

## Superminds, not substitutes: Designing human-machine collaboration for a better future of work

JAMES GUSZCZA AND JEFF SCHWARTZ WITH A NOTE BY JOE UCUZOGLU, CEO, DELOITTE US

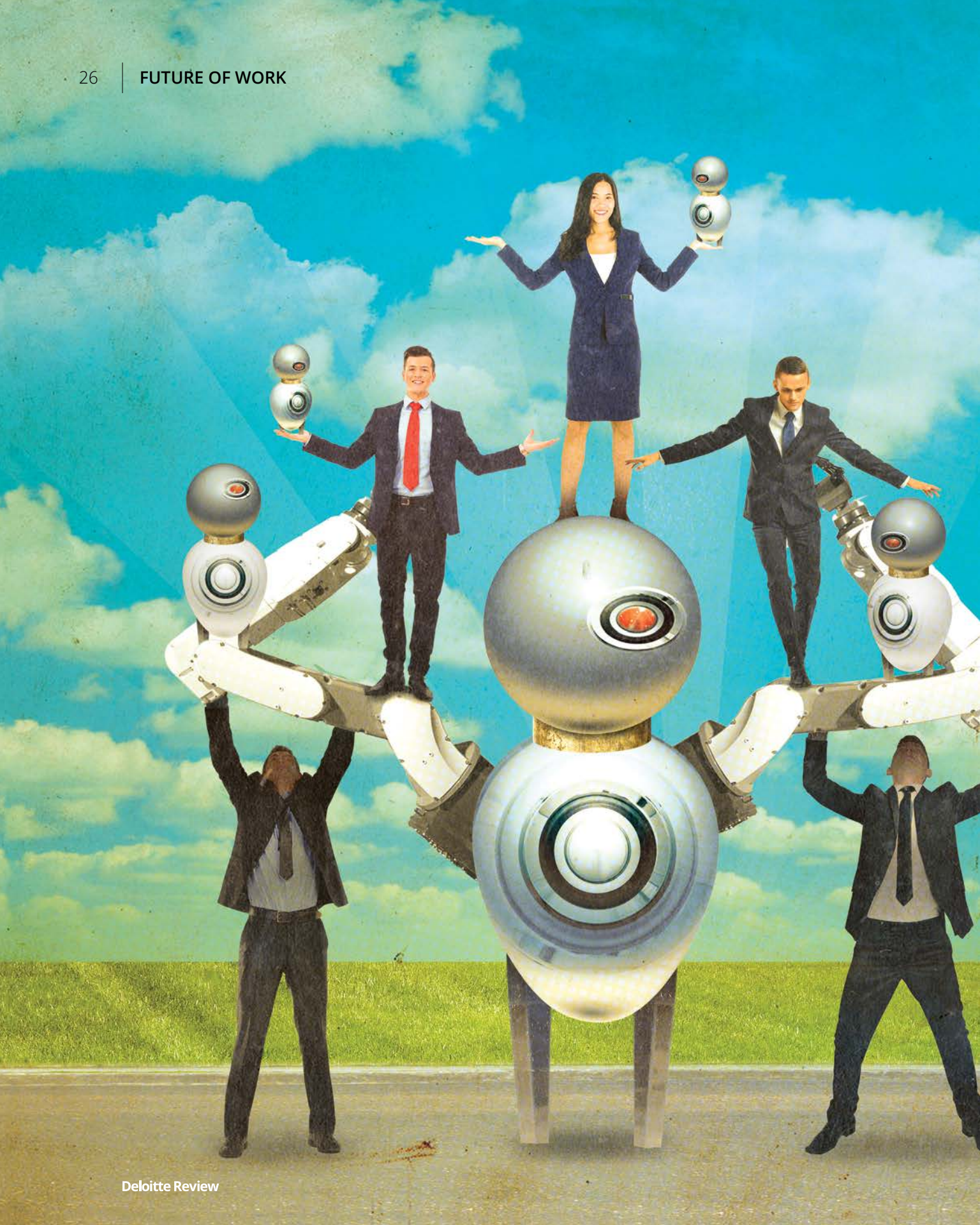
ILLUSTRATION BY DAVID VOGIN

---

**Deloitte.**  
Insights

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see <http://www.deloitte.com/about> for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms. Please see <http://www.deloitte.com/us/about> for a detailed description of the legal structure of the US member firms of Deloitte Touche Tohmatsu Limited and their respective subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting. For information on the Deloitte US Firms' privacy practices, see the US Privacy Notice on Deloitte.com.

Copyright © 2020. Deloitte Development LLC. All rights reserved.



**FUTURE OF WORK**

# Superminds, not substitutes

**DESIGNING HUMAN-MACHINE  
COLLABORATION FOR A  
BETTER FUTURE OF WORK**

JAMES GUSZCZA AND JEFF SCHWARTZ  
WITH A NOTE BY JOE UCUZOGLU,  
CEO, DELOITTE US  
ART BY DAVID VOGIN



“AI systems will need to be smart *and* to be good teammates.”

– Barbara Grosz<sup>1</sup>

**A**RTIFICIAL INTELLIGENCE (AI) is one of the signature issues of our time, but also one of the most easily misinterpreted. The prominent computer scientist Andrew Ng’s slogan “AI is the new electricity”<sup>2</sup> signals that AI is likely to be an economic blockbuster—a general-purpose technology<sup>3</sup> with the potential to reshape business and societal landscapes alike. Ng states:

Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don’t think AI will transform in the next several years.<sup>4</sup>

Such provocative statements naturally prompt the question: How will AI technologies change the role of humans in the workplaces of the future?

An implicit assumption shaping many discussions of this topic might be called the “substitution” view: namely, that AI and other technologies will perform a continually expanding set of tasks better and more cheaply than humans, while humans will remain employed to perform those tasks at which machines cannot (yet) excel. This view comports with the economic goal of achieving scalable efficiency.

The seductiveness of this received wisdom was put into sharp relief by this account of a prevailing attitude at the 2019 World Economic Forum in Davos:

People are looking to achieve very big numbers ... Earlier, they had incremental, 5 to 10 percent goals in reducing their workforce. Now they’re saying, “Why can’t we do it with 1 percent of the people we have?”<sup>5</sup>

But as the personal computing pioneer Alan Kay famously remarked, “A change in perspective is worth 80 IQ points.” This is especially true of discussions of the roles of humans and machines in the future of work. Making the most of human and machine capabilities will require moving beyond received wisdom about both the nature of work *and* the capabilities of real-world AI.

