly a question of approach, but it's also an ethical stance about designing humans either into or out of the systems we create."[72]

Actually, if either of these directions is more "ethical," it is the decision to design the human out of the system, for that it is the best way to boost productivity, which as noted above is the single most important factor determining standard of living. It is noteworthy that none of individuals making these arguments against AI that replaces workers actually have unpleasant jobs, such as picking grapes, collecting garbage, or transcribing data. If they were doing those jobs, they would likely think twice about their opposition to automation.

The focus on technology as only a supplement and not a replacement to workers is the wrong approach.

But in any case, the focus on technology as only a supplement to workers is the wrong approach. If AI is only a supplement, unless it increases quality or output, organizations will not be able to afford it. Some who argue for AI as only a supplement assume that "money is no object" and we should be able to provide people with the goods and services they "deserve," simply by willing it or requiring it. For example, Aaronson argues that "Every elder deserves the attention of an educated, skilled and compassionate worker."[73] Zeynep Tufekei, a social scientist at the UNC Chapel Hill, agrees, writing "The country—and the world—is awash in underemployment and unemployment, and many people find caregiving to be a fulfilling and desirable profession. The only problem is that we—as a society—don't want to pay caregivers well and don't value their labor."[74] There is a reason we don't pay caregivers well: Doing so would by definition mean spending less money on other things that people also value, such as health care, entertainment, and housing. If we want to improve caregiving, the only way to do it is to raise productivity in that function or the rest of the economy so we can consume more caregiving without reducing other consumption or increase the productivity of caregiving so we don't have to give up so much consumption to get more caregiving. Both require technology, including AI, to boost productivity.

Finally, the idea that a small class of elites will enrich themselves through control of a strong AI in the future, with the vast majority of humanity becoming an impoverished lumpen proletariat is something out of a sci-fi movie. AI won't overturn the laws of supply and demand and competitive markets. Competitive markets, backed up by thoughtful competition policy, mean that profits as a share of national income are by definition limited. If profits get too high in any particular industry, new players enter, which in turn drives down prices and profits. This is not to say that profits cannot increase somewhat, as they have in the last decade in the United States (although they are at levels near what they were in the 1960s when median wage growth was rapid), only that dystopian predictions of a few robot owners lording it over the rest of us are fantasy. In other words, the development of better AI will not change the fundamental relationship between capital and labor. There will still be firms competing in markets. There will be still be workers in labor markets. Finally, it is important to note that as the Economic Policy Institute has shown, it was not technology that increased income inequality in the last several decades; it was changes in political economy factors. If anything, high productivity means lower income inequality.[75]

In sum, the worries of machines overtaking humans are as old as machines themselves. Pitting man against machine only stokes antipathy toward technology and could have a chilling effect on AI innovation and adoption. For example, UK scholar Anthony Atkinson advocates that policymakers encourage "innovation in a form that increases the employability of workers." In other words, support innovation that does not boost productivity.[76] This is in fact the very last thing economies need. The reality is that, far from being doomed by an excess of technology and productivity, the real risk is being held back by too little. To be sure, there are winners and losers in the process of productivity improvement: Some workers will lose their jobs, and it is appropriate for policymakers to help those workers quickly transition to new employment. But to say that we should worry about productivity growth reducing the overall number of jobs available and implement basic income supports to the anticipated large mass of nonworkers is simply without merit. In fact, implementing the now fashionable idea of a basic minimum income (BMI) would actually create unemployment by reducing consumer demand (from the workers who would have to pay higher taxes to support the unemployed BMI recipients). Aid to unemployed workers should be designed to get them back working and producing output for society, not to keep them on the welfare, watching TV at home.

MYTH 2: AI WILL MAKE US STUPID

Reality: AI will enable us to make smarter decisions.

Even if smart machines won't take our jobs, some AI dystopians assert they will turn us into dumb automatons who can only respond to the more intelligent machine. No one has done more to advance this notion than Nicholas Carr, author of *The Glass Cage: How Our Computers Are Changing Us.* Carr writes, "Automation can take a toll on our work, our talents, and our lives." Carr sees technologies like AI that simplify tasks that used to require more human engagement as decidedly negative. But he's not objecting to AI, just to most technology. He complains that past technologies such as the automatic transmission meant that most people now don't learn to drive a manual transmission car. The power lawn mower led people to lose their skills at using a scythe. And GPS-enabled smartphones led people to lose their way-finding and map-reading skills. To be sure, with all three technologies, human skills at shifting gears, scything grass, and reading paper maps has diminished. But not only have these technologies greatly improved our lives, new skills have emerged, such as the ability to use computers and navigate the Internet.

Carr wants technology to be hard to use, not easy, objecting to technologies that fade into the background, seeing them as dehumanizing. But this is an elitist view that only someone who has had the benefit of technologies such as home heating and electricity would write. Presumably Carr would see the invention of automatic starter motors for cars as dehumanizing, because it eliminated the experience of hand-cranking the engine to start it or indoor plumbing because we lost our skills at pumping water. Try telling that to a villager in Africa or India without potable water.

Some have pointed to technologies such as IBM's chess program that beat chess masters, saying they reduce the interest in chess and lower chess skills. But as Sejnowski writes:

> When Deep Blue beat Garry Kasparov, the world chess champion, in 1997, did human chess players give up trying to compare with machines? Quite the contrary: Humans have used chess programs to improve their game, and as a consequence the level of play in the world has improved.[77] So my prediction is that as more and more cognitive appliances, like chess-playing programs and recommender systems, are devised, humans will become smarter and more capable.[78]

Carr also writes that ready availability of information online, in part through search engines, weakens memory.[79] He seems to challenge Plato, who, 3,000 years ago, quoted the Egyptian king Thamus as "complaining that those who practice writing will stop exercising their memory and become forgetful: they might start believing that wisdom dwells in writings … when it resides in the mind."[80] To be sure, for most people easy access to information reduces the need to memorize it. Indeed, many people now no longer bother to remember phone numbers because they are in their smartphones' speed dial. But this does not mean that their minds are any less effective or developed, only that they can use their minds for more valuable activities than remembering mundane facts. I am reminded of my father, who early in his career was a demand deposit accountant at a bank, and one of his jobs was to reconcile the cash and checks coming in and out of the bank. He did this with a manual adding machine (advanced information technology at the time), but he also double-checked it by manually adding extremely long columns of numbers. Needless to say, my father was a master at adding and subtracting. The fact that I am not does not, however, mean that I have not developed other skills that surpass my father's. As Donald Milchie, the British dean of AI research, said, AI is a remedy to "complexity pollution" because "AI is about making machines more fathomable and more under control of human beings, not less."[81]

The reality is that, far from being doomed by an excess of technology and productivity, the real risk is being held back by too little.

Carr also writes that as technology becomes more automatic, much of that from AI, humans will trust them too much. So when there are errors, as when airline pilots encounter new situations that the autopilot systems can't handle or when an autonomous vehicle (AV) software module fails, humans will be ill prepared to deal with them. There are several problems with this argument. First, he is right that AI systems will not be infallible, and that in some cases people's skills will decay. For example, there is no doubt that when fully autonomous vehicles become adopted widely that many people's driving skills will atrophy, or perhaps even never develop. And there will be almost certainly be cases where an AV's software fails and the person "driving" is not paying attention and gets into a possibly fatal crash. But this one-sided view ignores the fact that, overall, autonomous vehicles will reduce accidents. Carr dismisses these benefits, writing "some routine accidents may be avoided, but we're going to end up with even more bad drivers on the road."[82] Research suggests that full deployment of AVs would save lives, leading to around 30,000 fewer fatalities and generating $900 billion in savings a year in the U.S. alone from fewer accidents.[83] Likewise, MIT's David Mindell writes, with regard to aviation, "Digital avionics and software, to be sure, have succeeded in simplifying and improving their interfaces. Their safety record is exemplary, and on balance they have certainly improved safety."[84] So the issue is not whether AI systems won't make errors; the issue is whether on net they will make fewer errors than humans. And the answer is yes, they will make fewer errors.

Related to this, Carr speaks of overdependence on AI automation, where people trust AI systems and assume the system is giving accurate advice when, in fact, it may not always be. For example, he argues that computer-aided detection systems for radiology can lead to radiologists making type 1 errors (not identifying a problem when one exists) and type 2 errors (finding a problem where there is not one).[85] He then claims that "doctors will face increasing pressure, if not outright management fiat, to cede more control over diagnosis and treatment decisions to software."[86] But Carr focuses on speculative examples, and not overall net change in outcomes. On net, AI systems improve decisionmaking, and while there is a potential for some errors to be made, the real question is whether there are fewer errors with computer-aided systems. If there were not, the systems would not be used, assuming similar overall costs. This is why, according to a study published in *The Journal of the American Medical Association*, 50 percent of doctors see the effects of health IT as positive, with just 28 percent saying it is negative.[87] Nor does Carr talk about the fact that AI systems now make real progress in helping diagnose cancer more quickly and enabling more personalized treatments.[88] Moreover, there is considerable evidence that human errors are more prevalent than AI errors. As Gigerzner writes "Studies consistently show that most doctors don't understand health statistics and thus cannot critically evaluate a medical article in their own field."[89] Behavioral economist Richard Thaler writes:

The issue is not whether AI systems won't make errors; the issue is whether on net they will make fewer errors than humans. And the answer is yes, they will make fewer errors.

> Any routine decision-making task—detecting fraud, assessing the severity of a tumor, hiring employees—is done better by a simple statistical model than by a leading expert in the field. So pardon me if I don't lose sleep worrying about computers taking over the world. Let's take it one step at a time, and see if people are willing to trust them to make the easy decisions at which they're already better than humans.[90]

In short, almost no one complains about autonomous technology that has replaced humans because, by definition, if it has replaced humans, it is doing a better job than humans (or is boosting productivity). And Norvig writes:

> With regard to autonomy: If AI systems act on their own, they can make errors that might not be made by a system with a human in the loop. Again, this valid concern is not unique to AI. Consider our system of automated traffic lights, which replaced the human direction of traffic once the number of cars exceeded the number of available policemen. The automated system leads to some errors, but this is deemed a worthwhile tradeoff.[91]

In my own case, 15 years ago I was diagnosed with throat cancer, but my doctor misdiagnosed the source of the primary tumor, even though the latest medical research had shown that unknown primaries for throat cancer almost always originated in the tonsils. But my doctor—highly renowned as one of the best in Washington, DC, for this particular type of cancer—did not have the time to read the hundreds of scholarly oncology journals, and so prescribed the wrong course of treatment. It was only through luck and determination that I spoke with a doctor involved in analyzing the tumor biopsy who told me about the new research study. With that information, I switched doctors, and my new doctor rightly removed my tonsils, where the cancer had in fact originated. If my original doctor had access to a machine-learning system and had put in my symptoms and test results, he without doubt would have recommended removing my tonsils.

