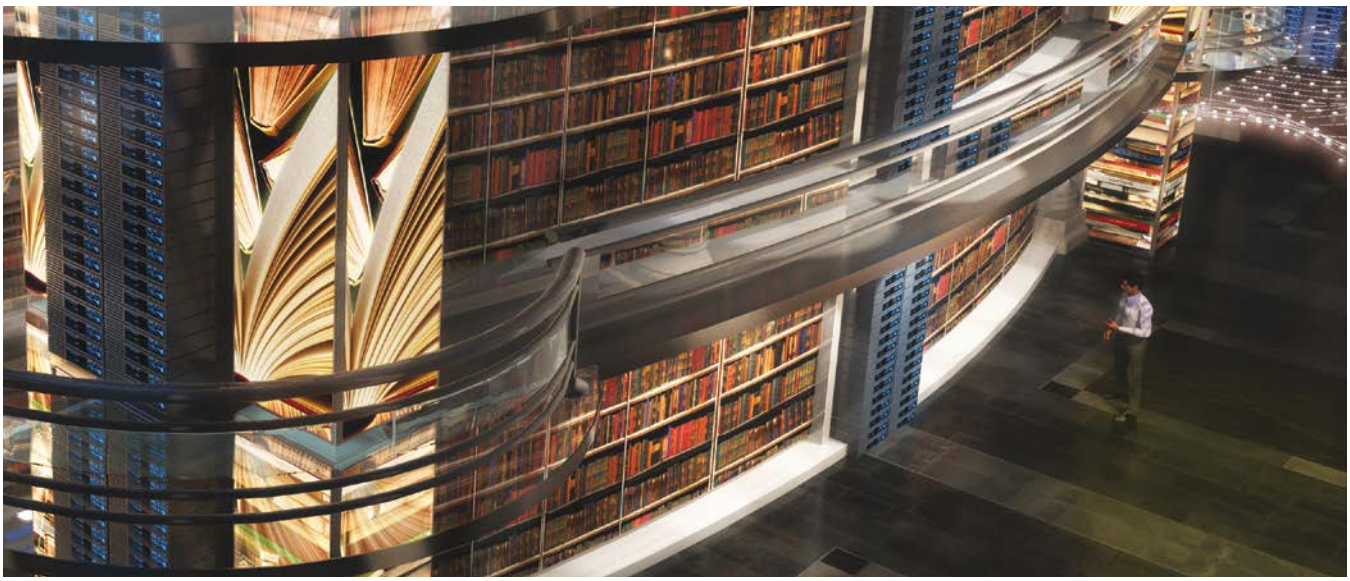
Deloitte Review

ISSUE 22, JANUARY 2018

COMPLIMENTARY ARTICLE REPRINT



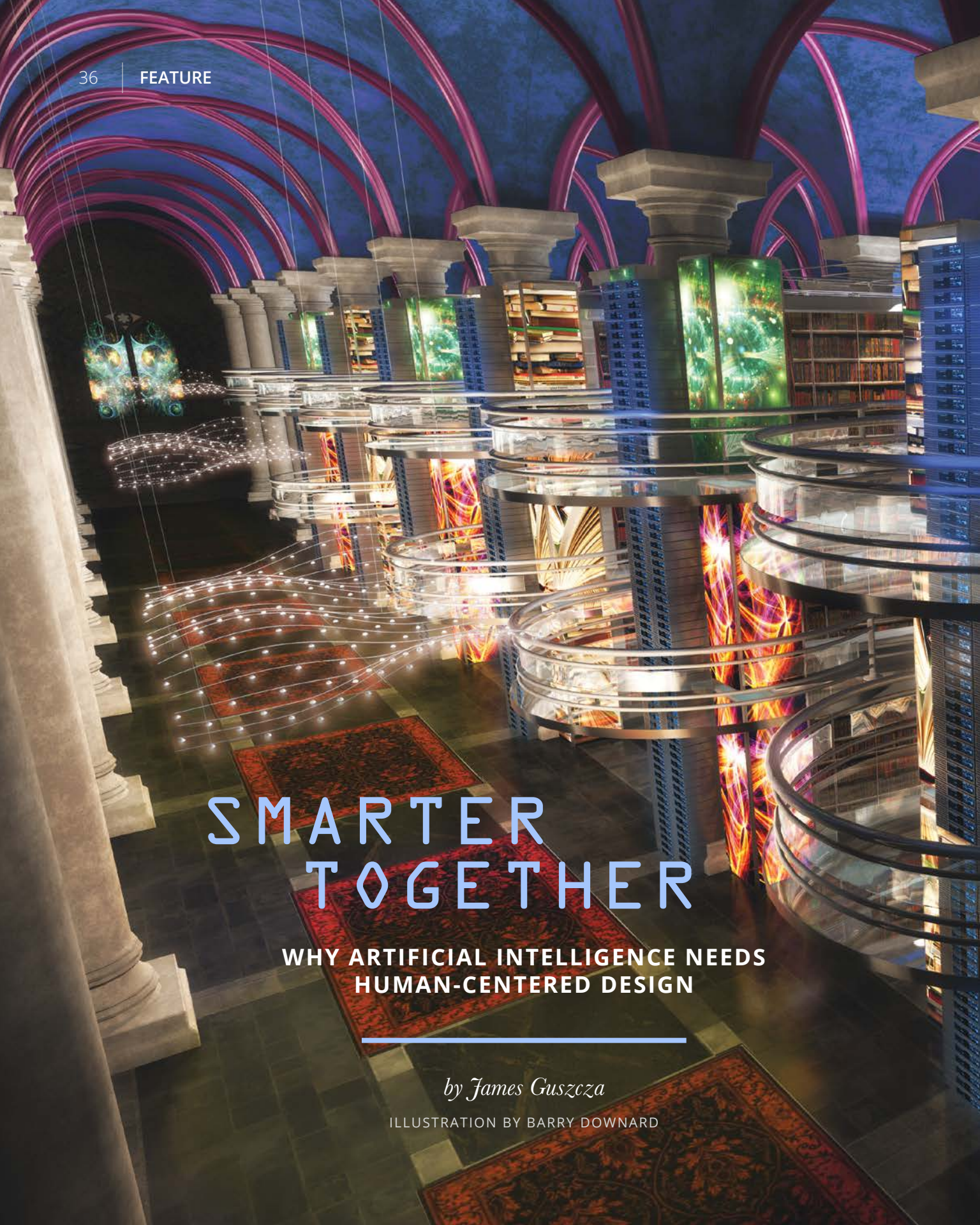
Smarter together

Why artificial intelligence needs human-centered design

By James Guszcza

ILLUSTRATION BY BARRY DOWNARD

Deloitte. Insights



SMARTER TOGETHER

WHY ARTIFICIAL INTELLIGENCE NEEDS HUMAN-CENTERED DESIGN

by James Guszcza

ILLUSTRATION BY BARRY DOWNARD

"Seekers after the glitter of intelligence are misguided in trying to cast it in the base metal of computing."

— Terry Winograd[1]

ARTIFICIAL INTELLIGENCE (AI) HAS emerged as a signature issue of our time, set to reshape business and society. The excitement is warranted, but so are concerns. At a business level, large "big data" and AI projects often fail to deliver. Many of the culprits are familiar and persistent: forcing technological square pegs into strategic round holes, overestimating the sufficiency of available data or underestimating the difficulty of wrangling it into usable shape, taking insufficient steps to ensure that algorithmic outputs result in the desired business outcomes. At a societal level, headlines are dominated by the issue of technological unemployment. Yet it is becoming increasingly clear that AI algorithms embedded in ubiquitous digital technology can encode societal biases, spread conspiracies and promulgate fake news, amplify echo chambers of public opinion, hijack our attention, and even impair our mental well-being.[2]