The zero-sum conception of jobs as fixed bundles of tasks, many of which will increasingly be performed by machines, limits one’s ability to reimagine jobs in ways that create new forms of value and meaning.<sup>6</sup> And framing AI as a kind of technology that imitates human cognition makes it easy to be misled by exaggerated claims about the ability of machines to replace humans.

We believe that a change in perspective about AI’s role in work is long overdue. Human and machine capabilities are most productively harnessed by designing *systems* in which humans and machines function collaboratively in ways that complement each other’s strengths and counterbalance each other’s limitations. Following MIT’s Thomas Malone, a pioneer in the study of collective intelligence, we call such hybrid human-machine systems *superminds*.<sup>7</sup>

The change in perspective from AI as a human substitute to an enabler of human-machine superminds has fundamental implications for how organizations should best harness AI technologies:

- Rather than focusing primarily on the ability of computer technologies to *automate* tasks, we would do well to explore their abilities to *augment* human capabilities.
- Rather than adopt a purely technological view of *deploying* technologies, we can cultivate a broader view of *designing systems* of human-computer collaboration. Malone calls this approach “supermind-centered design.”<sup>8</sup>
- Rather than approach AI only as a technology for reducing *costs*, we can also consider its potential for achieving the mutually reinforcing goals of creating more *value* for customers and work that provides more *meaning* for employees.

Compared with the economic logic of scalable growth, the superminds view may strike some as Pollyannaish wishful thinking. Yet it is anything but. Two complementary points—one scientific, one societal—are worth keeping in mind.

First, the superminds view is based on a contemporary, rather than decades-old, scientific understanding of the comparative strengths and limitations of human and machine intelligence. In contrast, much AI-related thought leadership in the business press has arguably been influenced by an understanding of AI rooted in the scientific

zeitgeist of the 1950s and the subsequent decades of science-fiction movies that it inspired.<sup>9</sup>

## Human and machine capabilities are most productively harnessed by designing *systems* in which humans and machines function collaboratively in ways that complement each other’s strengths and counterbalance each other’s limitations.

Second, the post-COVID world is likely to see increasing calls for new social contracts and institutional arrangements of the sort articulated by the Business Roundtable in August 2019.<sup>10</sup> In addition to being more scientifically grounded, a human-centered approach to AI in the future of work will better comport with the societal realities of the post-COVID world. A recent essay by *New America* chief executive Anne-Marie Slaughter conveys today’s moment of opportunity:

The coronavirus, and its economic and social fallout, is a time machine to the future. Changes that many of us predicted would happen over decades are instead taking place in the span of weeks. The future of work is here [and it’s] an opportunity to make the changes we knew we were going to have to make eventually.<sup>11</sup>

To start, let us ground the discussion in the relevant lessons of both computer and cognitive science.

## WHY DEEP LEARNING IS DIFFERENT FROM DEEP UNDERSTANDING

The view that AI will eventually be able to replace people reflects the aspiration—explicitly articulated by the field’s founders in the 1950s—to implement human cognition in machine form.<sup>12</sup> Since then, it has become common for major AI milestones to be framed as machine intelligence taking another step on a path to achieving full human intelligence. For example, the chess grandmaster Garry Kasparov’s defeat by IBM’s Deep Blue computer was popularly discussed as “the brain’s last stand.”<sup>13</sup> In the midst of his defeat by IBM Watson, the *Jeopardy* quiz show champion Ken Jennings joked, “I for one welcome my new computer overlords.”<sup>14</sup> More recently, a *Financial Times* profile of DeepMind CEO Demis Hassabis, published shortly after AlphaGo’s defeat of Go champion Lee Sedol, stated: “At DeepMind, engineers have created programs based on neural networks, modeled on the human brain ... The intelligence is general, not specific. This AI ‘thinks’ like humans do.”<sup>15</sup>

But the truth is considerably more prosaic than this decades-old narrative suggests. It is indeed true that powerful machine learning techniques such as deep learning neural networks and reinforcement learning are inspired by brain and cognitive science. But it does not follow that the resulting AI technologies understand or think in humanlike ways.

So-called “second wave” AI applications essentially result from large-scale statistical inference on massive data sets. This makes them powerful—and often economically game-changing—tools for performing narrow tasks in sufficiently controlled environments. But such AIs possess no common sense, conceptual understanding, awareness of other minds, notions of cause and effect, or intuitive understanding of physics.

What’s more, and even more crucially, these AI applications are reliable and trustworthy only to the extent that they are trained on data that adequately represents the scenarios in which they are to be deployed. If the data is insufficient or the world has changed in relevant ways, the technology cannot necessarily be trusted. For example, a machine translation algorithm would need to be exposed to many human-translated examples of a new bit of slang to hopefully get it right.<sup>16</sup> Similarly, a facial recognition algorithm trained only on images of light-skinned faces might fail to recognize dark-skinned individuals at all.<sup>17</sup>

In contrast, human intelligence is characterized by the ability to learn concepts from few examples, enabling them to function in unfamiliar or rapidly changing environments—essentially the opposite of brute-force pattern recognition learned from massive volumes of (human-)curated data. Think of the human ability to rapidly learn new slang words, work in physical environments that aren’t standardized, or navigate cars through unfamiliar surroundings. Even more telling is a toddler’s ability to learn language from a relative handful of examples.<sup>18</sup> In each case, human intelligence succeeds where today’s “second wave” AI fails because it relies on concepts, hypothesis formation, and causal understanding rather than pattern-matching against massive historical data sets.

It is therefore best to view AI technologies as focused, narrow applications that do not possess the flexibility of human thought. Such technologies will increasingly yield economic efficiencies, business innovations, and improved lives. Yet the old idea that “general” AI would mimic human cognition has, in practice, given way to today’s multitude of practical, narrow AIs that operate very differently from the human mind. Their ability to generally replace human workers is far from clear.