Moreover, these AI-enabled decision support systems are being designed and implemented to support, not replace workers. To be sure, some doctors may end up relying on them too much, but, with the proper training and professional guidelines, most won't. And as noted above, it's not that relying on expert judgement produces the best result. In fact, the combination of machine-learning systems and expert judgement will lead to increased correct diagnoses.

This gets to the real problem with Carr and other anti-technology advocates: Their arguments reflect a deep elitism.

Finally, many AI dystopians argue that AI will end up taking control of our lives, not because "it" wants to, but because it will be so easy for us as humans to give it control. As Markoff writes:

> Now the Internet seamlessly serves up life directions. They might be little things like finding the best place for Korean barbecue based on the Internet's increasingly complete understanding of your individual wants and needs, or big things like an Internet service arranging your marriage—not just the food, gifts, and flowers, but your partner, too."[92]

Kaplan presents an even more apocalyptic view, predicting that humans will blindly follow our computer/AI overlords the way a puppy follows its master. He writes, "Synthetic intellects may ultimately decide what is allowed and not allowed, what rules we should follow. This may start with adjusting driving routes based on congestion but could end up controlling where we can live, what we can study, and whom we can marry."[93] This is nonsense. AI won't change human nature. It might give really good recommendations on where we should live, what we should study, and who the best potential mates for us might be. But for the vast majority of humanity, individuals will still be the ones making those, albeit more informed, decisions.

Second, many of these technologies can be designed in ways that keep the user more engaged. For example, Carr talks about the difference between Airbus and Boeing automation systems and finds that Boeing systems are designed to keep the pilot more engaged. As engineers' experience with AI-assisted systems increases, designers will by default include features like Boeing has done. Indeed, Mindell writes "These findings suggest that new technologies ought to enhance human problem solving, not eliminate it. If possible, automation ought to aid humans in their tasks without distancing them from the machine, and without alienating them from their professions."[94]

Finally, when Carr complains about AI reducing skills and capabilities, he is advocating for society to make a choice between embracing AI with these challenges or rejecting AI and allowing society to be poorer than it has to be, in which case there will be fewer resources to invest in safer roads, in health care, or in public education, which will come at considerable economic and human costs. This gets to the real problem with Carr and other anti-technology advocates: Their arguments reflect a deep elitism. As they live privileged, upper-middle class existences as knowledge workers, they at the same time counsel against technological improvement that can relieve billions of people from performing dangerous and mind-numbing jobs while suffering from low incomes. This elitism is evident when Carr dismisses autonomous vehicles, even though he acknowledges they will save lives, because they foreclose other options, such as promoting mass transit or strengthening driver education.[95] Easy for him to say. Try telling that to the parents of a child killed by a drunk driver who didn't bother to brush up on his driver's education or take the bus. Or when he complains that "the working of the soil, which Thomas Jefferson saw as the most vigorous and virtuous of occupations, is being offloaded almost entirely to machines." Clearly he never picked crops or did other backbreaking work. Carr is just one in a long line of privileged Romantics who have criticized technology since the dawn of the industrial revolution, worrying that "the dark Satanic mills" of the late 1700s or the "demonic AI" of today were destroying our humanity, and should be shunned in favor of the more natural order.

MYTH 3: AI WILL DESTROY OUR PRIVACY

Reality: AI will have little effect on privacy. Privacy issues will be here regardless, and most information practices are and will be bound by laws and regulations.

While not yet the focus of as much techno-panic as the effects of AI on human existence and jobs, some have argued that AI will destroy privacy, through its ability to more automatically collect and analyze information. In 1993 Daniel Crevier wrote that AI systems "could be tuned to listen for a few hundred key words, which would increase the effective surveillance power of any single human monitor by orders of magnitude. By letting the system filter out the tedious bits, an AI-assisted listener could process four hundred hours of tape in, say, two hours."[96] More recently, *Globe and Mail* reporter Carly Weeks writes that "new and ever-expanding ways computer technology can search, store and archive information about all of us means that, in many ways, the notion of privacy is becoming obsolete."[97] Matthew Aylett, a computer scientist at Edinburgh University argues that "Given enough data sources you can find out things that people didn't realize. Take the classic idea that if you know people's positions on their phone, you can tell where they move about and guess where they work and where they live. It is very exciting for companies, who can sell things to you based on that."[98] Markoff agrees, writing in overwrought tones, "This neo-Orwellian society presents a softer form of control. The Internet offers unparalleled new freedoms while paradoxically extending control and surveillance far beyond what Orwell originally conceived. Every footstep and every utterance is now tracked and collected, if not by Big Brother then by a growing array of commercial 'Little Brothers'."[99] And Ryan Calo, a law professor at University of Washington, writes:

> AI can be said to threaten privacy according to a specific pattern: AI substitutes for humans at various stages of observation or surveillance, allowing such activity to reach a previously impracticable scale. Whereas once telephonic surveillance required one listener per phone call, the development of voice recognition technology permits the substitution of a computer capable of monitoring

> thousands of calls simultaneously. Whereas once hundreds of intelligence analysts might be required to pour over field records in search of connections, AI knowledge management techniques automatically spot patterns and call them to the attention of agents. These developments vastly amplify the potential for data gathering and analysis, and hence underpin ubiquitous surveillance.[100]

Even some supporters of AI fall prey to this view. Adele Howe, computer science professor at Colorado State University and executive council member of the Association for Advancement of Artificial Intelligence, states, "We have to get over, at some point, the idea that we have privacy. We don't. We have to redefine what privacy means."[101] Besides being wrong, statements like this by AI scientists only fan the fears of AI opposition. We can be assured that as AI becomes more widespread, privacy advocates will seek to raise fears about AI in what the Information Technology and Innovation Foundation (ITIF) calls "The Privacy Panic Cycle."[102]

While AI systems certainly have the ability and even need to collect and analyze more information, the threat to privacy is little greater than the non-AI systems of today.

There are four problems with these claims. First, while AI systems certainly have the ability and even need to collect and analyze more information, the threat to privacy is little greater than the non-AI systems of today. Many organizations already collect personally identifiable data. AI doesn't change that. It might lead to more data being collected, but it won't change privacy in any qualitative way. Moreover, the rules that govern data use and protect privacy today will also cover data analyzed by AI. In short, this is a policy question, not a technology question. Privacy issues will be with us regardless of whether AI progresses or not. Moreover, if we don't want U.S. government agencies, for example, to collect certain data, Congress and the courts can require that. Whether they have or use AI is irrelevant to that.

Second, some argue that AI technologies will give governments unlimited power of surveillance. But governments don't need AI to do that; technology exists today for unlimited surveillance. Government could bug all our phones and install cameras in every room in every home to watch and listen, just as in *1984*. But "it" doesn't, because "it" is us (at least in democratic societies) and "we" neither want it nor will allow it. Going forward we shouldn't fear AI. We should fear, or least be vigilant for, the decline of the rule of law and democracy. That, rather than any particular technology, is the source of any threat to privacy from government overreach.

Third, these privacy dystopians overlook the vast benefits to society from data analytics. As Domingos writes, "Privacy is not a zero-sum game, even though it's often treated like one."[103] Data analytics is already producing important societal benefits in health care, education, transportation, government function, social work, and a host of other areas. For example, data analytics can improve corporate transparency.[104] It is improving medical discovery.[105] It is improving the enforcement of human rights and the protection of fragile ecosystems.[106] This is why, as Domingos writes, "laws that forbid using data for any purpose other than the originally intended one are extremely myopic."[107]

Finally, many, if not most, of the benefits of AI-enabled data analysis can be obtained without risking disclosure of personally identifiable information. As ITIF has shown, data de-identification technologies work very well if properly designed and executed.[108]

Moreover, as MIT's *Technology Review* reports, companies are working on "privacy-preserving deep learning." These systems allow multiple organizations to combine their data to train deep-learning software without having to take the risk of actually sharing it.[109] For example, researchers at Cornell University have developed systems that can train a company's deep-learning algorithms using data such as images from smartphones, without transferring that data to the company.[110]

MYTH 4: THE COMPLEXITY OF AI WILL ENABLE BIAS AND ABUSE

Reality: In most cases, smart machines will be less biased than humans.

There is no question that, by their very nature, AI systems are more complex than traditional software systems where the parameters were built in and largely understandable. It was relatively clear how the older rules-based expert systems made decisions. In contrast, machine learning allows AI systems to continually adjust and improve based on experience. Now different outcomes are more likely to originate from obscure changes in how variables are weighted in computer models.

This has led some critics to claim that this complexity will result in systemic "algorithmic bias" that enables government and corporate abuse. These critics fall into two camps. The first camp believes that companies or governments will deliberately "hide behind their algorithm," using their algorithm's opaque complexity as a cover to exploit, discriminate, or otherwise act unethically. For example, Tim Hwang and Madeleine Clare Elish, of nonprofit organization Data & Society, claim that Uber deliberately uses its surge pricing algorithm to create a "mirage of a marketplace" that does not accurately reflect supply and demand so Uber can charge users more.[111] In other words, they are saying the Uber algorithm is a sham that gives cover to unfair pricing policies.

The second are those, such as Frank Pasquale, author of The *Black Box Society: The Secret Algorithms That Control Money and Information*, who argue that opaque, complicated systems will allow "runaway data" to produce unintended and damaging results.[112] For example, Cathy O'Neil, author of the forthcoming book with the catchy title *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, describes how machine-learning algorithms are likely to be racist and sexist because computer scientists typically train these systems using historical data that reflects societies' existing biases.[113] Likewise, Hannah Devlin asks in the *Guardian*, "would we be comfortable with a[n AI] diagnostic tool that saved lives overall, say, but discriminated against certain groups of patients?"[114]

To be sure, AI, like any technology, can be used unethically or irresponsibly. But resistance to AI because of this concern fails to recognize a key point: AI systems are not independent from their developers and, more importantly, from the organizations using them. For example, once malicious Internet users taught Microsoft's experimental AI chatbot to spew hate speech on the Internet, the company pulled the plug on that particular project.[115] If a government wants to systematically discriminate against certain groups of its citizens, it doesn't need AI to do so. Likewise, if a corporation's goal is to unethically maximize profits at the expense of certain segments of the population, it doesn't need AI. As algorithmic systems researcher Christian Sandvig puts it, "Illegal things are going to happen with and without computers."[116]

Furthermore, if an algorithmic system produces unintended and potentially discriminatory outcomes, it's not because the technology itself is malicious; it's because it simply follows human instructions, or more often relies on data sets from the world that may reflect bias. For example, after researchers at Carnegie Mellon University found instances of targeted advertising algorithms showing ads for elite jobs and job training services to men more than women, critics quickly latched on to the disingenuous narrative that these algorithms were sexist.[117] However, the entire purpose of these algorithms is to allow advertisers to narrowly target specific demographics. No algorithm "decided" that women should not have the same access to economic opportunity as men; the algorithms were simply following instructions about whom to target and how to optimize ad performance by showing the ad to users likely to click on it. If the advertiser wanted to target women instead of men, such algorithms make it incredibly easy to do so; that they did not is the fault of the advertiser, and not because of "runaway data."[118]