Effectively addressing such issues requires a realistic conception of AI, which is too often hyped as emerging "artificial minds" on an exponential path to generally out-thinking humans.[3] In reality, today's AI applications result from the same classes of algorithms that have been under development for decades, but implemented on considerably more powerful computers and trained on larger data sets. They are "smart" in narrow senses, not in the general way humans are smart. In functional terms, it is better to view them not as "thinking machines," but as cognitive prostheses that can help humans think better.[4]

In other words, AI algorithms are "mind tools," not artificial minds. This implies that successful applications of AI hinge on more than big data and powerful algorithms. *Human-centered design* is also crucial. AI applications must reflect realistic conceptions of user needs and human psychology. Paraphrasing the user-centered design pioneer Don Norman, AI needs to "accept human behavior the way it is, not the way we would wish it to be."[5]

This essay explores the idea that smart *technologies* are unlikely to engender smart *outcomes* unless they are designed to promote smart *adoption* on the part of human end users. Many of us have experienced the seemingly paradoxical effect of adding a highly intelligent individual to a team, only to witness the team's effectiveness—its "collective IQ"—diminish. Analogously, "smart" AI technology can inadvertently result in "artificial stupidity" if poorly designed, implemented, or adapted to the human social context. Human, organizational, and societal factors are crucial.