## Why humans are underrated

A key theme that has emerged from decades of work in AI and cognitive science serves as a useful touchstone for evaluating the relative strengths and limitations of human and computer capabilities in various future of work scenarios. This theme is known as “the AI paradox.”<sup>19</sup>

It is hardly news that it is often comparatively easy to automate a multitude of tasks that humans find difficult, such as memorizing facts and recalling information, accurately and consistently weighing risk factors, rapidly performing repetitive tasks, proving theorems, performing statistical procedures, or playing chess and Go. What’s seemingly paradoxical is that the inverse also holds true: Things that come naturally to most people—using common sense, understanding context, navigating unfamiliar landscapes, manipulating objects in uncontrolled environments, picking up slang, understanding human sentiment and emotions—are often the hardest to implement in machines.

The renowned Berkeley cognitive scientist Alison Gopnik states, “It turns out to be much easier to simulate the reasoning of a highly trained adult expert than to mimic the ordinary learning of every baby.”<sup>20</sup> The Harvard cognitive scientist Steven Pinker comments that the main lesson from decades of AI research is that “Difficult problems are easy, and the easy problems are difficult.”<sup>21</sup>

Far from being substitutes for each other, human and machine intelligence therefore turn out to be fundamentally *complementary* in nature. This basic observation turns the substitution view of AI on its head. In organizational psychology, what Scott Page calls a “diversity bonus” results from forming teams composed of different kinds of

thinkers. Heterogeneous teams outperform homogenous ones at solving problems, making predictions, and innovating solutions.<sup>22</sup> The heterogeneity of human and machine intelligences motivates the search for “diversity bonuses” resulting from well-designed teams of human and machine collaborators.

## Difficult problems are easy, and the easy problems are difficult.

A “twist ending” to an AI breakthrough typically used to illustrate the substitution view—Deep Blue’s defeat of the chess grandmaster Garry Kasparov—vividly illustrates the largely untapped potential of the human-machine superminds approach. After his defeat, Kasparov helped create a new game called “advanced chess” in which teams of humans using computer chess programs competed against other such teams. In 2005, a global advanced chess tournament called “freestyle chess” attracted grandmaster players using some of the most powerful computers of the time. The competition ended in an upset victory: Two amateur chess players using three ordinary laptops, each running a different chess program, beat their grandmaster opponents using supercomputers.

Writing in 2010, Kasparov commented that the winners’ “skill at manipulating and ‘coaching’ their computers to look very deeply into positions effectively counteracted the superior chess understanding of their grandmaster opponents and the greater computational power of other participants.” He went on to state what has come to be known as “Kasparov’s Law”:

Weak human + machine + *better process*  
was superior to a strong computer alone  
and, more remarkably, superior to a strong

human + machine + inferior process ...  
Human strategic guidance combined with  
the tactical acuity of a computer was  
overwhelming.<sup>23</sup>

In Thomas Malone’s vernacular, the system of two human players and three computer chess programs formed a *human-computer collective intelligence*—a supermind—that proved more powerful than competing group intelligences boasting stronger human and machine *components*, but inferior supermind *design*.

Though widespread, such phenomena are often hidden in plain sight and obscured by the substitution view of AI. Nonetheless, the evidence is steadily gathering that smart technologies are most effective and trustworthy when deployed in the context of well-designed systems of human-machine collaboration.

We illustrate different modes of collaboration<sup>24</sup>—and the various types of superminds that result—through a sequence of case studies below.

## The best way to predict the future of work is to invent it

### CHATBOTS AND CUSTOMER SERVICE

Call center operators handle billions of customer requests per year—changing flights, refunding purchases, reviewing insurance claims, and so on. To handle the flood of queries, organizations commonly implement chatbots to handle simple queries and escalate more complex ones to human agents.<sup>25</sup> A common refrain, echoing the substitution view, is that human call center operators remain employed to handle tasks beyond the capabilities of today’s chatbots, but that these

jobs will increasingly go by the wayside as chatbots become more sophisticated.

While we do not hazard a prediction of what *will* happen, we believe that call centers offer an excellent example of the surplus value, as well as more intrinsically meaningful work, that *can* be enabled by the superminds approach. In this approach, chatbots and other AI tools function as assistants to humans who increasingly function as problem-solvers. Chatbots offer uniformity and speed while handling massive volumes of routine queries (“Is my flight on time?”) without getting sick, tired, or burned out. In contrast, humans possess the common sense, humor, empathy, and contextual awareness needed to handle lower volumes of less routine or more open-ended tasks at which machines flounder (“My flight was canceled and I’m desperate. What do I do now?”). In addition, algorithms can further assist human agents by summarizing previous interactions, suggesting potential solutions, or identifying otherwise hidden customer needs.

This logic has recently been employed by a major health care provider to better deal with the COVID crisis. A chatbot presents patients with a sequence of questions from the US Centers for Disease Control and Prevention and in-house experts. The AI bot alleviates high volumes of hotline traffic, thereby enabling stretched health care workers to better focus on the most pressing cases.<sup>26</sup>

If this is done well, customers can benefit from more efficient, personalized service, while call center operators have the opportunity to perform less repetitive, more meaningful work involving problem-solving, engaging with the customer, and surfacing new opportunities. In contrast, relying excessively on virtual agents that are devoid of common sense, contextual awareness, genuine



empathy, or the ability to handle unexpected situations (consider the massive number of unexpected situations created by the COVID crisis) poses the risk of alienating customers.

Even if one grants the desirability of this “superminds” scenario, however, will AI technologies not inevitably decrease the number of such human jobs? Perhaps surprisingly, this is not a foregone conclusion. To illustrate, recall what happened to the demand for bank tellers after the introduction of automated teller machines (ATMs). Intuitively, one might think that ATMs dramatically reduced the need for human tellers. But the demand for tellers in fact *increased* after the introduction of ATMs: The technology made it economical for banks to open numerous smaller branches, each staffed with human tellers operating in more high-value customer service, less transactional roles.<sup>27</sup> Analogously, a recent

Bloomberg report told of a company that hired more call center operators to handle the increased volume of complex customer queries after its sales went up thanks to the introduction of chatbots.<sup>28</sup>

A further point is that the introduction of new technologies can give rise to entirely new job categories. In the case of call centers, *chatbot designers* write and continually revise the scripts that the chatbots use to handle routine customer interactions.<sup>29</sup>

This is not to minimize the threat of technological unemployment in a field that employs millions of people. We point out only that using technology to automate simple tasks need not inevitably lead to unemployment. As the ATM story illustrates, characteristically human skills can become *more* valuable when the introduction of a technology increases the number of nonautomatable tasks.