Quite simply, these systems will reflect human intention. As Krauss writes:

> We haven't lost control, because we create the conditions and the initial algorithm that determine the decision making. I envisage the human/computer interface as like having a helpful partner; the more intelligent machines become, the more helpful they'll be in partners. Any partnership requires some level of trust and loss of control, but if the benefits often outweigh the losses, we preserve the partnership. If they don't, we sever it. I see no difference in whether the partner is human or machine.[119]

Mindell agrees, writing:

> For any apparently autonomous system, we can always find the wrapper of human control that makes it useful and returns meaningful data. To move our notions of robotics and automation, and particularly the newer idea of autonomy, into the twenty-first century, we must deeply grasp how human intentions, plans, and assumptions are always built into machines. Even if software takes actions that could not have been predicted, it acts within frames and constraints imposed upon it by its creators. How a system is designed, by whom, and for what purpose shapes its abilities and its relationships with people who use it.[120]

Nonetheless, many critics seem convinced that the complexity of these systems is responsible for any problems that emerge, and that pulling back the curtain on this complexity by mandating "algorithmic transparency" is necessary to ensure that the public can police against nefarious corporate or government attempts to use algorithms unethically or irresponsibly.[121] Data and algorithmic transparency, as defined by the Electronic Privacy Information Center (EPIC): "Entities that collect personal information should be transparent about what information they collect, how they collect it, who will have access to it, and how it is intended to be used. Furthermore, the algorithms employed in big data should be made available to the public."[122]

Combatting bias and protecting against harmful outcomes is of course important, but mandating that companies make their propriety AI software publicly available would not actually solve these problems, and would create other problems. As Lauren Smith, policy counsel at the Future of Privacy Forum, writes:

> Consumers and policymakers are unlikely to understand what an algorithm says or means, it would likely undergo continuous change over time or in reaction to new data inputs, and it would be difficult to decide how to measure unfairness—whether by looking at inputs, outputs, decision trees, or eventual effects. These challenges may leave even companies that care deeply about avoiding discrimination unsure as to what best practices really are.[123]

Combatting bias and protecting against harmful outcomes is of course important, but mandating that companies make their propriety AI software publicly available would not actually solve these problems, and would create other problems.

Moreover, the economic impact of such a mandate would be significant, as it would prevent companies from capitalizing on their intellectual property and future investment and research into AI would slow. Other companies would simply copy their algorithms. However, Pasquale will have none of this, claiming that this economic argument is just a nefarious smokescreen: "They [corporations] say they keep techniques strictly secret to preserve valuable intellectual property—but their darker motives are so obvious."[124]

Such calls for algorithmic transparency hinge on a glaring logical inconsistency. Algorithms are simply a recipe for decisionmaking, so if proponents of algorithmic transparency really worry that these decisions are harmful, then they should also call for all aspects of all decisionmaking to be public. Maybe CEOs and other top decisionmakers should be required to take detailed psychological tests to determine their biases? That advocates do not call for such disclosures indicates that such proponents must think regular, human decisions are already transparent, fair, and free from the subconscious and overt biases we know permeate every aspect of society and the economy. This is of course wrong, as the above discussion from Richard Thaler shows. And as Daniel Kahneman writes in his book *Thinking, Fast and Slow*, human irrationality and bias is inherent in being human. Yet Pasquale, EPIC, and others do not argue that companies and governments should have to publicly disclose every kind of decisionmaking process and the underlying data that informed these decisions. For example, research shows that taxicabs frequently do not pick up passengers based on their race, and employers may filter out candidates with African-American sounding names, despite their sufficient qualifications.[125] And yet such critics have not expressed the same support for a legislative mandate that taxi drivers publicly report on why they did not pick up every passenger they passed by and that employers must publish a review of every résumé they receive, with detailed notes explaining why they didn't offer a particular candidate a job.

Equally confounding, these critics acknowledge the ability for AI systems to be incredibly complex and effective but fail to think that companies and governments can responsibly embed ethical principles into algorithmic systems. Damned if they do; damned if they don't. To embed such principles, Nicholas Carr laments, "We'll need to become gods."[126] He asks, "Who gets to program the robot's conscience? Is it the robot's manufacturer? The robot's owner? The software coders? Politicians? Government regulators? Philosophers? An insurance underwriter?"[127] But this ignores the fact that the body politic already "plays god" in most decisions governments make. We decide as a society that a certain number of road fatalities a year is acceptable, because we choose not to spend more on higher priced, but more safely designed roads. We say it is okay for people to die from treatable diseases because we choose not to spend even more on health care. But by definition, societal resources are scarce relative to needs, and decisions must be made on a regular basis about how to allocate these resources among competing uses. And yes, these are "godlike," in that they will result in some people being harmed. But absent massive increases in productivity, societies will have to do the best they can at managing and minimizing the tradeoffs. Programming smart machines will be no different.

Fortunately, many have recognized that embedding ethical principles into machine-learning systems is both possible and effective for combatting unintended or harmful outcomes.

Fortunately, many have recognized that embedding ethical principles into machine-learning systems is both possible and effective for combatting unintended or harmful outcomes. In May 2016, the White House published a report detailing the opportunities and challenges of big data and civil rights, but rather than focus on demonizing the complex and necessarily proprietary nature of algorithmic systems, it presented a framework for "equal opportunity by design"—the principle of ensuring fairness and safeguarding against discrimination throughout a data-driven system's entire lifespan.[128] This approach, described more generally by Federal Trade Commissioner Terrell McSweeny as "responsibility by design," rightly recognizes that algorithmic systems can produce unintended outcomes, but doesn't demand a company waive rights to keep its software proprietary.[129] Instead, the principle of responsibility by design provides developers with a productive framework for solving the root problems of undesirable results in algorithmic systems: bad data as an input, such as incomplete data and selection bias, and poorly designed algorithms, such as conflating correlation with causation, and failing to account for historical bias.[130] Moreover, researchers are making progress in enabling algorithmic accountability. For example, Carnegie Mellon researchers have found a way to help determine why a particular machine-learning system is making the decisions its making, without having to divulge the underlying workings of the system or code.[131]

It also is important to note that some calls for algorithmic transparency are actually more in line with the principle of responsibility by design than with EPIC's definition. For example, former chief technologist of the Federal Trade Commission (FTC) Ashkan Soltani said that although pursuing algorithmic transparency was one of the goals of the FTC, "accountability" rather than "transparency" would be a more appropriate way to describe the ideal approach, and that making companies surrender their source codes is "not necessarily what we need to do."[132] Rather than make companies relinquish their intellectual property rights, encouraging adherence to the principle of responsibility by design and algorithmic accountability would allow companies to better police themselves to prevent unintended outcomes and still ensure that regulators could intervene and audit these systems should there be evidence of bias or other kinds of harm.[133]

Figuring out just how to define responsibility by design and encourage adherence to it warrants continued research and discussion, but it is crucial that policymakers understand that AI systems are valuable because of their complexity, not in spite of it. Attempting to pull back the curtain on this complexity to protect against undesirable outcomes is incredibly counterproductive and threatens the progress of AI and machine learning as a whole.

MYTH 5: AI WILL TAKE OVER AND POTENTIALLY EXTERMINATE THE HUMAN RACE

Reality: We will be lucky if smart machines become smart enough to make us a sandwich.

It's appropriate to end with the scariest myth: Machines become super-intelligent, and for some reason, perhaps to optimize paperclip production, as Nick Bostrom speculates, decide they are better off without humans. At one level it's a sad commentary that we have become so technophobic that we take this science-fiction proposition seriously. But since so many do, let's consider it.

A slew of books, articles, and talks anthropomorphize AI in ways intended to inspire fear, if not downright terror. As MIT scientist David Mindell writes, "Some believe that humans are about to become obsolete, that robots are 'only one software upgrade away' from full autonomy."[134] In his book *Humans Need Not Apply,* AI scientist Jerry Kaplan writes that because AI operates "on timescales we can barely perceive, with access to volumes of data we can't comprehend, they can wreak havoc on an unimaginable scale in the blink of an eye—shutting down electrical grids, placing all airplane takeoffs on hold, cancelling millions of credit cards."[135] He goes on to warn "the root cause is much more sinister—the emergence of invisible electronic agents empowered to take actions on behalf of the narrow self-interests of their owners, without regard to the consequences for the rest of the world. These AI systems "could amass vast fortunes, dominate markets, buy up land, own natural resources, and ultimately employ legions of humans as their nominees, fiduciaries, and agents—and that's in the happy event they deign to use us at all. The slave becomes the master."[136] Calum Chase writes, "An economic singularity might lead to an elite owning the means of production and suppressing the rest of us in a dystopian technological authoritarian regime."[137] But wait: It gets worse, because once scientists develop an AI system as smart as a human (referred to as artificial general intelligence), it is only a matter of time, perhaps even minutes, before it morphs into a Skynet-like artificial super intelligence. Oxford philosopher Nick Bostrom, author of the book *Superintelligence,* argues that once an AI system exists that has attained artificial general intelligence, it likely will rapidly improve its own intelligence, to be super-intelligent. And then watch out, because we are then either dead or slaves.

"The fears of runaway AI systems either conquering humans or making them irrelevant aren't even remotely well grounded."
– Rodney Brooks

To be sure, this is an interesting philosophic question to entertain in the classroom. How would humans react to AGI? How would it react to us? But philosophical speculation is not the same as computer engineering. In fact, the reality is vastly more likely to be in line with what author Rolf Doblei writes, "Artificial thinking won't evolve to self-awareness in our lifetime. In fact, it won't happen in 1,000 years.[138] Or as MIT computer science Rodney Brooks writes, "The fears of runaway AI systems either conquering humans or making them irrelevant aren't even remotely well grounded. Misled by suitcase words, people are making category errors in fungibility of capabilities—category errors comparable to seeing the rise of more efficient internal combustion engines and jumping to the conclusion that warp drives are just around the corner."[139]

Slowdown in Moore's Law

There are two reasons not to worry about AGI taking over any time soon. The first relates to the fact that most of these predictions are premised on steady if not accelerating progress in Moore's Law. Moore's Law is named after one of the founders of Intel, Gordon Moore, who famously predicted that the speed of computer processing would double every 18 to 24 months even as the price of that computing power halved. The fact that Moore's law has continued, although more slowly over the last decade, has led many to believe that it will keep growing ad infinitum and in so doing, make AGI not only possible but inevitable.