An AI framework

It is common to identify AI with machines that think like humans or simulate aspects of the human brain (for a discussion of these potentially misleading starting points, see the sidebar, "The past and present meanings of 'AI'," on page 43). Perhaps even more common is the identification of AI with various machine learning *techniques*. It is true that machine learning applied to big data enables powerful AI applications ranging from self-driving cars to speech-enabled personal assistants. But not all forms of AI involve machine learning being applied to big data. It is better to start with a *functional* definition of AI. "Any program can be considered AI if it does something that we would normally think of as intelligent in humans," writes the computer scientist Kris Hammond. "How the program does it is not the issue, just that is able to do it at all. That is, it is AI if it is smart, but it doesn't have to be smart like us."[6]

Under this expansive definition, the computer automation of routine, explicitly defined "robotic process" tasks such as cashing checks and pre-populating HR forms count as AI. So does the insightful application of data science products, such as using a predictive decision tree algorithm to triage emergency room patients. In each case, an algorithm performs a task previously done only by humans. Yet it is obvious that neither case involves mimicking human intelligence, nor applying machine learning to massive data sets.

Starting with Hammond's definition, it is useful to adopt a framework that distinguishes between AI for *automation* and AI for human *augmentation*.

AI for automation

AI is now capable of automating tasks associated with both *explicit* and *tacit* human knowledge. The former is "textbook" knowledge that can be documented in manuals and rulebooks. It is increasingly practical to capture such knowledge in computer code to achieve robotic process automation (RPA): building software "robots" that perform boring, repetitive, error-prone, or time-consuming tasks, such as processing changes of address, insurance claims, hospital bills, or human resources forms. Because RPA enjoys both low risk and high economic return, it is often a natural starting point for organizations wishing to achieve efficiencies and cost savings through AI. Ideally, it can also free up valuable human time for more complex, meaningful, or customer-facing tasks.

Tacit knowledge might naively seem impervious to AI automation: It is automatic, intuitive "know-how" that is learned by doing, not purely through study or rule-following. Most human knowledge is tacit knowledge: a nurse intuiting that a child has the flu, a firefighter with a gut feel that a burning building is about to collapse, or a data scientist intuiting that a variable reflects a suspicious proxy relationship. Yet the ability of AI applications to automate tasks associated with human tacit knowledge is rapidly progressing. Examples include facial recognition, sensing emotions, driving cars, interpreting spoken language, reading text, writing reports, grading student papers, and even setting people up on dates. In many cases, newer forms of AI can perform such tasks more accurately than humans.

The uncanny quality of such applications make it tempting to conclude that computers are implementing—or rapidly approaching—a kind of human intelligence in the sense that they "understand" what they are doing. That's an illusion. Algorithms "demonstrate human-like tacit knowledge" only in the weak sense that they are constructed or trained using data that encodes the tacit knowledge of a large number of humans working behind the scenes. The term "human-in-the-loop machine learning" is often used to connote this process.[7] While big data and machine learning enable the creation of algorithms that can capture and transmit meaning, this is very different from understanding or originating meaning.

It is tempting to conclude that computers are implementing—or rapidly approaching—a kind of human intelligence in the sense that they "understand" what they are doing. That's an illusion.

Given that automation eliminates the need for human involvement, why should autonomous AI systems require human-centered design? There are several reasons:

Goal-relevance. Data science products and AI applications are most valuable when insightfully designed to satisfy the needs of human end users. For example, typing "area of Poland" into the search engine Bing returns the literal answer (120,728 square miles) along with the note: "About equal to the size of Nevada." The numeric answer is the more accurate, but the intuitive answer will often be more useful.[8] This exemplifies the broader point that "optimal" from the perspective of computer algorithms is not necessarily the same as "optimal" from the perspective of end-user goals or psychology.

Handoff. Many AI systems can run on "autopilot" much of the time, but require human intervention in exceptional or ambiguous situations that

require common sense or contextual understanding. Human-centered design is needed to ensure that this "handoff" from computer to human happens when it should, and that it goes smoothly when it does happen. Here's an admittedly low-stakes personal example of how AI can give rise to "artificial stupidity" if the handoff doesn't go well. I recently hailed a cab for a trip that required only common sense and a tiny amount of local knowledge—driving down a single major boulevard. Yet the driver got lost because he was following the (as it turned out, garbled) indications of a smartphone app. A "low confidence" or "potentially high interference" warning might have nudged the driver to rethink his actions rather than suppressing his common sense in favor of the algorithmic indication.