**Characteristically human skills can become *more* valuable when the introduction of a technology increases the number of nonautomatable tasks.**

## RADIOLOGISTS AND “DEEP MEDICINE”

Radiology is another field commonly assumed to be threatened by technological unemployment. Much of radiology involves interpreting medical images—a task at which deep learning algorithms excel. It is therefore natural to anticipate that much of the work currently done by radiologists will be displaced.<sup>30</sup> In a 2017 tweet publicizing a recent paper, Andrew Ng asked, “Should radiologists be worried about their jobs? Breaking news: We can now diagnose pneumonia from chest X-rays better than radiologists.”<sup>31</sup> A year earlier, the deep learning pioneer Geoffrey Hinton declared that it’s “quite obvious that we should stop training radiologists.”<sup>32</sup>

But further reflection reveals a “superminds” logic strikingly analogous to the scenario just discussed in the very different realm of call centers. In his recent book *Deep Medicine*, Eric Topol quotes a number of experts who discuss radiology algorithms as *assistants* to expert radiologists.<sup>33</sup> The Penn Medicine radiology professor Nick Bryan predicts that “within 10 years, no medical imaging study will be reviewed by a radiologist until it has been pre-analyzed by a machine.” Writing with Michael Recht, Bryan states that:

We believe that machine learning and AI will enhance both the value and the professional satisfaction of radiologists by allowing us to spend more time performing functions that add value and influence patient care, and less time doing rote tasks that we neither enjoy nor perform as well as machines.<sup>34</sup>

The deep learning pioneer Yann LeCun articulates a consistent idea, stating that algorithms can automate simple cases and enable radiologists to avoid errors that arise from boredom, inattention,

or fatigue. Unlike Ng and Hinton, LeCun does not anticipate a reduction in the demand for radiologists.<sup>35</sup>

Using AI to automate voluminous and error-prone tasks so that doctors can spend more time providing personalized, high-value care to patients is the central theme of Topol’s book. In the specific case of radiologists, Topol anticipates that these value-adding tasks will include explaining probabilistic outputs of algorithms both to patients and to other medical professionals. For Topol, the “renaissance radiologists” of the future will act less as technicians and more as “real doctors” (Topol’s phrase), and also serve as “master explainers” who display the solid grasp of data science and statistical thinking needed to effectively communicate risks and results to patients.

This value-adding scenario, closely analogous to the chatbot and ATM scenarios, involves the deployment of algorithms as physician *assistants*. But other human-machine arrangements are possible. A recent study combined human and algorithmic diagnoses using a “swarm” tool that mimics the collective intelligence of animals such as honeybees in a swarm.

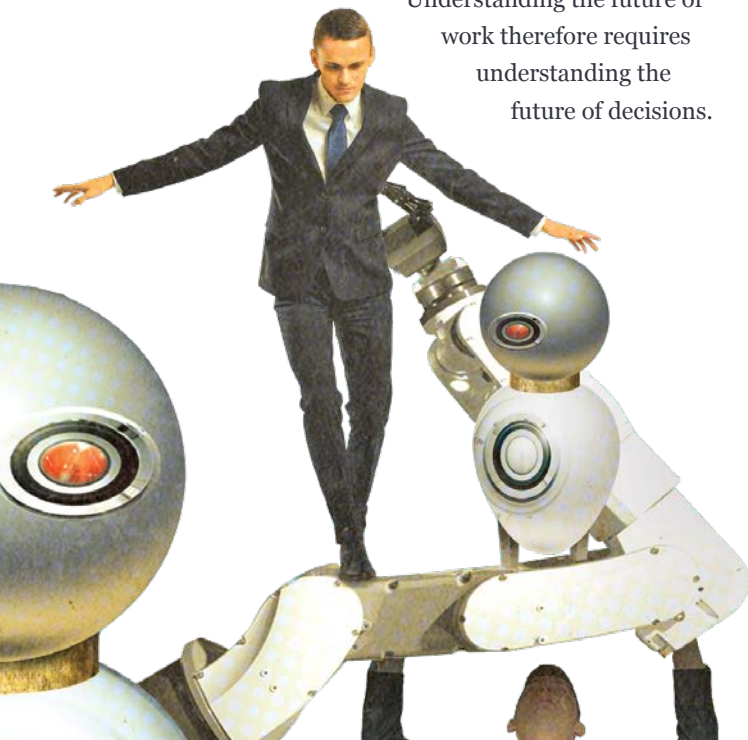


(Previous studies have suggested that honeybee swarms make decisions through a process that is similar to neurological brains.<sup>36</sup>) The investigators found that the hybrid human-machine system—which teamed 13 radiologists with two deep learning AI algorithms—outperformed both the radiologists and the AIs making diagnoses in isolation. Paraphrasing Kasparov’s law, humans + machines + a better process of working together (the swarm intelligence tool) outperforms the inferior process of either humans or machines working alone.<sup>37</sup>

## MACHINE PREDICTIONS AND HUMAN DECISIONS

Using the mechanism of swarm intelligence to create a human-machine collective intelligence possesses the thought-provoking appeal of good science fiction. But more straightforward forms of human-machine partnerships for making better judgments and decisions have been around for decades—and will become increasingly important in the future. The AI pioneer and proto-behavioral economist Herbert Simon wrote that “decision-making is the heart of administration.”<sup>38</sup>

Understanding the future of work therefore requires understanding the future of decisions.



Algorithms are increasingly used to improve economically or societally weighty decisions in such domains as hiring, lending, insurance underwriting, jurisprudence, and public sector casework. Similar to the widespread suggestion of algorithms threatening to put radiologists out of work, the use of algorithms to improve expert decision-making is often framed as an application of machine learning to automate decisions.

In fact, the use of data to improve decisions has as much to do with human psychology and ethics as it does statistics and computer science. Once again, it pays to remember the AI paradox and consider the relative strengths and weaknesses of human and machine intelligence.

The systematic shortcomings of human decision-making—and corresponding relative strengths in algorithmic prediction—have been much discussed in recent years thanks to the pioneering work of Simon’s behavioral science successors Daniel Kahneman and Amos Tversky. Two major sorts of errors plague human decisions:

- **Bias.** Herbert Simon was awarded the Nobel Prize in economics partly for his realization that human-bounded cognition is such that we must rely on heuristics (mental rules of thumb) to quickly make decisions without getting bogged down in analysis paralysis. Kahneman, Tversky, and their collaborators and successors demonstrated that these heuristics are often systematically biased. For example, we confuse ease of imagining a scenario with the likelihood of its happening (the availability heuristic); we cherry-pick evidence that comports with our prior beliefs (confirmation bias) or emotional attachments (the affect heuristic); we ascribe unrelated capabilities to people who possess specific traits we admire (the halo effect); and

we make decisions based on stereotypes rather than careful assessments of evidence (the representativeness heuristic). And another bias—overconfidence bias—ironically blinds us to such shortcomings.