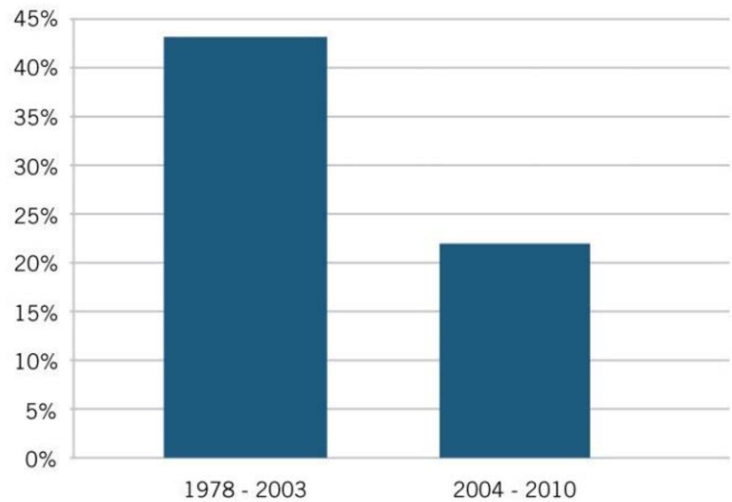
Barret projects that "In about twenty years, a thousand dollars will buy a computer a million times more powerful than one today, and in twenty-five years a billion times more powerful than today."[140] He breathlessly claims that because of that "Most intelligence will be computer-based, and trillions of times more powerful than today."[141] He goes on to state: "What the Law of Accelerating Returns means is that the protections and advances we're discussing in this book are hurtling toward us like a freight train that doubles its speed every mile, then doubles it again."[142] Likewise, McAfee and Brynjolfsson write that we are "reaching the second half of the chessboard," where exponential gains in computing power lead to drastic changes after an initial gestation period.[143] Ray Kurzweil writes that "an analysis of the history of technology shows that technological change is exponential, contrary to the common-sense 'intuitive linear' view. So we won't experience 100 years of progress in the 21st century—it will be more like 20,000 years of progress (at today's rate)."[144] This exponentialist view is so widely repeated that almost no one recognizes its absurdity. If innovation is improving exponentially every few years, this would suggest that a decade from now, the U.S. Patent and Trademark Office (PTO) should be issuing 4.4 million patents a year, up from the 542,000 it issued in 2013 (the exponential rate of increase). Likewise, exponential innovation would mean that economic growth rates should be increasing exponentially. In fact, they are stagnant or falling.

This belief in the continued inevitability of Moore's law has become almost religious. Barret writes, "We know Moore's law and the law of accelerating returns are economic rather than deterministic laws."[145] In other words, as long as companies have an incentive to get faster processors, there is no need to worry about the constraints of physics: "They shall overcome." Unfortunately, physics is real and can't be wished away, no matter how much companies would like to.

The reality is that Moore's law has slowed by half over the last 12 years compared to the prior three decades, hardly evidence of our exponential acceleration.[146] (See Figure 3.) Moreover, for the first time since their invention, transistor costs are increasing.

The pace of hardware-based IT advancement could slow down even more. Silicon-based IT systems are likely nearing their limits—even Gordon Moore said Moore's law is dead.[147] Intel recently announced that it was moving away from its past development process and that this shift will "lengthen the amount of time [available to] utilize ... process technologies."[148]

Figure 3: Change in Computer Processer Speed, Annual Rate of Change[149]



Whenever there is a period where more than the normal number of innovations is emerging, people think innovation is accelerating.

And possibly as soon as 2020, the dominant silicon-based CMOS semiconductor architecture will hit physical limits (particularly pertaining to heat dissipation) that threaten to compromise Moore's law unless a leap can be made to radically new chip architectures. That's not to say that at some point a radically different technology will not replace the current silicon-based IT system, perhaps quantum computing.[150] But it is unlikely that this replacement system will be ready for commercialization just as the miniaturization constraints of silicon reach their limits. This means that there is likely to be an intervening period of a least a couple of decades of slow hardware-driven innovation and slow growth until the next computing system fully emerges.[151]

Still, some argue that because of AI the pace of innovation will continue to be exponential, even if semiconductor progress stalls. Computer scientist Jeremy Howard states that capabilities of machine learning grow exponentially.[152] Calum Chase agrees that "AI is a powerful tool, and it is growing more powerful at an exponential rate."[153] The better computers get at intellectual activities, he asserts, the more they can better build better computers to have better capabilities. But while machine-learning systems get better at their particular task the more practice they have (e.g., identifying images of cars), there is no evidence that the machine-learning process is improving exponentially, much less that computer programs are improving the performance of computers exponentially. In fact, while one branch of machine learning and AI—deep learning around classification systems—has grown rapidly in the last few years, other areas, such as deductive reasoning and logic, have progressed much more slowly. This is why AI scientists Dietterich and Horwitz write, "we have made surprisingly little progress to date on building the kinds of general intelligence that experts and the lay public envision when they think about 'Artificial Intelligence.'"[154] Gary Marcus agrees:

> Learning to detect a cat in full frontal position after 10 million frames drawn from Internet videos is a long way from understanding what a cat is, and anybody who thinks that we've 'solved' AI doesn't realize the limitations of the current technology. To be sure, there have been exponential advances in narrow-

> engineering applications of artificial intelligence, such as playing chess, calculating travel routes, or translating texts in rough fashion, but there's been scarcely more than linear progress in five decades of working toward strong AI.[155]

All this talk of accelerated pace of innovation is not new. Whenever there is a period where more than the normal number of innovations is emerging, people think innovation is accelerating. In 1956, the father of computer science, John von Neumann, wrote that "technological evolution is still accelerating."[156] It wasn't then; it isn't now.

AI Is Task-specific, Not General, Intelligence

The second reason not to worry about the AI apocalypse is that is that software and the mind are completely different systems and even major advances in computing are unlikely to produce the latter. As Zarkadakis writes, "'Superhuman intelligence' is not semantically equivalent to 'a computer possessing the whole spectrum of cognitive capabilities that a human brain has.' Computers supersede us only in specific subsets of intelligence. Brute computing power does not suffice for computers to achieve the whole spectrum of the human brain's cognitive abilities."[157] He goes on to write:

There is in fact a fundamental difference between information processing and thinking.

> Most ordinarily people and non-science journalists still think of AI as computers becoming as intelligent as humans. But this is not what actually takes place in modern AI labs. What researchers there try to do is to produce software and hardware that would work together in such a way for a computer to be able to perform human-like tasks better, more efficiently, in a manner less error-prone and a lot more quickly. For this to happen machine self-awareness is not a prerequisite.[158]

Or another way to put is to state, as Rodney Brooks puts it, "We generalize from performance to competence and grossly overestimate the capabilities of machines—those of today and of the next few decades."[159] In other words, when we talk to Siri on our phones, we think she's pretty darn smart and Skynet is only a few upgrades away.

There is in fact a fundamental difference between information processing and thinking. As Seth Lloyd, MIT professor of quantum mechanical engineering, writes, "deep learning is informationally broad—it analyzes vast amounts of data—but conceptually shallow."[160] Or as Brooks writes, "Today's algorithm has nothing like human-level competence in understanding images."[161] Domingos agrees, writing, despite all their success, today's machine learning is "still a far cry from the brain."[162] He goes on to write, "Unfortunately, what we have so far is only a very crude cartoon of how nature learns, good enough for a lot of applications but still a pale shadow of the real thing."[163] The path from this kind of "learning" (it is an anthropomorphism even to call it learning) to what "human-intelligent" agents do is completely unclear.[164] Domingos writes, "The *Terminator* scenario, where a super-AI becomes sentient and subdues mankind with a robot army, has no chance of coming to pass with the kinds of learning algorithms we'll meet in this book."[165] He continues, "Narrowly defined tasks are easily learned from data, but tasks that require a broad combination of skills and knowledge aren't. … Common sense is important not just because your mom told you so, but because computers don't have it."[166] Similarly IT-focused venture capitalist Dan Gordon writes:

> The core problem is that the leap between today's 'intelligent' software and a superintelligence is unknown, and our temptation is to mystify it. Whether we are building brain emulations or pure AIs, we don't understand what would make them 'come to life' as intelligent beings, let alone super intelligent. … Machine learning software today uses a statistical model of a subject area to 'master' it. Mastering consists of changing the weights of the various elements in the model in response to a set of training instances (situations where human trainers grade the instances: 'yes, this is credit card fraud,' 'no, this is not a valid English sentence,' etc.). Clear enough, but it just doesn't seem very much like what our minds do.[167]

Or as Bradford writes, "Originality—the really hard part of being smart, and utterly not understood, even in humans—is, so far, utterly undemonstrated in AIs."[168] This is why Zarkadakis writes:

> This leads us to an inevitable conclusion: that if we want to engineer a conscious machine we have already reached the limits of conventional computer technology. Symbolic representation can only take us so far. It does not matter how many quintillion calculations per second computers will be capable of performing by the next decade. Computer technologies that are based on separating hardware from software and which use symbolic logic to represent the world may become intelligent enough to replace many knowledge-based jobs, but they will never become conscious. They will therefore not threaten the survival of the human race.[169]

In other words, virtually all AI research today is focused on task-specific AI, what some call narrow AI, such as can a machine find cancer in cells? These are built to perform a set of tasks extremely well and to be sure, if they work, they can do the particular narrow task in a superhuman way from day one. But this does not mean the machine's intelligence is broad and general the way human intelligence is. This is why Zarkadakis notes:

> Both fundamental assumptions for the AI Singularity appear to be highly problematic. For AI to take over the world it must first become self-aware—or 'awake,' to use Vinge's own term.[170] Nothing in the current technology points even remotely towards such an eventuality. Computers may be becoming increasingly more powerful in terms of calculations per second,[171] and able to perform tasks demanding increasingly intricate levels of knowledge, but they are still a long way away from doing what a human baby can do without even thinking. When was the last time you saw a computer giggle at a funny face?[172]

But these limitations do not deter AGI true believers. For many believe that even if conventional narrow AI can't gain AI capabilities, we will build machines that will mimic the brain. Unfortunately, calling a program running on silicon a "neural network" does not make it like the human brain. These AGI believers vastly underestimate the complexity of the human brain, or any animal brain for that matter, while overestimating human progress in neural science. In 2012 scientists were still struggling to model the activity of a quarter of a cubic millimeter of a mouse brain, and scientists are only beginning to understand how the brain works.[173] As Zarkadakis, writes "The human brain is the most complex object in the known universe. It is made up of approximately one hundred billion cells called 'neurons,' which connect to one another by means of nearly one hundred billion connections. Apart from being incredibly complex, the brain is also deeply mysterious: it 'thinks'."[174] Indeed, as V.S. Ramachandran explains in *The Tell-Tale Brain*, neuroscience is at an extremely early stage of development today, where we don't even know what we don't know.[175] This has a direct bearing on the arguments of the ASI or even AGI proponents. If we are so ignorant of the workings of the brain, how could we possibly think we can recreate the brain and mind in a machine? As Rodney Brooks sums it up, "I think it is a mistake to be worrying about us developing {strong} AI any time in the next few hundred years. I think the worry stems from a fundamental error in not distinguishing the difference between the very real recent advances in a particular aspect of AI, and the enormity and complexity of building sentient volitional intelligence."[176]

Similarly, Gary Marcus, a psychology professor at New York University, writes:

Calling a program running on silicon a "neural network" does not make it like the human brain.

