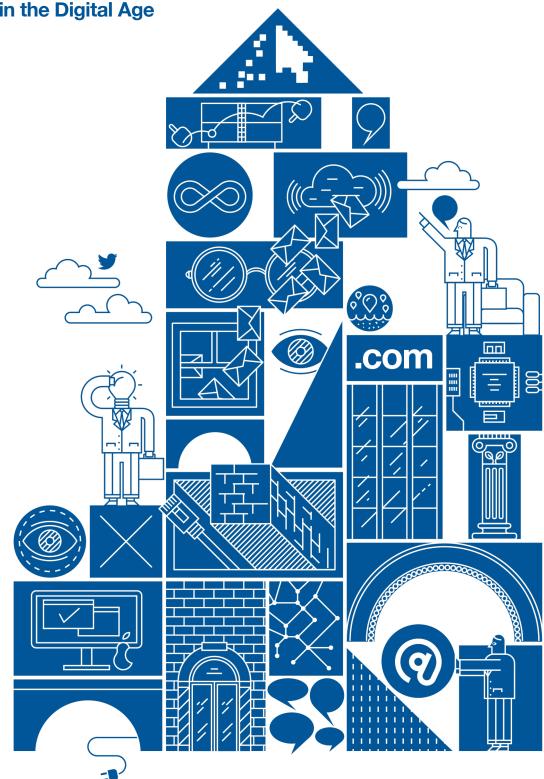
Reinventing the Company in the Digital Age



Reinventing the Company

in the Digital Age



Big Data and the Future of Business

Kenneth Cukier

Cukier makes the point that no area of human endeavour or industrial sector will be immune from the complete shakeup that Big Data is about to bring, as it transforms society, politics, and business. As he says neatly "More isn't just more. More is new. More is better. More is different."

Although there are still limitations (usually based on privacy issues) on what one can get and do with data, most of our assumptions about the cost of collecting and the difficulty of processing data need to be completely overturned.

He sees this new world of data impacting on two main areas of public policy and regulation:

- Employment: we can expect a wave of structural unemployment to spring from the technology in the medium term.
- Privacy, as collecting data happens invisibly and passively, as a by-product of another service.

Cukier envisions Big Data changing business, and business changing society. He hopes that the benefits outweigh the drawbacks, but remains ultimately cautious as he sees society as not very proficient at handling all the data that we can already collect.

Kenneth Cukier The Economist



Kenneth Cukier is the Data Editor of *The Economist* in London and co-author with Viktor Mayer-Schönberger of the award-winning book *Big Data: A Revolution That Transforms How We Live, Work, and Think* (2013), a *New York Times* Bestseller translated into twenty languages. He is a regular commentator on the BBC, CNN and NPR, and a member of the World Economic Forum's council on data-driven development. Mr Cukier was a research fellow at Harvard's Kennedy School of Government in 2002-04. He is a board director of International Bridges to Justice and a member of the Council on Foreign Relations.

Key Features for the Company of the Future:

Harness Big Data or "Moonshot" Innovations

Companies typically look for a 10% improvement. But this is only a "sustaining innovation" learning from data to do the same thing that is already done, only marginally better. Over time, this approach distracts from the chances of major breakthroughs. Clever companies of the future will constantly be on the lookout to use Big Data to attain the "moonshot": disruptive innovations that create radically novel products and new markets. These do not improve what already exists. but open up ways to do things in entirely new ways.

Use Big Data for "Small Wins" That Scale Up

Alongside the "moonshot" is its contrary: data used to identify tiny improvements that on their own are insignificant, but when combined together, add up to an outsized gain in performance. In the past, it wasn't worthwhile for a company to go after such "small wins." But as the cost of handling data falls and it is easy to operationalize what is learned, the ability to find and put into practice these "small wins" becomes feasible. It will be the backbone of how companies of the future compete.