- **Noise.** Completely extraneous factors, such as becoming tired or distracted, routinely affect decisions. For example, when shown the same biopsy results twice, pathologists produced severity assessments that were only 0.61 correlated. (Perfect consistency would result in a correlation of 1.0.)<sup>39</sup> Or simply think about whether you’d prefer to be interviewed or considered for promotion at the end of the day after a very strong job candidate, or closer to the beginning of the day after a very weak candidate.

Regarding noise, algorithms have a clear advantage. Unlike humans, algorithms can make limitless predictions or recommendations without getting tired or distracted by unrelated factors. Indeed, Kahneman—who is currently writing a book about noise—suggests that noise might be a more serious culprit than bias in causing decision traps, and views this as a major argument in favor of algorithmic decision-making.<sup>40</sup>

Bias is the more subtle issue. It is well known that training predictive algorithms on data sets that reflect human or societal biases can encode, and potentially amplify, those biases. For example, using historical data to build an algorithm to predict who should be made a job offer might well be biased against females or minorities if past decisions reflected such biases.<sup>41</sup> Analogously, an algorithm used to target health care “super-utilizers” in order to offer preventative concierge

health services might be biased against minorities who have historically lacked access to health care.<sup>42</sup>

As a result, the topic of machine predictions and human decisions is often implicitly framed as a debate between AI boosters arguing for the superiority of algorithmic to human intelligence on the one side, and AI skeptics warning of “weapons of math destruction” on the other. Adopting a superminds rather than a substitution approach can help people move beyond such unproductive debates.

## Unlike humans, algorithms can make limitless predictions or recommendations without getting tired or distracted by unrelated factors.

One of us (Jim Guszczka) has learned from firsthand experience how predictive algorithms can be productively used as inputs into, not replacements for, human decisions. Many years ago, Deloitte’s Data Science practice pioneered the application of predictive algorithms to help insurance underwriters better select business insurance risks (for example, workers’ compensation or commercial general liability insurance) and price the necessary contracts.

Crucially, the predictive algorithms were designed to meet the end-user underwriters halfway, and the underwriters were also properly trained so that they could meet the algorithms halfway. Black-box machine learning models were typically used only as interim data exploration tools or benchmarks for the more interpretable and easily documented linear models that were usually put

into production. Furthermore, algorithmic outputs were complemented with natural language messages designed to explain to the end user “why” the algorithmic prediction was what it was for a specific case.<sup>43</sup> These are all aspects of what might be called a “human-centered design” approach to AI.<sup>44</sup>

In addition, the end users were given clear training to help them understand when to trust a machine prediction, when to complement it with other information, and when to ignore it altogether. After all, an algorithm can only weigh the inputs presented to it. It cannot judge the accuracy or completeness of those inputs in any specific case. Nor can it use common sense to evaluate context and judge how, or if, the prediction should inform the ultimate decision.

Such considerations, often buried by discussions that emphasize big data and the latest machine learning methods, become all the more pressing in the wake of such world-altering events as the COVID crisis.<sup>45</sup> In such times, human judgment is more important than ever to assess the adequacy of algorithms trained on historical data that might be unrepresentative of the future. Recall that, unlike humans, algorithms possess neither the common sense nor the conceptual understanding needed to handle unfamiliar environments, edge cases, ethical considerations, or changing situations.

Another point is ethical in nature. Most people simply would not want to see decisions in certain domains—such as hiring, university admissions, public sector caseworker decisions, or judicial decisions—meted out by machines incapable of judgment. Yet at the same time, electing *not* to use

algorithms in such scenarios also has ethical implications. Unlike human decisions, machine predictions are consistent over time, and the statistical assumptions and ethical judgments made in algorithm design can be clearly documented. Machine predictions can therefore be systematically audited, debated, and improved in ways that human decisions cannot.<sup>46</sup>

## Human judgment is more important than ever to assess the adequacy of algorithms trained on historical data that might be unrepresentative of the future.

Indeed, the distinguished behavioral economist Sendhil Mullainathan points out that the applications in which people worry most about algorithmic bias are also the very situations in which algorithms—if properly constructed, implemented, and audited—also have the greatest potential to reduce the effects of implicit human biases.<sup>47</sup>

The above account provides a way of understanding the increasingly popular “human-centered AI” tagline: Algorithms are designed not to replace people but rather to extend their capabilities. Just as eyeglasses help myopic eyes see better, algorithms can be designed to help biased and bounded human minds make better judgments and decisions. This is achieved through a blend of statistics and human-centered design. The goal is not merely to optimize an algorithm in a technical statistical sense, but rather to optimize (in a broader sense) a system of humans working with algorithms.<sup>48</sup> In Malone’s vernacular, this is “supermind design thinking.”

## CAREGIVING

New America's Anne-Marie Slaughter comments:

Many of the jobs of the future should also be in caregiving, broadly defined to include not only the physical care of the very old and very young, but also education, coaching, mentoring, and advising. [The COVID] crisis is a reminder of just how indispensable these workers are.<sup>49</sup>

In a well-known essay about health coaches, the prominent medical researcher and author Atul Gawande provides an illuminating example of Slaughter's point. Gawande describes the impact of a health coach (Jayshree) working with a patient (Vibha) with multiple serious comorbidities and a poor track record of improving her diet, exercise, and medical compliance behaviors:

"I didn't think I would live this long," Vibha said through [her husband] Bharat, who translated her Gujarati for me. "I didn't want to live." I asked her what had made her better. The couple credited exercise, dietary changes, medication adjustments, and strict monitoring of her diabetes. But surely she had been encouraged to do these things after her first two heart attacks. What made the difference this time? "Jayshree," Vibha said, naming the health coach from Dunkin' Donuts, who also speaks Gujarati. "Jayshree pushes her, and she listens to her only and not to me," Bharat said. "Why do you listen to Jayshree?" I asked Vibha. "Because she talks like my mother," she said.<sup>50</sup>

The skills of caregivers such as Jayshree are at the opposite end of the pay and education spectra from such fields as radiology. And the AI paradox

suggests that such skills are unlikely to be implemented in machine form anytime soon.