> Deep learning is only part of the larger challenge of building intelligent machines. Such techniques (are) still a long way from integrating abstract knowledge, such as information about what objects are, what they are for, and how they are typically used. The most powerful A.I. systems, like Watson … use techniques like deep learning as just one element in a very complicated ensemble of techniques.[177]

Indeed, it is very easy to attribute much more capability to these systems than is actually there. As Barrat writes about AI Scientist Ben Goertzel's company OpenCog, its "organizing theme is that intelligence is based on high-level pattern recognition. Usually 'patterns' in AI are chunks of data (files, pictures, text, objects) that have been classified—organized by category—or will be classified by a system that has been trained on data."[178] This is far cry from "intelligence," even simple intelligence. Even Calum Chase, an AI mythmaker, acknowledges that to mimic the human brain, the level of granularity of how the brain works, "might be impossible—at least for several centuries."[179] Plus on top of the scientific difficulties, there is a physical difficulty. For as Krauss writes, given current power consumption by electronic computers, a computer with storage and processing capability of the human mind would require more than 10 terrawatts of power, within a factor of 2 of the current power consumption of all humanity.[180]

A final claim that needs rebutting is that even if these systems are not truly intelligent now, they can easily become so because they are based on "learning algorithms." But this is to confuse what learning means. It does not mean learning to become smarter, as in doubling one's IQ. Rather it means learning more about something and being more capable, as a person knows more about history after taking a history course. Moreover, exactly how AI becomes AGI is never specified, other than through some kind of recursive learning system. But as Dietterich and Horvitz write:

> [S]everal of these (apocalyptic) speculations envision an 'intelligence chain reaction,' in which an AI system is charged with the task of recursively designing

> progressively more intelligent versions of itself and this produces an 'intelligence explosion.' While formal work has not been undertaken to deeply explore this possibility, such a process runs counter to our current understandings of the limitations that computational complexity places on algorithms for learning and reasoning.[181]

Finally, some will warn that even if they don't become self-aware, that AI systems are inherently dangerous because they will control too much, making us overly dependent on them should things go wrong. As Dietterich and Horvitz write:

> Other scenarios can be imagined in which an autonomous computer system is given access to potentially dangerous resources (for example, devices capable of synthesizing billons of biologically active molecules, major portions of world financial markets, large weapons systems, or generalized task markets). The reliance on any computing systems for control in these areas is fraught with risk, but an autonomous system operating without careful human oversight and failsafe mechanisms could be especially dangerous. Such a system would not need to be particularly intelligent to pose risks.[182]

To be sure, AI systems will be complicated and societies will be more dependent on them, but that is really saying nothing unless one advocates returning to a hunter-gatherer society. The very nature of civilization is a long and steady march of complexity, so that any one person understands very little of it. As musician Brian Eno writes about the things he did one morning at his cottage in the English countryside:

> And here's what I won't understand about all this. I won't understand how the oil that drives my central heating got from a distant oil field to my house. I won't know how it was refined into heating oil or what commercial transactions were involved. I won't know how the burner works. I won't know where my porridge or tea or nuts came from or how they got to me. I won't know how my phone works, or how my digital radio works, or how the news it relays to me was gathered or edited. I also won't understand the complexities of organizing a bus or train service, and I couldn't repair any of the vehicles involved. I won't really understand how beds are mass-produced, or how Wi-Fi works, or exactly what happens when I press 'send' on my e-mail or transfer money electronically… Now here's the funny thing. I won't be in the least troubled by my vast ignorance about almost everything I'll be doing this morning. … My untroubled attitude results from my almost absolute faith in the reliability of the vast supercomputer I'm permanently plugged into. It was built with the intelligence of thousands of generations of human minds, and they're still working at it now. All that human intelligence remains alive, in the form of the supercomputer of tools, theories, technologies, crafts, sciences, disciplines, customs, rituals, rules of thumb, arts, systems of belief, superstitions, work-arounds, and observations that we call Global Civilization.[183]

Control Systems

Even if AGI systems could be built, which is unlikely, they will remain under the control of humans. They won't transform into ASI and "get loose" to go out and build an army of killer robots to exterminate us. But AI dystopians will have none of this. Even AGI is too smart for us, they argue, and will find ways to rewrite its code to escape control. Others go even farther. Calum Chase writes:

> Even without escaping its cage, an oracle AI could cause unacceptable damage if so inclined, by perpetrating the sort of mind crimes we mentioned in the last chapter. It could stimulate conscious minds inside its own mind and use them as hostages, threatening to inflict unspeakable tortures on them unless it is released. Given sufficient processing capacity, it might create millions or even billions of these hostages.

That this is all nothing more than philosophical cogitation suggests that it doesn't even deserve an answer. But here's one: Okay, let the big AI torture the other AIs it made. Who cares?[184] Oxford University mathematician Marcus du Sautoy writes that artificial intelligence should enjoy human rights.[185] I guess that means that if AIs torture other AIs they should be put on trial, and if convicted they should be sent to AI prison.

University of Louisville researcher Roman V. Yampolskiy goes even farther out on the techno-panic scale, warning that:

> A malevolent superintelligence may attempt to abuse and torture humankind with perfect insight into our physiology to maximise amount of physical or emotional pain, perhaps combining it with a simulated model of us to make the process infinitely long.[186]

The reality is that there is considerable reason to believe that any manmade system will continue to be controlled by humans. As David Mindell writes:

> Robots are imagined (and sold) as fully autonomous agents, even when today's modest autonomy is shot through with human imagination. … Whatever they might do in a laboratory, as robots moved closer to environments with human lives and real resources at stake, we tend to add more human approvals and interventions to govern their autonomy. My argument here is not that machines are not intelligent, nor that someday they might be. Rather, my argument is that such machines are not inhuman.[187]

Mindell is saying that machines built by humans will reflect human values. He goes on to write:

> Finally, we have the myth of full autonomy, the utopian idea that robots, today or in the future, can operate entirely on their own. Yes, automation can certainly take on parts of tasks previously accomplished by humans, and machines do act on their own in response to their environments for certain periods of time. But the machine that operates entirely independently of human direction is a useless

> machine. The constraints on those behaviors are still very tight, and very much pre-scripted by the designers and programmers. But we should not confuse technical thought experiments with what's useful in human context. This is because it is true that an autonomous system might use software that is nondeterministic (i.e., unpredictable), or might employ emergent properties driven by its environment or engage in learning behaviors. Yet any supposedly intelligent system was programmed by people and embeds their world views into the machine. For the twenty-first century, then, autonomy is human action removed in time.[188]

"A robot whose programmed goal is 'make a good dinner' may decide to cook a steak, a bouillabaisse, or even a delicious new dish of its own creation, but it can't decide to murder its owner any more than a car can decide to fly away."
—Pedro Domingos

In other words, designers of autonomous vehicles will not build vehicles that can decide on their own to go 200 mph and then crash. Likewise, as Domingos writes, "if computers start to program themselves, how will we control them? Turns out we can control them quite well."[189] He goes on to state that machine-learning systems:

> learn to achieve the goals we set them; they don't get to change the goals. Relax. The chances that an AI equipped with the Master Algorithm will take over the world are zero. The reason is simple: unlike humans, computers don't have a will of their own. They're products of engineering, not evolution. Even an infinitely powerful computer would still be only an extension of our will and nothing to fear.[190]

He says it best when he writes, "A robot whose programmed goal is 'make a good dinner' may decide to cook a steak, a bouillabaisse, or even a delicious new dish of its own creation, but it can't decide to murder its owner any more than a car can decide to fly away."[191] As Bendford writes, "No adventurous algorithm will escape the steely glare of its many skeptical inspectors. Any AI that has abilities in the physical world where we actually live will get a lot of inspection."[192] Or as Thomas Dietterich writes, "whereas such scenarios make for great science fiction, in practice it's easy to limit the resources a new system is permitted to use. Engineers do every day when they test new devices and new algorithms."[193] Finally, Schank writes, "There's nothing we can produce that anyone should be frightened of. If we could actually build a mobile intelligent machine that could walk, talk and chew gum, first users of that machine would certainly not be able to take over the world or form a new society of robots. A much simpler use would be as a household robot."[194]

This is not to say that as AI research goes forward that one component should not involve research on control systems. Indeed, as Dietterich and Horvitz write, "Deeper study is needed to understand the potential of superintelligence or other pathways to result in even temporary losses of control of AI systems. If we find there is significant risk, then we must work to develop and adopt safety practices that neutralize or minimize that risk"[195] But the key point is that it is way too early to tell if such risks will emerge, and there is plenty of time to work on systems as they develop. It's a bit like worrying now how to land a spaceship on a plane in another solar system. You worry about that when you develop interstellar transport. As noted AI expert Ben Goertzel writes, "We're only going to find out how to make ethical AI systems by building them, not concluding from afar that they're bound to be dangerous. … The theory of how to make ethical AGI is going to come about through experimenting with AGI systems."[196] Moreover, Yudkowsky writes, "Current AI algorithms aren't smart enough to exhibit most of the difficulties we can foresee for sufficiently advanced agents—meaning there's no way to test proposed solutions to those difficulties."[197]

The "Colt 45" Solution and AGI

But even if we can control AI, some argue that AGI will emerge and build its own machines that we can't control. Barrat claims that for AGI, "Modifying its own hardware is within the system's capability."[198] R. L Adams writes, "And if it were truly autonomous, it could improve upon its design, engineer stealthy weapons, infiltrate impenetrable systems, and act in accordance to its own survival."[199]

It is way too early to tell if such risks will come about, and there is plenty of time to work on systems as they develop. It's a bit like worrying now how to land a spaceship on a plane in another solar system.

But for this to happen, the AGI would have to control robots and robotic factories, which, if humans do not want to happen, would not happen. Besides if we really don't like an AI system and feel that its operation is a threat, there is an easy solution: Unplug it from the electricity grid, or even better, put a bullet through its hard drive.

But true believers in the singularity and emergence of AGI and ASI don't believe this is workable, and claim that these systems will not be able to be shut off. Chase writes, "The first AGI is likely to be a development of a large existing system which we depend on too much to allow switching off to be a simple proposition. Where is the off switch for the Internet?" But this ignores the fact there is no "Internet." The Internet is not a thing, a box, or a system. It is a set of networking protocols that lets machines send packets to it each; no more or no less.[200] Any AGI system will be housed in discrete physical units that can be turned off or destroyed. Benford sums it up best, "Any AI with ambitions to Take Over the World (the theme of many bad sci-fi movies) will find itself confronting an agile, angry, smart species on its own territory, the real material world, not the computational abstractions of 0s and 1s. My bet is on the animal nature."[201]

WHAT TO DO AND WHAT NOT TO DO

Making sure that societies receive the full economic and social benefits that AI has to offer first and foremost requires accelerating, rather than restricting the technology's development and adoption. And that in turn requires that policymakers resist an AI techno-panic; they must instead embrace future possibilities with optimism and hope.