This illustrates the general issue known as "the paradox of automation":[9] The more reliant we become on technology, the less prepared we are to take control in the exceptional cases when the technology fails. The problem is thorny because the conditions under which humans must take control require *more*, not less, skill than the situations that can be handled by algorithms—and automation technologies can erode precisely the skills needed in such scenarios. Keeping human skills sufficiently fresh to handle such situations might sometimes involve relying on automation less than the technology makes practical. Once again, "optimal" from a narrowly technological perspective might differ from "optimal" for a human-computer system.

Feedback loops. Automated algorithmic decisions can reflect and amplify undesirable patterns in the data they are trained on. A vivid recent example is Tay, a chatbot designed to learn about the world through conversations with its users. The chatbot had to be switched off within 24 hours after pranksters trained it to utter racist, sexist, and fascist statements.[10] Other examples of algorithms reflecting and amplifying undesirable societal biases are by now ubiquitous. For such reasons, there is an increasing call for chatbot and search-engine design to optimize not only for speed and algorithmic accuracy, but also user behavior and societal biases encoded in data.[11]

Psychological impact. Just as user behavior can impair algorithms, so can algorithms impair user behavior. Two serious contemporary issues illustrate the point. First, it is becoming increasingly clear that AI-enabled entertainment and social media applications can impair human well-being in a number of ways. Compulsive email checking can cause people to shortchange themselves on sleep and distract themselves on the job; excessive social media use has been linked with feelings of unhappiness and "fear of missing out"; and Silicon Valley insiders increasingly worry about people's minds being "hijacked" by addictive technologies.[12]

Second, there is increasing concern that the collaborative filtering of news and commentary can lead to "filter bubbles" and "epistemic gated communities" of opinion. In his recent book *#Republic*, legal scholar Cass Sunstein argues this can exacerbate group polarization and undermine reasoned deliberation, a prerequisite to a well-functioning democracy. He suggests social media recommendation engines be imbued with a form of human-centered design: the spontaneous, serendipitous discoveries of alternate news stories and opinion pieces to help ward off polarization and groupthink.[13] Sunstein analogizes this with the perspective-altering serendipitous encounters and discoveries characteristic of living in a dense, diverse, walkable urban environment.

> "The technology is the easy part. The hard part is figuring out the social and institutional structures around the technology."

In short, it can be counterproductive to deploy technologically sophisticated autonomous AI systems without a correspondingly sophisticated approach to human-centered design. As John Seely Brown presciently remarked, "The technology is the easy part. The hard part is figuring out the social and institutional structures around the technology."[14]

Yet automation is only part of the story. Algorithms can also be used to augment human cognitive capabilities—both System 1 "thinking fast," and System 2 "thinking slow." It is possible to achieve forms of human-computer collective intelligence—provided we adopt a human-centered approach to AI.

AI for augmented thinking slow

Psychologists have long known that even simple algorithms can outperform expert judgments at predictive tasks ranging from making medical diagnoses to estimating the odds a parolee will recidivate to scouting baseball players to underwriting insurance risks. The field was initiated in 1954, with the publication of the book *Clinical Versus Statistical Prediction* by psychologist and philosopher Paul Meehl.

Meehl was a hero to the young Daniel Kahneman, the author of *Thinking, Fast and Slow*,[15] whose work with Amos Tversky uncovered the human mind's surprising tendency to rely on intuitively coherent but predictively dubious narratives, rather than logical assessments of evidence. Behavioral economists such as Richard Thaler point out that this systematic feature of human psychology results in persistently inefficient markets and business processes that can be rationalized through the use of algorithm-assisted decision-making—"playing Moneyball."[16] Just as eyeglasses compensate for myopic vision, data and algorithms can compensate for cognitive myopia.

Meehl's and Kahneman's work implies that in many situations, algorithms should be used to *automate* decisions. Overconfident humans tend to override predictive algorithms more often than they should.[17] When possible, it is therefore best to employ human judgment in the design of algorithms, and remove humans from case-by-case decision-making. But this is not always possible. For example, procedural justice implies that it would be unacceptable to replace a judge making parole decisions with the mechanical outputs of a recidivism prediction algorithm. A second issue is epistemic in nature. Many decisions, such as making a complex medical diagnosis, underwriting a rare insurance risk, making an important hiring decision, and so on are not associated with a rich enough body of historical data to enable the construction of a sufficiently reliable predictive algorithm. In such scenarios, an imperfect algorithm can be used not to *automate* decisions, but rather to generate anchor points to *augment and improve* human decisions.