Even so, AI can perhaps play a role in composing purely human superminds such as the one Gawande describes. In Gawande's example, the value wasn't created by generally "human" contact, but rather by the sympathetic engagement of a *specific* human—in this case one, with a similar language and cultural background. AI algorithms have long been used to match friends and romantic partners based on cultural and attitudinal similarities. Such matching could also be explored to improve the quality of various forms of caregiving in fields such as health care, education, customer service, insurance claim adjusting, personal finance, and public sector casework.<sup>51</sup> This illustrates another facet of Malone's superminds concept: Algorithms can serve not only as human collaborators, but also as human connectors.

## Start with why

As Neils Bohr and Yogi Berra each said, it is very hard to predict—especially about the future. This essay is not a series of predictions, but a call to action. Realizing the full benefits of AI technologies will require graduating from a narrow "substitution" focus on automating tasks to a broader "superminds" focus on designing and operationalizing systems of human-machine collaboration.

The superminds view has important implications for workers, business leaders, and societies. Workers and leaders alike must remember that jobs are not mere bundles of skills, and nor are they static. They can and should be creatively reimaged to make the most of new technologies

in ways that simultaneously create more value for customers and more meaningful work for people.<sup>52</sup>

To do this well, it is best to start with first principles. What is the ultimate goal of the job for which the smart technology is intended? Is the purpose of a call center to process calls or to help cultivate enthusiastic, high-lifetime-value customers? Is the purpose of a radiologist to flag problematic tumors, or to participate in the curing, counseling, and comforting of a patient? Is the purpose of a decision-maker to make predictions, or to make wise and informed judgments? Is the purpose of a store clerk to ring up purchases, or to enhance customers' shopping experiences and help them make smart purchases? Once such ultimate goals have been articulated and possibly reframed, we can go about the business of redesigning jobs in ways that make the most of the new possibilities afforded by human-machine superminds.

## Workers and leaders alike must remember that jobs can and should be creatively reimaged to make the most of new technologies in ways that simultaneously create more value for customers and more meaningful work for people.

An analogy from MIT labor economist David Autor conveys the economic logic of why characteristically human skills will remain valued in the highly computerized workplaces of the future. In 1986, the space shuttle Challenger blew up, killing

its entire crew. A highly complex piece of machinery with many interlocking parts and dependencies, the Challenger's demise was due to the failure of a single part—the O-ring. From an economist's perspective, the marginal utility of making this single part more resilient would have been virtually infinite. Autor states that by analogy:

In much of the work that we do, we are the O-rings ... As our tools improve, technology magnifies our leverage and increases the importance of our expertise, judgment, and creativity.<sup>53</sup>

In discussing the logic of human-machine superminds, we do not mean to suggest that achieving them will be easy. To the contrary, such forces as status quo bias, risk aversion, short-term economic incentives, and organizational friction will have to be overcome. Still, the need to overcome such challenges is common to many forms of innovation.

A further challenge relates to the AI paradox: Organizations must learn to better measure, manage, and reward the intangible skills that come naturally to humans but at which machines flounder. Examples include empathy for a call center operator or caregiver; scientific judgment for a data scientist; common sense and alertness for a taxi driver or factory worker; and so

on. Such characteristically human—and often under-rewarded—skills will become more, not less, important in the highly computerized workplaces of the future.

---

## PAIRING HUMANS AND MACHINES TO FORM SUPERMINDS: THRIVING IN A CHANGING WORLD

By Joe Ucuzoglu, CEO, Deloitte US



The unique capabilities of humans matter now more than ever, even in the face of rapid technological progress. In the C-suite and boardrooms, a range of complex topics dominate the agenda: from understanding the practical implications of AI, cloud, and all things digital, to questions of purpose, inclusion, shareholder primacy versus stakeholder capitalism, trust in institutions, and rising populism—and now, the challenges of a global pandemic. In all of these areas, organizations must navigate an unprecedented pace of change while keeping human capabilities and values front and center.

We know from recent years of technological advancement that machines are typically far better than people at looking at huge data sets and making connections. But data is all about the past. What is being created here in the Fourth Industrial Revolution—and in the era of COVID-19—is a future for which past data can be an unreliable guide. Pablo Picasso once said, “Computers are useless. All they can do is provide us with the answers.” The key is seeing the right questions, the new questions—and that’s where humans excel.

What’s more, the importance of asking and answering innovative questions extends up and down entire organizations. It’s not just for C-suites and boardrooms, as Jim Guszczka and Jeff Schwartz share in their examples. It’s about effectively designing systems in which two kinds of intelligence, human and machine, work together in complementary balance, forming *superminds*.

Embracing the concept of superminds and looking holistically at systems of human-machine collaboration provides a way forward for executives. The question is, “What next?” The adjustments all of us have had to make in light of COVID-19 show that we are capable of fast, massive shifts when required, and innovating new ways of working with technology. Eventually, this pandemic will subside, but the currents of digital transformation that have been accelerated out of necessity over the past few months are likely to play out for the rest of our working lives.

How will your organizations become a master of rapid experimentation and learning, of developing and rewarding essential human skills, and of aligning AI-augmented work with human potential and aspirations? ●

---



---

*The authors thank **John Seely Brown, John Hagel, Margaret Levi, Tom Malone, and Maggie Wooll** for helpful conversations. We also thank **Siri Anderson, Abha Kishore Kulkarni, Susan Hogan, and Tim Murphy** for invaluable research assistance and support.*

---

**James Guszczka**, Deloitte Services LP, is Deloitte's US chief data scientist. He lives in Santa Monica, California.

**Jeff Schwartz**, a principal with Deloitte Consulting LLP, is the US leader for the Future of Work and the US leader of Deloitte Catalyst, Tel Aviv. He is based in New York City.



[Read more on www.deloitte.com/insights](http://www.deloitte.com/insights)  
**Talent and workforce effects in the age of AI**

Over the past few years, artificial intelligence has matured into a collection of powerful technologies that are delivering competitive advantage to businesses across industries. But will AI-driven automation render most jobs obsolete, or will humans be working in collaboration with the technology?