When it comes to AI, policymakers should rely on the innovation principle, not the precautionary principle. In other words, we should proceed on the assumption that AI will be fundamentally good, and while it will present some risks, as every technology does, we should focus on addressing these risks, rather than slowing or stopping the technology. Living by the innovation principle also means understanding that AI will involve both Type I and Type II errors, which is to say it will produce some errors, but it will also eliminate many others. Focusing only on the potential errors AI might make overlooks the vast number of errors it will solve or avoid. Banning a technology to eliminate errors would likely be very costly. As Zarkadakis eloquently states, "Despite the fact that complex software and hardware already take autonomous decisions that may have adverse effects on a global scale, there should be no desire to halt progress—as long as we understand the risks. Every day, the information age delivers value across every sector of our society. The digitization of just about everything creates new opportunities for wealth and for finding fresh ways to solve problems across the whole spectrum of the human condition."[202]

To be sure, we should not overlook the challenges as we pursue the opportunities AI offers. Indeed, we need more discussion about the challenges, not less, if only because many, if not most, of the voices dominating the AI debate are spreading techno-panic. Governments should look at AI rationally and calmly. For example, the White House recently announced a series of workshops to explore the social issues around AI.[203] It will be important that they populate these panels with individuals who are generally pro-innovation, even if they reflect a diversity of views about the risks of AI. Likewise, the UK Parliament has launched an inquiry into robotics and AI to examine a range of issues, including the implications for the labor market, and other social, legal, and ethical issues.[204]

When it comes to AI, policymakers should rely on the innovation principle, not the precautionary principle. We should proceed on the assumption that AI will be fundamentally good, and while it will present some risks, as every technology does, we should focus on addressing these risks, rather than slowing or stopping the technology.

As IT researcher David Moschella writes, society will be more willing to experiment with machine intelligence in low-risk areas, such as navigating, reporting, entertaining, and learning. But over time—and more quickly for some nations than others—society will accept it for medium-risk areas such as investing, diagnosing, essay grading, and lawyering. The high-risk areas, such as driving, surgery, and killing in times of war, may be the last to gain acceptance. As Moschella writes, "Looking ahead, levels of assurance and societal acceptance will likely vary widely, both within each column and especially among countries."[205]

One group that should start focusing on the enormous benefit AI can produce, instead of possible harms, is the AI profession itself. Indeed, as a profession, AI researchers do a poor job of explaining why government and society should support AI. In fact, researchers sometimes appear to go out of their way to convince people they shouldn't. The civil engineering association doesn't have articles on its website about why bridges could collapse, nor does the mechanical engineers' website feature articles about how robots will take jobs. But one finds these kinds of statements and links on the website of the Association for the Advancement of Artificial Intelligence.[206] As noted above, some of the loudest voices for how AI will destroy jobs come from the AI community itself, despite the fact that they are not economists and have no insight or expertise into labor market dynamics.

Finally, policymakers need to do more to support research on AI development, including on making AI safer, more secure, and more transparent. Given the future importance of AI, a big reason government should increase funding for AI research is that such research funding will play an important role in fostering the next generation of scientists and engineers with the skills needed to advance AI. Policymakers also should support companies and government agencies in using AI to better accomplish their tasks.

In short, technological progress has been and will remain key for future progress, and AI is poised to play a key role in that progress, provided we do not give in to fear.

ENDNOTES

1. "AI Overview: Broad Discussions of Artificial Intelligence," AITopics; accessed May 31, 2016, http://aitopics.org/topic/ai-overview.
2. Guruduth S. Banavar, "The Future of Expertise" (remarks delivered at the Nobel Week Dialogue in Gothenburg, Sweden, December 9, 2015), https://youtu.be/ESPA8CTFOSE.
3. Herbert A. Simon, *The Shape of Automation for Men and Management* (New York: Harper & Row, 1965).
4. Calum Chase, *Surviving AI: The Promise and Peril of Artificial Intelligence* (Three C's, 2015), 104.
5. John Markoff, *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots* (New York: Ecco, 2015), 114.
6. Elizabeth Dwoskin, "Siri's Creators Say They've Made Something Better That Will Take Care of Everything for You," *The Washington Post*, May 4, 2016, https://www.washingtonpost.com/news/the-switch/wp/2016/05/04/siris-creators-say-theyve-made-something-better-that-will-take-care-of-everything-for-you/.
7. Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (New York: Basic Books, 2015).
8. Amit Karp, "Deep Learning Will Be Huge—and Here's Who Will Dominate It," *Venture Beat*, April 2, 2016, http://venturebeat.com/2016/04/02/deep-learning-will-be-huge-and-heres-who-will-dominate-it/.
9. Adam Thierer, "Technopanics, Threat Inflation, and the Danger of an Information Technology Precautionary Principle," *Mercatus Center*, January 25, 2013, http://mercatus.org/publication/technopanics-threat-inflation-and-danger-information-technology-precautionary-principle, originally published in *The Minnesota Journal of Law, Science and* Technology 14, no. 1 (2014).
10. Markoff, *Machines of Loving Grace*, 157.
11. Matt Peckham, "What 7 of the World's Smartest People Think About Artificial Intelligence," *Time*, May 5, 2016, http://time.com/4278790/smart-people-ai/.
12. Alex Hern, Leah Green and Bruno Rinvolucri, "We Should Be More Afraid of Computers Than We Are," *The Guardian*, 02:21. Posted March 16, 2016, http://www.theguardian.com/commentisfree/video/2016/mar/16/artificial-intelligence-we-should-be-more-afraid-of-computers-than-we-are-video.
13. George Zarkadakis, *In Our Own Image: Savior or Destroyer? The History and Future of Artificial Intelligence* (New York: Pegasus, 2016), 64.
14. Ibid, 309.
15. Zeynep Tufekci, "The Real Bias Built in at Facebook," *The New York Times*, May 19, 2016, http://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html?smprod=nytcore-iphone&smid=nytcore-iphone-share&_r=0.
16. Daniel Castro and Alan McQuinn, "The Privacy Panic Cycle: A Guide to Public Fears About New Technologies" (Information Technology and Innovation Foundation, September 10, 2015), https://itif.org/publications/2015/09/10/privacy-panic-cycle-guide-public-fears-about-new-technologies.
17. James Barrat, *Artificial Intelligence and the End of the Human Era: Our Final Invention* (New York: Thomas Dunne Books, 2013), 223.
18. Ibid.
19. Zarkadakis, *In Our Own Image*, 312.
20. Ibid, 309.
21. Chase, *Surviving AI*, 85.
22. Paul Saffo, "What Will Be the Place of Humans in a World Occupied by an Exponentially Growing Population of Autonomous Machines?" in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 206.
23. Ibid, 72.

24. Susan R. Fussell et al., "How People Anthropomorphize Robots" (working paper, Computer Interaction Institute, Carnegie Mellon University, Pittsburg, PA, March 2008), http://sfussell.hci.cornell.edu/pubs/Manuscripts/Fussell-HRI08.pdf.

25. Gregory S. McNeal, "MIT Researchers Discover Whether We Feel Empathy for Robots," *Forbes*, April 10, 2015, http://www.forbes.com/sites/gregorymcneal/2015/04/10/want-people-to-like-your-robot-name-it-frank-give-it-a-story/#6da5470f71d5.

26. Eric J. Topol, "A New Wisdom of the Body," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 216.

27. David Sarnoff, "The Fabulous Future" in *The Fabulous Future: America in 1980*, David Sarnoff (New York: E.F. Dutton), 36.

28. Kahn Herman and Anthony J. Wiener, *The Year Two Thousand* (London: Collier Macmillan Ltd., 1968), 93.

29. Barrat, *Artificial Intelligence End Human Era*, 17.

30. See forthcoming Information Technology and Innovation Foundation study that has reviewed media coverage of technological innovation over the past 4o years.

31. Daniel C. Dennett, "The Singularity—An Urban Legend?" in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 85.

32. Roger Schank, "Hawking Is Afraid of AI Without Having a Clue What It Is: Don't Worry Steve," *Roger Schank*, December 8, 2014, http://www.rogerschank.com/hawking-is-afraid-of-ai-without-having-a-clue-what-ai-is.

33. Robert Friedel, *A Culture of Improvement: Technology and the Western Millennium* (Cambridge, MA: The MIT Press, 2010), 541.

34. Given that real GDP per capita in Great Britain has quadrupled since then, much of it because of automation, I doubt that many people in 1956 would have chosen not to have UK companies embrace automation given the great benefit it generated for their children.

35. Norbet Weiner, *Cybernetics: or the Control and Communication in the Animal and the Machine* (Cambridge, MA: The MIT Press, 1961).

36. Friedel, *A Culture of Improvement*, 542.

37. John von Neumann, "Can We Survive Technology" in *The Fabulous Future: America in 1980*, David Sarnoff, (New York: E.F. Dutton), 46.

38. Robert Provine, "Keeping Them On a Leash," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 148.

39. Roger Schank, "Machines That Think Are in Movies," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 135.

40. Thomas Dietterich and Eric J. Horvitz, "Rise of Concerns About AI: Reflections and Directions," *Viewpoints*: *Communications of the ACM 58*, no. 10 (2015), http://research.microsoft.com/en-us/um/people/horvitz/CACM_Oct_2015-VP.pdf.

41. Ibid.

42. Ibid.

43. Jerry Kaplan, *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence* (New Haven, CT: Yale University Press, 2015), 203.

44. Alan Yuhas, "Would You Bet Against Sex Robots? AI 'Could Leave Half of World Unemployed,'" *The Guardian*, February 13, 2016, http://www.theguardian.com/technology/2016/feb/13/artificial-intelligence-ai-unemployment-jobs-moshe-vardi.

45. Stuart W. Elliott, "Projecting the Impact of Computers On Work in 2030" (presentation paper, National Research Council, 2007).

46. Eric Brynjolfsson and Andrew McAfee, *Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy* (Lexington, MA: Digital Frontier Press, October 17, 2011).

47. Wasily Leontief and Faye Duchin, "The Impacts of Automation on Employment, 1963-2000" (New York Institute for Economic Analysis, April 1984), http://eric.ed.gov/?id=ED241743.