How might this work? A suggestive illustration comes from the world of chess. Several years after IBM Deep Blue defeated the world chess champion Garry Kasparov, a "freestyle chess" competition was held, in which any combination of human and computer chess players could compete. The competition ended with an upset victory that Kasparov subsequently discussed:

> The winner was revealed to be not a grandmaster with a state-of-the-art PC but a pair of amateur American chess players using three computers at the same time. Their 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. 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.[18]

This idea that weak human + machine + better process outperforms strong human + machine + inferior process has been called "Kasparov's law." A

corollary is that user-centered design is necessary to both the creation and deployment of algorithms intended to improve expert judgment. Just as a cyclist can perform better with a bicycle that was designed for her and that she has been trained to use, an expert can make better decisions with an algorithm built with her needs in mind, and which she has been trained to use.[19]

To that end, human-centric AI algorithms should suitably reflect the information, goals, and constraints that the decision-maker tends to weigh when arriving at a decision; the data should be analyzed from a position of domain and institutional knowledge, and an understanding of the process that generated it; an algorithm's design should anticipate the realities of the environment in which it is to be used; it should avoid societally vexed predictors; it should be peer-reviewed or audited to ensure that unwanted biases have not inadvertently crept in; and it should be accompanied by measures of confidence and "why" messages (ideally expressed in intuitive language) explaining why a certain algorithmic indication is what it is. For example, one would not wish to receive a black-box algorithmic indication of the odds of a serious disease without the ability to investigate the reasons why the indication is what it is.

But even these sorts of algorithm design considerations are not sufficient. The overall decision *environment*—which includes both the algorithm and human decision-makers—must be similarly well-designed. Just as the freestyle chess winners triumphed because of their deep familiarity and experience with both chess and their chess programs, algorithm end users should have a sufficiently detailed understanding of their tool to use it effectively. The algorithm's assumptions, limitations, and data features should therefore be clearly communicated through writing and information visualization. Furthermore, guidelines and business rules should be established to convert predictions into prescriptions and to suggest when and how the end user might either override the algorithm or complement its recommendations with other information. End users can also be trained to "think slow," more like statisticians. Psychologists Philip Tetlock and Barbara Mellors have found that training decision-makers in probabilistic reasoning and avoiding cognitive biases improves their forecasting abilities.[20] Building accurate algorithms is not enough; user-centered design is also essential.

3D: Data, digital, and design for augmented thinking fast

Economic value comes not from AI algorithms, but from AI algorithms that have been properly designed for, and adapted to, human environments. For example, consider the "last mile problem" of predictive algorithms: No algorithm will yield economic value unless it is properly acted upon to drive results. While this is a truism, it is also one of the easiest things for organizations to get wrong. One recent study estimated that 60 percent of "big data" projects fail to become operationalized.[21]

A good example of model operationalization is the predictive algorithm used to rank all of the building sites in New York City in order of riskiness. Prior to the algorithm's deployment, roughly 10 percent of building inspections resulted in an order to vacate. After deployment, the number rose to 70 percent.[22] This is a classic example of predictive analytics being used to improve "System 2" decision-making, as discussed in the previous section. Still more value can be derived through the application of what behavioral economists call *choice architecture*, aka "nudges."[23] Consider risks that are either ambiguous or not quite dangerous enough (yet) to warrant a visit from the city's limited cadre of building inspectors. Such lesser risks could be prompted to "self-cure" through, for example, nudge letters that have been field-tested and optimized using randomized controlled trials (RCTs). Analogous "push the worst, nudge the rest" strategies can be adopted for algorithms designed to identify unhygienic restaurants, inefficient programs, unsafe workplaces, episodes of waste, fraud, abuse, or expense or tax policy noncompliance.