**Visit [www.deloitte.com/insights/workforce-ai-adoption](http://www.deloitte.com/insights/workforce-ai-adoption)**

---

## Superminds, not substitutes

page 26

1. Barbara J. Grosz, "Some reflections on Michael Jordan's article 'Artificial intelligence—the revolution hasn't happened yet,'" *Harvard Data Science Review*, July 2, 2019.
2. Shana Lynch, "Andrew Ng: Why AI is the new electricity," Insights by Standard Business, March 11, 2017.
3. A general-purpose technology (GPT) is a type of technology whose breadth of applications and spillover effects can drastically alter entire economies and social structures. Previous examples of GPTs include the invention of writing, the steam engine, the automobile, the mass production system, the computer, the internet, and, of course, electricity.
4. Lynch, "Andrew Ng: Why AI is the new electricity."
5. Kevin Roose, "The hidden automation agenda of the Davos elite," *New York Times*, January 25, 2019.
6. For a discussion on reimagining jobs and "reconstructing work," see: Peter Evans-Greenwood, Harvey Lewis, and Jim Guszcza, "Reconstructing work: Automation, artificial intelligence, and the essential role of humans," *Deloitte Review* 21, July 31, 2017; for more on the theme of redefining and redesigning work to create new sources of value, see: John Hagel, Jeff Schwartz, and Maggie Wooll, "Redefining work for new value: The next opportunity," *MIT Sloan Management Review*, December 3, 2019; for a discussion on the business logic of focusing on value creation rather than cost reduction, see: Jeff Schwartz et al., "Reframing the future of work," *MIT Sloan Management Review*, February 20, 2019.
7. Thomas W. Malone, *Superminds* (New York: Little, Brown Spark, 2018). For an interview with Malone, see: Jim Guszcza and Jeff Schwartz, "Superminds: How humans and machines can work together," *Deloitte Review* 24, January 28, 2019. Malone's and others' work in the emerging multidisciplinary field of collective intelligence is surveyed in the *Handbook of Collective Intelligence*, edited by Thomas W. Malone and Michael S. Bernstein (Cambridge, Massachusetts: MIT Press, 2015). *Superminds* explores the newer forms of collective intelligence that can emerge from groups of humans connected and otherwise augmented by digital and AI technologies.
8. Personal communication from Thomas Malone to Jim Guszcza.
9. For example, the AI pioneer Marvin Minsky served as an adviser to Stanley Kubrick's and Arthur C. Clarke's *2001: A Space Odyssey*. Perhaps that movie's most memorable character was HAL 9000, a computer that spoke fluent English, used commonsense reasoning, experienced jealousy, and tried to escape termination by doing away with the ship's crew. In short, HAL was a computer that implemented a very general form of human intelligence. Minsky and other AI leaders of the day believed that such general, human-imitative artificial intelligences would be achievable by the year 2001.
10. David Gelles and David Yaffe-Bellany, "Shareholder value is no longer everything, top CEOs say," *New York Times*, August 19, 2019.

- 
11. Anne-Marie Slaughter, "Forget the Trump administration. America will save America," *New York Times*, March 21, 2020.
  12. It is commonly agreed that the field of AI originated at a 1956 summer conference at Dartmouth University, attended by such scientific luminaries as John McCarthy, Claude Shannon, Alan Newell, Herbert Simon, and Marvin Minsky. The conference's proposal stated: "The study is to proceed on the basis of the conjecture that *every aspect of learning or any other feature of intelligence* [emphasis added] can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves." See: John McCarthy et al., "A proposal for the Dartmouth Summer Research Project on artificial intelligence," *AI Magazine* 27, no. 4 (2006). Regarding the time frame, the proposal went on to state, "We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it for a summer."
  13. Garry Kasparov, "The brain's last stand," keynote address at DefCon 2017, Kasparov.com, August 18, 2017.
  14. YouTube, "Jeopardy IBM challenge day 3 Computer Overlords clip," video, July 28, 2012.
  15. Murad Ahmed, "Demis Hassabis, master of the new machine age," *Financial Times*, March 11, 2016. This was not an isolated statement. Two days earlier, the *New York Times* carried an opinion piece stating that "Google's AlphaGo is demonstrating for the first time that machines can truly learn and think in a human way." See: Howard Yu, "AlphaGo's success shows the human advantage is eroding fast," *New York Times*, March 9, 2016.
  16. Gary Marcus and Ernest Davis, *Rebooting AI* (New York: Pantheon, 2019). For a relevant excerpt, see: Gary Marcus and Ernest Davis, "If computers are so smart, how come they can't read?," *Wired*, September 10, 2019.
  17. Steve Lohr, "Facial recognition is accurate, if you're a white guy," *New York Times*, February 9, 2018.
  18. The linguist Noam Chomsky famously observed that children can learn grammar based on surprisingly little data, and argued that knowledge of grammar is innate. This is the so-called "poverty of the stimulus" argument. For a modern discussion that points toward potential developments in third-wave AI, see: Amy Perfors, Joshua Tennenbaum, and Terry Regier, "Poverty of the stimulus? A rational approach," MIT.edu, January 2006.
  19. In computer science, the phenomenon goes by the name "Moravec's Paradox" after Hans Moravec, who stated that "it is comparatively easy to make computers exhibit adult-level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility." See: Hans Moravec, *Mind Children* (Cambridge, Massachusetts: Harvard University Press, 1988).
  20. Alison Gopnik, "What do you think about machines that think?," Edge, 2015. Gopnik also states, "One of the fascinating things about the
-