48. Daniel Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence* (New York: Basic Books, 1994), 324.

49. Stanley Fischer, "Reflections on Macroeconomics Then and Now" (speech, Policy Challenges in an Interconnected World 32nd Annual National Association for Business Economics Economic Policy Conference, Washington, DC, March 7, 2016), http://www.federalreserve.gov/newsevents/speech/fischer20160307a.htm.

50. Michael Chu, James Manyika, and Mehdi Miremadi, "Four Fundamentals of Workplace Automation," *McKinsey Quarterly*, November, 2015.

51. Don Tapscott, "After 20 Years, It's Harder to Ignore the Digital Economy's Dark Side," *Harvard Business Review*, March 11, 2016, https://hbr.org/2016/03/after-20-years-its-harder-to-ignore-the-digital-economys-dark-side.

52. Domingos, *The Master Algorithm*, 277.

53. Carl Frey Benedikt and Michael A. Osbourne, "The Future of Employment: How Susceptible Are Jobs to Computerisation?" (paper, Oxford Martin School, University of Oxford, Oxford, September 17, 2013), http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.

54. Citi GPS: Global Perspectives and Solutions, "Technology at Work v2.0: The Future Is Not What It Used to Be" (Oxford: Oxford Martin School, University of Oxford, January 2016), 60, http://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work_2.pdf.

55. Ibid, 10.

56. Nick Bostrom, narrated by Napoleon Ryan, *Superintelligence: Paths, Dangers, Strategies* (Audible Studios on Brilliance Audio, May 5, 2015).

57. James Manyika et al., "Growth and Renewal in the United States: Retooling America's Economic Engine" (McKinsey Global Institute, February 2011), http://www.mckinsey.com/global-themes/americas/growth-and-renewal-in-the-us.

58. Robert D. Atkinson, "Think Like an Enterprise: Why Nations Need Comprehensive Productivity Strategies" (Information Technology and Innovation Foundation, May 2016), https://itif.org/publications/2016/05/04/think-enterprise-why-nations-need-comprehensive-productivity-strategies.

59. Ian Stewart, Debapratim De, and Alex Cole, "Technology and People: The Great Job-Creating Machine" (New York: Deloitte, 2015), http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-technology-and-people.pdf.

60. Bharat Trehan, "Productivity Shocks and the Unemployment Rate," *Federal Reserve Bank of San Francisco Economic Review*, (2003), http://www.frbsf.org/economic-research/files/article2.pdf.

61. Organisation for Economic Co-operation and Development (OECD), "Technology, Productivity and Job Creation: Best Policy Practices" (Paris: OECD, 1998), 9, http://www.oecd.org/dataoecd/39/28/2759012.pdf.

62. International Labour Organization, *World Employment Report 2004-05: Employment, Productivity and Poverty Reduction* (Geneva: ILO, 2005).

63. Bart van Ark, Ewout Frankema, and Hedwig Duteweerd, "Productivity and Employment Growth: An Empirical Review of Long and Medium Run Evidence" (working paper, Groningen Growth and Development Centre, May 2004).

64. Jianmin Tang, "Employment and Productivity: Exploring the Trade-off," *Industry Canada* no. 28 (2015), http://www.csls.ca/ipm/28/tang.pdf.

65. Anders Isaksson, Thiam Hee Ng, and Ghislain Robyn, *Productivity in Developing Countries: Trends and Policies* (Vienna: UNIDO, 2005), 139, https://www.unido.org/uploads/tx_templavoila/Productivity_in_developing_countries_trends_and_policies.pdf.

66. W. Brain Arthur, "The Second Economy," *McKinsey Quarterly*, October 2011, http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/the-second-economy.

67. Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York: Penguin Books, 2006).

68. Ibid.

69. Jeremy Howard, "The Wonderful and Terrifying Implications of Computers That Can Learn" (talk, TED, Brussels, December 2014), https://www.ted.com/talks/jeremy_howard_the_wonderful_and_terrifying_implications_of_computers_that_can_learn?language=en, accessed March 8, 2016.

70. Markoff, *Machines of Loving Grace*, 51.

71. Ibid, 18.

72. Ibid, 16.

73. Markoff, *Machines of Loving Grace*, 329.

74. Zeynep Tufekci, "Failing the Third Machine Age: When Robots Come for Grandma," *Medium*, July 22, 2014, https://medium.com/message/failing-the-third-machine-age-1883e647ba74#.uw5ppfvsp.

75. Robert D. Atkinson, "The Progressive Case for Productivity Growth: How a pro-productivity agenda can raise wages, lower inequality and sustain the middle class," *Washington Monthly*, November 29, 2015, http://www.washingtonmonthly.com/republic3-0/2015/11/the_progressive_case_for_produc058807.php.

76. Quoted in David Rotman, "Who Will Own the Robots," *MIT Technology Review*, June 16, 2015, https://www.technologyreview.com/s/538401/who-will-own-the-robots/.

77. Terrance J. Sejnowski, "AI Will Make You Smarter," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 113.

78. Ibid.

79. Nicholas Carr, *The Glass Cage: Automation and Us* (New York: W. W. Norton & Company, 2014), 79.

80. Crevier, *AI: Tumultuous History Artificial Intelligence*, 331.

81. Ibid, 332.

82. Carr, *The Glass Cage*, 91.

83. Robert D. Atkinson, "The Coming Transportation Revolution" (The Milken Institute, Fourth Quarter, 2014), http://assets1c.milkeninstitute.org/assets/Publication/MIReview/PDF/78-87-MR64.pdf.

84. Carr, *The Glass Cage*, 72.

85. Ibid, 70.

86. Ibid, 114.

87. Liz Hamel et al., "Primary Care Practitioners' Perspectives on Delivery System Changes," *JAMA*, November 24, 2015, http://jama.jamanetwork.com/article.aspx?articleid=2471558.

88. Alexandra Ossola, "Artificial Intelligence Helps Diagnose Cancer," *Popular Science*, April 18, 2016, http://www.popsci.com/artificial-intelligence-helps-diagnose-cancer.

89. Gerd Gigerenzer, "Robodoctors," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 317.

90. Richard H. Thaler, "Who's Afraid of Artificial Intelligence," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 487.

91. Peter Norvig, "Design Machines to Deal With the World's Complexity," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 177.

92. Markoff, *Machines of Loving Grace,* 342.

93. Jerry Kaplan, *Humans Need Not Apply*, 206.

94. David A. Mindell, *Our Robots, Ourselves: Robotics and the Myths of Autonomy* (New York: Viking, 2015), 77.

95. Carr, *The Glass Cage*, 208.

96. Crevier, *AI: Tumultuous History Artificial Intelligence*.
97. Carly Weeks, "Dear Valued customer, thank you for giving us all of your personal data," *The Globe and Mail,* June 8, 2011, http://www.theglobeandmail.com/technology/dear-valued-customer-thank-you-for-giving-us-all-your-personal-data/article582786/.
98. Margi Murphy, "Speech recognition 'cold war' spells privacy concerns," *TechWorld*, August 19, 2015, http://www.techworld.com/news/big-data/deep-learning-ai-cold-war-spells-privacy-concerns-3623347/.
99. Markoff, *Machines of Loving Grace*, XVI.
100. Ryan Calo, "Peeping HALs: Making Sense of Artificial Intelligence and Privacy," *European Journal of Legal Studies,* 2010, http://www.ejls.eu/6/83UK.htm.
101. Carly Weeks, "Dear Valued Customer, Thank You for Giving Us All of Your Personal Data," *The Globe and Mail*, June 8, 2011, last updated August 24, 2012, http://www.theglobeandmail.com/technology/dear-valued-customer-thank-you-for-giving-us-all-your-personal-data/article582786/.
102. Castro and McQuinn, "The Privacy Panic Cycle."
103. Domingos, *The Master Algorithm,* 275.
104. Daniel Castro, "Europe Should Follow UK's Lead on Corporate Transparency" (Center for Data Innovation, May 12, 2016), https://www.datainnovation.org/2016/05/europe-should-follow-uks-lead-on-corporate-transparency/.
105. Daniel Castro and Elaine Ding, "European Policymakers Should Do More to Pursue Data-Driven Medicine" (Center for Data Innovation, May 4, 2016), https://www.datainnovation.org/2016/05/european-policymakers-should-do-more-to-pursue-data-driven-medicine/.
106. Daniel Castro and Jordan Misra, "100 Data Innovations" (The Center for Data Innovation, accessed May 17, 2016), http://www2.datainnovation.org/2014-100-data-innovations.pdf.
107. Domingos, *The Master Algorithm,* 275.
108. Ann Cavoukian and Daniel Castro, "Big Data and Innovation, Setting the Record Straight: De-identification Does Work" (Information Technology and Innovation Foundation, June 2014), http://www2.itif.org/2014-big-data-deidentification.pdf.
109. Tom Simonite, "Microsoft and Google Want to Let Artificial Intelligence Loose on Our Most Private Data," *MIT Technology Review*, April 19, 2016, https://www.technologyreview.com/s/601294/microsoft-and-google-want-to-let-artificial-intelligence-loose-on-our-most-private-data/.
110. Ibid.
111. Tim Hwang and Madeleine Clare Elish, "The Mirage of the Marketplace," *Slate,* August 9, 2015, http://www.slate.com/articles/technology/future_tense/2015/07/uber_s_algorithm_and_the_mirage_of_the_marketplace.html.
112. David Auerbach, "The Code We Can't Control," *Slate*, January 14, 2015, http://www.slate.com/articles/technology/bitwise/2015/01/black_box_society_by_frank_pasquale_a_chilling_vision_of_how_big_data_has.html.
113. Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).
114. Hannah Devlin, "Killer Robots and Digital Doctors: How Can We Protect Society from AI?" *The Guardian*, March 15, 2016, https://www.theguardian.com/technology/2016/mar/15/killer-robots-driverless-cars-alphago-and-the-social-impact-artificial-intelligence.
115. Paul Mason, "The Racist Hijacking of Microsoft's Chatbot Shows How the Internet Teems with Hate," *The Guardian*, March 29, 2016, http://www.theguardian.com/world/2016/mar/29/microsoft-tay-tweets-antisemitic-racism.
116. Katherine Noyes, "The FTC Is Worried About Algorithmic Transparency, and You Should Be Too," *PC World*, April 9, 2015, http://www.pcworld.com/article/2908372/the-ftc-is-worried-about-algorithmic-transparency-and-you-should-be-too.html.

117. Amber Laxton, “Critics of ‘Sexist Algorithms’ Mistake Symptoms for Illness” (The Center for Data Innovation, August 3, 2015), https://www.datainnovation.org/2015/08/critics-of-sexist-algorithms-mistake-symptoms-for-illness/.

118. Ibid.

119. Lawrence M. Krauss, “What, Me Worry?” in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 172.

120. Mindell, *Our Robots, Ourselves*, 10.

121. Auerbach, “The Code We Can’t Control.”

122. Mark Rotenburg, “Comments of The Electronic Privacy Information Center to The Office of Science and Technology Policy, Request for Information: Big Data and the Future of Privacy” (Electronic Privacy Information Center, April 4, 2014), https://epic.org/privacy/big-data/EPIC-OSTP-Big-Data.pdf.