In certain cases, applying choice architecture will be crucial to the economic success and societal

acceptability of an AI project. For example, the state of New Mexico recently adopted a machine learning algorithm designed to flag unemployment insurance recipients who are *relatively* likely to be improperly collecting large unemployment insurance (UI) benefits. The word "relatively" is important. While the highest-scoring cases were many times more likely than average to be improperly collecting UI benefits, most were (inevitably) false positives. This counterintuitive result is known as the "false positive paradox."[24] The crucial implication is that naively using the algorithm to cut off benefits would harm a large number of citizens in genuine need of them. Rather than adopt this naive strategy, the state therefore field-tested a number of pop-up nudge messages on the computer screens of UI recipients performing their weekly certifications. The most effective such message cut improper payments in half: informing recipients that "99 out of 100 people in <your county> accurately report earnings each week."[25]

The human-centered nature of choice architecture can therefore enable AI applications that are at once economically beneficial and pro-social.[26] Furthermore, the case for choice architecture is stronger than ever in our era of big data and ubiquitous digital technologies. Fine-grained behavioral data of large populations may increasingly enable personalized interventions appropriate to individual cases. Imbuing our ever-present digital

THE PAST AND PRESENT MEANINGS OF "AI"

While the term "AI" has made a major comeback, the term has come to mean something quite different from what its founders had in mind. Today's AI technologies are not generally intelligent thinking machines; they are applications that help humans think better.

The field of artificial intelligence dates back to a specific place and time: a conference held at Dartmouth University in the summer of 1956. The conference was convened by John McCarthy, who coined the term "artificial intelligence" and defined it as the science of creating machines "with the ability to achieve goals in the world."[27]

McCarthy's definition is still very useful. But the conference attendees—including legendary figures such as Marvin Minsky, Alan Newell, Claude Shannon, and Herbert Simon—aspired to a much more ambitious goal: to implement a complete version of human thought and language within computer technology. In other words, they wished to create *general* artificial intelligence, modeled on human general intelligence. Their proposal stated:

> The study is to proceed on the basis of the conjecture that *every aspect* of learning or any other feature of intelligence 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.[28]

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." This optimism might seem surprising in hindsight. But it is worth remembering that the authors were writing in the heyday of both B. F. Skinner's behaviorist psychology and the logical positivist school of philosophy. In this intellectual climate, it was natural to assume that human thought was ultimately a form of logical calculation. Our understanding of both human psychology and the challenges of encoding knowledge in logically perfect languages has evolved considerably since the 1950s.

It is a telling historical footnote that Minsky subsequently advised the director Stanley Kubrick during the movie adaptation of Arthur C. Clarke's novel *2001: A Space Odyssey*. That story's most memorable

(continued) >

THE PAST AND PRESENT MEANINGS OF "AI" *(continued)*

character was HAL—a sapient machine capable of conceptual thinking, commonsense reasoning, and a fluid command of human language. Minsky and the other Dartmouth Conference attendees believed that such generally intelligent computers would be available by the year 2001.

Today, AI denotes a collection of technologies that, paraphrasing McCarthy's original definition, excel at specific *tasks* that could previously only be performed by humans. Although it is common for commentators to state that such technologies as the DeepFace facial recognition system or DeepMind's AlphaGo are "modeled on the human brain" or can "think like humans do," such statements are misleading. An obvious point is that today's AI technologies—and all on the foreseeable horizon—are *narrow* AI point solutions. An algorithm designed to drive a car is useless for diagnosing a patient, and vice versa.

Furthermore, such applications are far from the popular vision of computers that implement (super) human thought. For example, deep learning neural network algorithms can identify tumors in X-rays, label photographs with English phrases, distinguish between breeds of animals, and distinguish people who are genuinely smiling from those who are faking it—often more accurately than we can.[29] But this does not involve algorithmically representing such concepts as "tumor," "pinscher," or "smile." Rather, deep learning neural network models are trained on large numbers of digitized photographs that have already been labeled by humans.[30] Such models neither imitate the mind nor simulate the brain. They are predictive models—akin to regression models—typically trained on millions of examples and containing millions of uninterpretable parameters. The technology can perform tasks hitherto performed only by humans; but it does not result from emulating the human brain or mimicking the human mind.

While such data-driven AI applications have massive practical applications and economic potential, they are also "rigid" in the sense that they lack contextual awareness, causal understanding, and commonsense reasoning capabilities. A crucial implication is that they cannot be relied on in "black swan" scenarios or environments significantly different from those they were trained in. Just as a credit scoring algorithm trained on data about US consumers would not yield a reliable score for an immigrant from another country, a self-driving car trained in Palo Alto would not necessary perform as well in Pondicherry.

technologies with choice architecture better can improve both engagement and outcomes. Health wearables are a familiar example. Prominent behavioral health experts point out that such devices are facilitators—but not drivers—of better health behaviors.[31] Using such wearables to merely gather data and generate information reports is simply not enough to prompt most of us to follow through and change our behaviors. A more promising strategy is to use data gathered by wearables to target, inform, and personalize such nudge tactics as peer comparisons, commitment contracts, gamification interventions, and habit-formation programs.[32]