- search for AI is that it's been so hard to predict which parts would be easy or hard. At first, we thought that the quintessential preoccupations of the officially smart few, like playing chess or proving theorems—the corridas of nerd machismo—would prove to be hardest for computers. In fact, they turn out to be easy. Things every dummy can do like recognizing objects or picking them up are much harder.”
21. Steven Pinker, *The Language Instinct* (New York: Harper Perennial Modern Classics, 1994).
  22. Scott E. Page, *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy* (Princeton, New Jersey: Princeton University Press, 2017).
  23. Garry Kasparov, “The chess master and the computer,” *New York Review of Books*, February 11, 2010.
  24. Malone provides a useful taxonomy for designing such human-machine partnerships. Computers can serve as *tools* for humans, *assistants* to humans, *peers* of humans, or *managers* of humans. (See: Malone, *Superminds*.) Malone also discusses these modes of collaboration in “Superminds: How humans and machines can work together,” *Deloitte Review* 24, January 28, 2019.
  25. A 2016 research report found that 80 percent of businesses were either using chatbots or planned to adopt them by 2020. Business Insider Intelligence, “80% of businesses want chatbots by 2020,” December 14, 2016.
  26. Kelley A. Wittbold et al., “How hospitals are using AI to battle COVID-19,” *Harvard Business Review*, April 3, 2020.
  27. This analysis was done by the economist James Bessen. See: James Pethokoukis, “What the story of ATMs and bank tellers reveals about the ‘rise of the robots’ and jobs,” blog, AEI, June 6, 2016.
  28. Ironically, the report is entitled, “How bots will automate call center jobs.” See Bloomberg, “How bots will automate call center jobs,” August 15, 2019.
  29. Ibid.
  30. Ziad Obermeyer and Ezekiel Emanuel, “Predicting the future—big data, machine learning, and clinical medicine,” *New England Journal of Medicine* 375 (2016): pp. 1,216–9, DOI: 10.1056/NEJMp1606181. The authors comment that, because their work largely involves interpreting digitized images, “machine learning will displace much of the work of radiologists and anatomical pathologists.”
  31. Tweet by Eric Topol, Twitter, 7:35 AM, November 16, 2017.
  32. Geff Hinton, “On radiology,” YouTube video, November 24, 2016.
  33. Eric Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (New York: Basic Books, 2019).
  34. Ibid.
  35. All quotes in the above paragraph are from Eric Topol, “Doctors and Patterns,” *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (New York: Basic Books, 2019).
  36. James A.R. Marshall et al., “On optimal decision-making in brains and social insect colonies,” *Journal of the Royal Society* 6 (2009): pp. 1,065–74, DOI: 10.1098/rsif.2008.0511.
-

- 
37. Bhavik N. Patel et al., “Human-machine partnership with artificial intelligence for chest radiograph diagnosis,” *Nature Digital Medicine* 2 (2019). Interestingly, the authors report that the machine diagnoses had a higher true positive rate (sensitivity), while the human diagnosticians had a lower false negative rate (specificity).
  38. Herbert Simon, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, 4th ed. (New York: Free Press, 1997).
  39. Daniel Kahneman et al., “Noise: How to overcome the high, hidden cost of inconsistent decision making,” *Harvard Business Review*, October 2016.
  40. For example: Paul McCaffrey, “Daniel Kahneman: Four keys to better decision making,” *Enterprising Investor*, June 8, 2018; Heloise Wood, “William Collins splashes seven figures on Kahneman’s book about noise,” *Bookseller*, March 20, 2018.
  41. For example: Miranda Bogen, “All the ways hiring algorithms can introduce bias,” *Harvard Business Review*, May 6, 2019.
  42. For example: Ziad Obermeyer et al., “Dissecting racial bias in an algorithm used to manage the health of populations,” *Science* 366, no. 6,464 (2019): pp. 447–53, DOI: 10.1126/science.aax2342.
  43. Upon joining the practice, Jim’s reaction—similar to that of many of our clients—channeled the substitution view: “This is impossible: The data is way too sparse and messy to eliminate the need for human underwriters!” In fact, the solution was superminds, not substitution, in nature. To ensure effective human-algorithm collaboration, expert human judgment was infused into both the design and the daily operation of the solution. Underwriters used their domain and institutional knowledge to help the actuaries and data scientists to create data features and understand data limitations; the actuaries and data scientists carefully sampled and adjusted the historical data to mirror the scenarios in which the algorithm was to be applied. In addition, they collaborated with business owners to establish “guardrails” and business rules for when and how the algorithms should be used for various risks. In contrast to the extreme form of machine learning-based AI in which prior knowledge is thrown away in favor of algorithmic pattern-matching, human knowledge was infused into the solution at multiple steps. And in contrast with the substitution narrative, the solution was explicitly designed to meet the needs of human end users.
  44. For further discussion of this conception of human-centered AI, see: Jim Guszczka, “Smarter together: Why artificial intelligence needs human-centered design,” *Deloitte Review* 22, January 22, 2018.
  45. For example: Matissa Hollister, “AI can help with the COVID-19 crisis—but the right human input is key,” World Economic Forum, March 30, 2020. The article notes: “AI systems need a lot of data, with relevant examples in that data, in order to find these patterns. Machine learning also implicitly assumes that conditions today are the same as the conditions represented in the training data. In other words, AI systems implicitly assume that what has worked in the past will still work in the future.”
  46. For a broader discussion of AI ethics, see: Jim Guszczka et al., “Human values in the loop: Design principles for ethical AI,” *Deloitte Review* 26, January 28, 2020.
  47. *Chicago Booth Review*, “Sendhil Mullainathan says AI can counter human biases,” August 7, 2019.
-

48. A physical analog to this discussion of human-machine “cognitive collaboration” for improved judgments and decisions is the concept of “cobots” developed by the prominent roboticist Rodney Brooks. Brooks and his collaborators have designed robots that can be trained without writing code, give human collaborators visual cues (such as animated “eyes” that move in the direction of a robot arm), and are designed to be safe for humans to collaborate with. See: Erico Guizzo and Evan Ackerman, “Rethink Robotics, pioneer of collaborative robots, shuts down,” *IEEE Spectrum*, October 4, 2018.
  49. Slaughter, “Forget the Trump administration. America will save America.”
  50. Atul Gawande, “The hot spotters: Can we lower medical costs by giving the neediest patients better care?,” *New Yorker*, January 17, 2011.
  51. For example, CareLinx is a startup that uses algorithms to match caregivers with families. See: Kerry Hannon, “Finding the right caregiver, eHarmony style,” *Money*, July 11, 2016.
  52. Hagel, Schwartz, and Wooll, “Redefining work for new value: The next opportunity”; Schwartz et al., “Reframing the future of work.”
  53. David H. Autor, “Why are there still so many jobs? The history and future of workplace automation,” *Journal of Economic Perspectives* 29, no. 3 (2015): pp. 3–30.
-



# Deloitte.

## Insights

 Follow @DeloitteInsight #DeloitteReview

 [www.linkedin.com/company/deloitte-insights](http://www.linkedin.com/company/deloitte-insights)



Go straight to smart.  
Download the Deloitte Insights app

Subscribe to receive email updates at [www.deloitte.com/insights](http://www.deloitte.com/insights)