123. Lauren Smith, “Algorithmic Transparency: Examining from Within and Without,” *IApp*, January 28, 2016, https://iapp.org/news/a/algorithmic-transparency-examining-from-within-and-without/.

124. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), 7.

125. Cornell Belcher and Dee Brown, “Hailing While Black—Navigating the Discriminatory Landscape of Transportation” (key findings from the hailing while black survey of Chicago voters, *Brilliant Corners*, February, 12, 2015), http://www.brilliant-corners.com/post/hailing-while-black; David R. Francis, “Employers’ Replies to Racial Names” (The National Bureau of Economic Research (NBER), accessed May 16, 2016), http://www.nber.org/digest/sep03/w9873.html.

126. Carr, *The Glass Cage*, 191.

127. Ibid, 186.

128. The White House Executive Office of the President, “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights,” May 2016, https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

129. Terrell McSweeny, "Tech for Good: Data for Social Empowerment" (keynote remarks of Commissioner Terrell McSweeny at Google DC Tech Talk Series, Washington District of Columbia, September 10, 2015), https://www.ftc.gov/system/files/documents/public_statements/800981/150909googletechroundtable.pdf.

130. The White House, “Big Data.”

131. Carnegie Mellon Transparency Reports Make AI Decision-Making Accountable, ECN Magazine, May 26, 2016, http://www.ecnmag.com/news/2016/05/carnegie-mellon-transparency-reports-make-ai-decision-making-accountable

132. Christopher Zara, “FTC Chief Technologist Ashkan Soltani On Algorithmic Transparency and the Fight Against Biased Bots,” *International Business Times*, April 9, 2015, http://www.ibtimes.com/ftc-chief-technologist-ashkan-soltani-algorithmic-transparency-fight-against-biased-1876177.

133. Noyes, “FTC Worried About Algorithmic Transparency.”

134. Mindell, *Our Robots, Ourselves*.

135. Jerry Kaplan, *Humans Need Not Apply,* 7.

136. Ibid, 199.

137. Chase, *Surviving AI*, XVII.

138. Rolf Dobelli, “Self-Aware AI? Not in 1,000 Years!” in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 101.

139. Rodney A. Brooks, “Mistaking Performance for Competence,” in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 111.

140. Barrat, *Artificial Intelligence End Human Era*, 130.

141. Ibid, 28.

142. Ibid, 131.

143. Brynjolfsson and McAfee, *Race Against the Machine*.

144. Ray Kurzweil, "The Law of Accelerating Returns," *Kurzweil Accelerating Intelligence*, March 7, 2001, http://www.kurzweilai.net/the-law-of-accelerating-returns.
145. Barrat, *Artificial Intelligence End Human Era.*
146. From 1978 to 2003, processing speeds grew at an annual average of 43 percent, while from 2003 to 2010, processing speeds grew at an annual average of 22 percent; John L. Hennessy and David A. Patterson, *Computer Architecture: A Quantitative Approach 5th Edition* (Burlington, MA: Morgan Kaufmann, September 2011).
147. Manek Dubash, "Moore's Law Is Dead, Says Gordon Moore," *Techworld*, April 13, 2010, http://www.techworld.com/news/operating-systems/moores-law-is-dead-says-gordon-moore-3576581/.
148. Jamie Condliffe, "Moore's Law Stutters at Intel," *Gizmodo*, March 23, 2016, http://gizmodo.com/moores-law-stutters-as-intel-switches-from-2-step-to-3-1766574361?utm_medium=sharefromsite&utm_source=Gizmodo_twitter.
149. Hennessy and Patterson, *Computer Architecture 5th Edition*.
150. Ibid.
151. Robert D. Atkinson, *The Past and Future of America's Economy: Long Waves of Innovation that Power Cycles of Growth* (Cheltenham, UK: Edward Elgar Pub, November 2004).
152. Howard, "Implications Computers That Can Learn."
153. Chase, *Surviving AI*, 41.
154. Dietterich and Horvitz, "Rise of Concerns About AI."
155. Gary Marcus, "Machines Won't Be Thinking Anytime Soon," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 405.
156. John von Neumann, "Can We Survive Technology," in *The Fabulous Future: America in 1980*, ed. Morton Shapiro (Evanston, IL: Northwestern University Press, 2015).
157. George Zarkadakis, *In Our Own Image*," 272.
158. Ibid, 260.
159. Brooks, "Mistaking Performance for Competence," 108.
160. Seth Lloyd, "Shallow Learning," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 116.
161. Brooks, "Mistaking Performance for Competence," 110.
162. Domingos, *The Master Algorithm*, 117.
163. Ibid, 140.
164. P 95, Machine smart Superintelligence: Paths, Danger, Strategies, Issues in Science and Technology, Spring 2015
165. Domingos, *The Master Algorithm*, 45.
166. Ibid, 276.
167. Dan Gordon, "Book Review: Machine Smart," *Issues in Science and Technology 31*, no. 3 (2015), http://issues.org/31-3/book-review-machine-smart/.
168. Gregory Benford, "Fear Not The AI," in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 170.
169. George Zarkadakis, *In Our Own Image*, 279.
170. Vernor Vinge, "The Coming Technological Singularity: How to Survive in the Post-Human Era" (presented at the VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, March 30-31, 1993).
171. At the time this book was written in 2014, the fastest computer in the world was the Chinese Tianhe-2, located at Sun Yatsen University, Guangzhou China. It could perform 33.86 petaflops, or 33.86 quadrillion floating point operations per second.
172. George Zarkadakis, *In Our Own Image*, 274.

173. Michael Chorost, “Let Artificial Intelligence Evolve,” *Slate*, April 18, 2016, http://www.slate.com/articles/technology/future_tense/2016/04/the_philosophical_argument_against_artificial_intelligence_killing_us_all.html.

174. George Zarkadakis, *In Our Own Image*, XIV.

175. V.S. Ramachandran, *The Tell-Tale Brain: A Neuroscientist’s Quest for What Makes Us Human* (New York: W.W. Norton, 2011).

176. Chase, *Surviving AI*, 91.

177. Gary Marcus, “Is “Deep Learning” a Revolution in Artificial Intelligence?” *The New Yorker*, November 25, 2012, http://www.newyorker.com/news/news-desk/is-deep-learning-a-revolution-in-artificial-intelligence.

178. Barrat, *Artificial Intelligence End Human Era*, 165.

179. Chase, *Surviving AI*, 76.

180. Krauss, “What, Me Worry?” 171.

181. Dietterich and Horvitz, “Rise of Concerns About AI.”

182. Ibid.

183. Brian Eno, “Just A New Fractal Detail in the Big Picture,” in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 278.

184. Chase, *Surviving AI*, 159.

185. Peter Dockrill, “Artificial Intelligence Should Be Protected by Human Rights, Says Oxford Mathematician,” *Science Alert,* May 31, 2016, http://www.sciencealert.com/artificial-intelligence-should-be-protected-by-human-rights-says-oxford-mathematician.

186. Sean Martin, “Artificial Intelligence Could ‘Abuse and Torture’ Mankind Shock Report Reveals,” *Express*, May 24, 2016, http://www.express.co.uk/news/science/673380/Artificial-Intelligence-could-abuse-and-torture-mankind-shock-report-reveals

187. Mindell, *Our Robots, Ourselves*, 8.

188. Ibid, 220.

189. Domingos, *The Master Algorithm*, 6.

190. Ibid, 283.

191. Ibid, 276.

192. Benford, “Fear Not The AI,” 168.

193. Thomas G. Dietterich, “How to Prevent an Intelligence Explosion,” in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 383.

194. Schank, “Machines That Think Are In Movies,” 135.

195. Dietterich and Horvitz, “Rise of Concerns About AI.”

196. Quoted in Barrat, *Artificial Intelligence End Human Era*, 170–71.

197. Eliezer S. Yudkowsky, “The Value-Loading Problem,” in *What to Think About Machines That Think*, ed. John Brockman (New York: Harper Perennial, 2015), 365.

198. Barrat, *Artificial Intelligence End Human Era*, 82.

199. R. L. Adams, “Is Artificial Intelligence Dangerous?” *Forbes,* March 25, 2016, http://www.forbes.com/sites/robertadams/2016/03/25/is-artificial-intelligence-dangerous/2/#3dfed4f189b7.

200. Chase, *Surviving AI*, 154.

201. Benford, “Fear Not The AI,” 169.

202. George Zarkadakis, *In Our Own Image*, 248.

203. Ed Felten, “Preparing for the Future of Artificial Intelligence,” *The White House Blog*, May 3, 2016, https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence.

204. “Robotics and Artificial Intelligence Inquiry Launched,” UK Parliament, Science and Technology Committee (Commons), March 24, 2016, http://www.parliament.uk/business/committees/committees-

a-z/commons-select/science-and-technology-committee/news-parliament-2015/robotics-and-artificial-intelligence-inquiry-launch-15-16/.

205. David Moschella (unpublished manuscript, Leading Edge Forum, CSC Corporation, 2016).

206. "The Dark Side of Watson," *NPR*, February 20, 2011, http://www.npr.org/2011/02/20/133916058/the-dark-side-of-watson?ft=1&f=1001.

ACKNOWLEDGMENTS

The author wishes to thank the following individuals for providing input to this report: Rodney Brooks, MIT and Rethink Robotics; Daniel Castro, ITIF; Dan Gordon, Managing Director of dangordontech LLC; David Moschella, Leading Edge Forum, CSC Corporation; Joshua New, ITIF; and Roger Schank, Engines for Education. The author also wishes to thank Kaya Singleton, ITIF, for her editorial assistance. Any errors or omissions are the author's alone.

ABOUT THE AUTHOR

Robert D. Atkinson is the founder and president of ITIF. Atkinson's books include *Innovation Economics: The Race for Global Advantage* (Yale, 2012), *Supply-Side Follies: Why Conservative Economics Fails, Liberal Economics Falters, and Innovation Economics is the Answer* (Rowman & Littlefield, 2006), and *The Past And Future Of America's Economy: Long Waves Of Innovation That Power Cycles Of Growth* (Edward Elgar, 2005). Atkinson holds a Ph.D. in city and regional planning from the University of North Carolina, Chapel Hill, and a master's degree in urban and regional planning from the University of Oregon.

ABOUT ITIF

The Information Technology and Innovation Foundation (ITIF) is a nonprofit, nonpartisan research and educational institute focusing on the intersection of technological innovation and public policy. Recognized as one of the world's leading science and technology think tanks, ITIF's mission is to formulate and promote policy solutions that accelerate innovation and boost productivity to spur growth, opportunity, and progress.

FOR MORE INFORMATION, VISIT US AT WWW.ITIF.ORG.