This illustrates a general principle that might be called "3D": *Data* and *digital* tech are facilitators; psychologically informed *design* is also needed to drive better engagement and outcomes. 3D thinking can enable innovative products and business models. Consider, for example, the telematics data emanating from cars connected to the Internet of Things, which insurers already use to more accurately price personal and commercial auto insurance contracts. This data can also be used to spur loss prevention; a young male driver might be given a discount on his expensive auto insurance policy if he follows data-generated prescriptions to improve his driving behaviors. Choice architecture enables a further idea: Natural language generation tools could be used to automatically produce periodic data-rich reports containing both helpful tips as well as peer comparison nudge messages. For example, being informed that his highway-driving is riskier

than that of most of his peers might be a highly effective, low-cost way to prompt safer driving. Such strategies can enable insurers to be less product-centric and more customer-centric in a way that benefits the company, the policyholder, and society as a whole.

Whether intended for automation or human augmentation, AI systems are more likely to yield economic benefits and societal acceptability if user needs and psychological factors are taken into account. Design can help close the gap between AI algorithm *outputs* and improved *outcomes* by enabling better modes of human-computer collaboration. It is therefore fitting to give the last word to Garry Kasparov, from his recent book, *Deep Thinking*: "Many jobs will continue to be lost to intelligent automation. But if you're looking for a field that will be booming for many years, get into human-machine collaboration and process architecture and design."

Both figuratively and literally, the last word is: design.

JAMES GUSZCZA is Deloitte Consulting's US Chief Data Scientist, based in Santa Monica, California.

SMARTER TOGETHER

page 36

1. See Terry Winograd, "Thinking machines: Can there be? Are we?," originally published in James Sheehan and Morton Sosna, *The Boundaries of Humanity: Humans, Animals, Machines* (Berkeley: The University of California Press, 1991).

2. See, for example, Matthew Hutson, "Even artificial intelligence can acquire biases against race and gender," *Science*, April 13, 2017; "Fake news: You ain't seen nothing yet," *Economist*, July 1, 2017; Paul Lewis, "'Our minds can be hijacked': The tech insiders who fear a smartphone dystopia," *Guardian*, October 6, 2017; Holly B. Shakya and Nicholas A. Christakis, "A new, more rigorous study confirms: The more you use Facebook, the worse you feel," *Harvard Business Review*, April 10, 2017.

3. The prominent AI researcher Yann LeCun discusses AI hype and reality in an interview with James Vincent, "Facebook's head of AI wants us to stop using the Terminator to talk about AI," *The Verge*, October 26, 2017.

4. For further discussion of this theme, see James Guszcza, Harvey Lewis, and Peter Evans-Greenwood, "*Cognitive collaboration: Why humans and computers think better together*," Deloitte University Press, January 23, 2017.

5. See Don Norman, *The Design of Everyday Things* (New York: Basic Books, 2013).

6. See Kris Hammond, "What is artificial intelligence?," *Computerworld*, April 10, 2015.

7. See Lukas Biewald, "Why human-in-the-loop computing is the future of machine learning," *Computerworld*, November 13, 2015.

8. Bing search performed on October 11, 2017. This is the result of work led by cognitive scientists Dan Goldstein and Jake Hoffman on helping people better grasp large numbers (see Elizabeth Landau, "How to understand extreme numbers," *Nautilus*, February 17, 2017).

9. I owe this point to Harvey Lewis. For a discussion, see Tim Harford, "Crash: How computers are

setting us up for disaster," *Guardian*, October 11, 2016.

10. See Antonio Regalado, "The biggest technology failures of 2016," *MIT Technology Review*, December 27, 2016.

11. Many examples of algorithmic bias are discussed in April Glaser, "Who trained your AI?," *Slate*, October 24, 2017.

12. See the references in the first endnote. In their recent book *The Distracted Mind: Ancient Brains in a High-Tech World* (MIT Press, 2016), Adam Gazzaley and Larry Rosen explore the evolutionary and neuroscientific reasons why we are "wired for" being distracted by digital technology, and explore the cognitive and behavioral interventions to promote healthier relationships with technology.

13. *#Republic* by Cass R. Sunstein (New Jersey: Princeton University Press, 2017).