Embrace Data Alongside Values and Intuition

Successful companies of the future will learn from Big Data and accept its findings as a check on managers' cognitive limitations, biases and the like. But at the same time, the firms cannot blindly accept its answers, but integrate Big Data into human values of justice, decency and common sense. People must remain masters of the technology, not its servants, for the data is always just a simulacrum of reality, not the real thing. Companies of the future must embrace data alongside a healthy respect for its limitations.

Big Data and the Future of Business

Part I: More

The basis of commercial enterprise is information. Indeed, some of the earliest forms of writing and accounting come from Sumerian merchants around 8,000 BC, who used small clay beads to denote goods for trade and later kept written records of transactions. So when we look at the role of data today, it is easy to say that not much has changed. We may collect, store and use more information—but the nature of data and its importance isn't much different. In this view, Big Data is just a fancy term to describe how society can harness more data than ever, but it doesn't alter the timeless fundamentals of commerce from antiquity to today.

This view, however, would be terribly wrong. For lots of areas of life, when one changes the amount, one changes the form. For example, no one would suggest that because symbols had been pressed into clay tablets, and then words formed and written with ink on scrolls, that the printing press wasn't a major revolution when it was developed around 1450. Yes, there had been words and books before, and yes there were now more words and more books. But it wasn't the same. More wasn't just more: more was different.

The effects of the printing press were the dramatic increase in written materials and the decline in cost of producing them. It was so monumental that the era of "more words" was responsible for sweeping changes. It diluted the authority of the church and the power of monarchies; it gave rise to mass literacy, democracy, capitalism, and a society based on knowledge as an ingredient of labor, rather than just muscle.

Today, the notion of written material—"the book"—changes again, when we see digital tablet computers like the iPad that can store all the books in

a major university library in a single device. And it can search it, index it, and allow portions to be easily copied and shared instantaneously. Here too, more isn't just more. More is new. More is better. More is different.

So much for words. Now, think of communications. Society was able to send messages long distancesin the past. Carrier pigeons were used in ancient Rome. To communicate with We have more information than ever. The change in scale leads to a change in state. The quantitative shift leads to a qualitative shift

his officers, Genghis Khan created relay posts for carrier pigeons throughout Asia and parts of Eastern Europe. In business, in the 1800s the Rothschild banking family sent their messages by pigeon, as did the market news service Reuters.

But at the dawn of the telegraph, no one could possibly claim that the wires and electric pulses were just an improved version of carrier pigeons. More was different. And then with the telephone: the greater communications, lower cost, and increased ease weren't just more of the same. Likewise, radio. Today, the internet is so fundamentally different than carrier pigeons that it seems ludicrous to compare the two. But that just underscores the degree to which more isn't just more; more is new, better, and different.

Like words and communications, so too data. We have more information than ever. But its importance is not that we can do more of what we already do, or know more about what we already examine. Rather, the change in scale leads to a change in state. The quantitative shift leads to a qualitative shift. By having more data, we can fundamentally do new things—things that we couldn't achieve when we only had lesser amounts.

In fact, we are just at the outset of learning what those things are, since we have always self-censored our imagination about what is possible with data. We did this, unawares, because we could never contemplate the notion of having so much of it around, since we had no idea it would become so easy and inexpensive to collect, store, process, and share. On what basis could we have extrapolated to divine this?

The wisest man with an abacus probably could never imagine the mechanical calculator with dials into the billions. The savant working those dials probably could scarcely imagine the electronic computer. And even once the transistor was invented several years after the first computers, it would have been hard for all but the most visionary engineer to fathom the pace of Moore's Law. As a principle of the digital age, it states that the number of transistors on a chip doubles about every two years, which has meant exponential reductions in cost and increases in power over time.

These changes in the degree to which society can collect and interact with information have had profound effects on how we understood the economy. The very idea of an economy is a relatively recent concept. When the classical economists emerged in Britain in the mid 1700s, their discipline was called political philosophy; the term economics only emerged later. Its veritable founding father, Adam Smith, was a moral philosopher whose major work before *The Wealth of Nations* was *The Theory of Moral Sentiments*.