14. See John Seely Brown, "Cultivating the entrepreneurial learner in the 21st century," March 22, 2015. Available at www.johnseelybrown.com.

15. See Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011).

16. See Richard Thaler and Cass Sunstein, "Who's on first: A review of Michael Lewis's 'Moneyball: The Art of Winning an Unfair Game,'" University of Chicago Law School website, September 1, 2003.

17. See Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," *Journal of Experimental Psychology*, 2014. In recent work, Dietvorst and Massey have found that humans are more likely to embrace algorithmic decision-making if they are given even a small amount of latitude for occasionally overriding them.

18. Garry Kasparov, "The chess master and the computer," *New York Review of Books*, February 11, 2010.

19. This analogy is not accidental. Steve Jobs described computers as "bicycles for the mind" in the 1990 documentary *Memory & Imagination: New Pathways to the Library of Congress*. Clip available at "Steve Jobs, 'Computers are like a bicycle for our minds.' - Michael Lawrence Films," YouTube video, 1:39, posted by Michael Lawrence, June 1, 2006.

20. This is discussed by Philip Tetlock in *Superforecasting: The Art and Science of Prediction* (Portland: Broadway Books, 2015).

21. See Gartner, "Gartner says business intelligence and analytics leaders must focus on mindsets and culture to kick start advanced analytics," press release, September 15, 2015.

22. See Viktor Schönberger and Kenneth Cukier, "Big data in the big apple," *Slate*, March 6, 2013.

23. Choice architecture can be viewed as the human-centered design of choice environments. The goal is to make it easy for us to make better decisions through automatic "System 1" thinking, rather than through willpower or effortful System 2 thinking. For example, if a person is automatically enrolled in a retirement savings plan but able to opt out, they are considerably more likely to save

more for retirement than if the default is set to not-enrolled. From the perspective of classical economics, such minor details in the way information is provided or choices are arranged—Richard Thaler, Nobelist and the co-author of *Nudge: Improving Decisions About Health, Wealth and Happiness* (Yale University Press, 2008) calls them "supposedly irrelevant factors"—should have little or no effect on people's decisions. Yet the core insight of choice architecture is that the long list of systematic human cognitive and behavioral quirks can be used as design elements for choice environments that make it easy and natural for people to make the choices they would make if they had unlimited willpower and cognitive resources. In his book *Misbehaving* (W. W. Norton & Company, 2015), Thaler acknowledged the influence of Don Norman's approach to human-centered design in the writing of *Nudge*.

24. When the population-level base rate of an event is low, and the test or algorithm used to flag the event is imperfect, false positives can outnumber true positives. For example, suppose 2 percent of the population has a rare disease and the test used to identify it is 95 percent accurate. If a specific patient from this population tests positive, that person has a roughly 29 percent chance of having the disease. This results from a simple application of Bayes' Theorem. A well-known cognitive bias is "base rate neglect"—many people would assume that the chance of having the disease is not 29 percent but 95 percent.

25. For more details, see Joy Forehand and Michael Greene, "Nudging New Mexico: Kindling compliance among unemployment claimants," *Deloitte Review* 18, January 2016.

26. This theme is explored in Jim Guszcza, David Schweidel, and Shantanu Dutta, "The personalized and the personal: Socially responsible uses of big data," *Deloitte Review* 14, January 2014.

27. McCarthy defined artificial intelligence as "the science and engineering of making intelligent machines, especially intelligent computer programs," and defined intelligence as "the computational part of the ability to achieve goals in the world." He noted that "varying kinds and degrees of intelligence occur in people, many animals, and some machines." See John McCarthy, "What is artificial intelligence?," Stanford University website, accessed October 7, 2017.

28. The original proposal can be found in John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, "A proposal for the Dartmouth Summer Research Project on artificial intelligence," *AI Magazine* 27, no. 4 (2006).

29. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," arXiv, February 6, 2015.

30. It is common for such algorithms to fail in certain ambiguous cases that can be correctly labeled by human experts. These new data points can then be used to retrain the models, resulting in improved accuracy. This virtuous cycle of human labeling and machine learning is called "human-in-the-loop computing." See, for example, Biewald,

"Why human-in-the-loop computing is the future of machine learning."

31. See Mitesh Patel, David Asch, and Kevin Volpp, "Wearable devices as facilitators, not drivers, of health behavior change," *Journal of the American Medical Association* 313, no. 5 (2015).

32. For further discussion of the themes in this section, see James Guszcza, "The last-mile problem: How data science and behavioral science can work together," *Deloitte Review* 16, January 2015.