It is easy to read passages from the classical economists and be led to appreciate the degree to which they were living in an observational and proseladen world, where commercial affairs were described with the majesty of words rather than the nakedness of numbers—a world of ideas mostly free of data. But this would be incorrect. In fact, Smith's *Wealth of Nations* is teaming with page after page of wheat yields. The earliest thinkers on the economy in the 1700s relied on data significantly to form their ideas.

Yet when it came time to define the factors of production, they identified three: land, labor, and capital. They did not include "information" as a distinct component, even if Smith and others wrote eloquently on how markets rely on information. It is easy to understand why they excluded it. At the time, it was so blindly hard to collect, store and use information that the idea it could be a raw material of business in and of itself would have sounded

These changes in the degree to which society can collect information have had profound effects on how we understood the economy preposterous. After all, the data would have had to be recorded by a person with a feathery quill pen on stiff parchment. It was expensive and cumbersome to handle and use information. Note that at this time, even basic statistics had yet to be invented. So even if one had the data, there wasn't much one could do with it.

Obviously, the situation is totally different today. Of course there are still limitations on what one can get and do with data. But most of our assumptions about the cost of collecting and the difficulty of

processing data need to be completely overturned. We still live with a "scarcity" mindset, like old people who hurry to the phone and keep the conversation short because a relative is calling "long distance"—a legacy behavior from the

days of expensive phone calls before market liberalization and new technologies would change the cost of telecoms forever.

And our institutions are still founded on the idea of information scarcity and high cost. Our airplane flight recorders maintain only a tiny amount of data, just several hours' worth of sparse mechanical and cockpit information—a legacy of the era in which they were designed. The recovery signal is weak and the battery is short, about 30 days. The world is now on track to fix these things after the tragedy of Malaysia Airlines flight MH370 that went missing in March 2014.

Yet the "black box" approach could help society in numerous ways: for instance, installing them on police vehicles and onto officers would help courts settle charges of police aggression versus the legitimate use of force. But only few places use them. Likewise, black boxes could enter operating rooms to help surgeons learn from mistakes, help patients harmed by negligence receive fair compensation, or prove that doctors performed flawlessly.

Yet doctors fear that it will open the door to a tsunami of malpractice suits, so have resisted their introduction. And neither the police nor doctors are wrong to hold their quasi anti-data views: it takes time for society to come to terms with how to accept and integrate a new technology and to develop the new culture that it requires. We are only just now getting comfortable with computers a half-century after their mainstream introduction.

In this regard, the experience of social media is instructive. In the critically acclaimed book *Delete: The Virtue of Forgetting in the Digital Age*,¹ Viktor Mayer-Schönberger of Oxford University (and my co-author of two books on Big Data) relates horrendous anecdotes of people denied jobs because of things like a photo of revelry that appeared years earlier on the job candidate's Facebook page. It highlighted the degree to which hiring managers hadn't recalibrated their practices for a world in which our past is ever-present online, and one's juvenile antics need to be "discounted" in a way that they never needed to before.

It will take a while for society to change practices and attitudes to find a reasonable way to bring the technology into our lives and institutions and our values

Likewise, in the Big Data world, many things will be passively recorded just because they exist or they happen. It will take a while for society to figure out how to manage this, and change practices and attitudes to find a reasonable way to bring the technology into our lives and institutions and our values.

Importantly, this tension—between what the technology is capable of and our attitudes and rules in which it exists—marks one of the main frictions the American political establishment has had to grapple with regarding the

Snowden disclosures on mass surveillance by America's National Security Agency. The inherent tension is this: the law was designed for an era when collecting and analyzing data was hard and costly, so embodies those presumptions. Once the same practices became easy and cheap, such as reviewing telephone metadata, activities that might have been considered impossible or at least exceptionally rare in the 1970s when the laws were codified could be considered commonplace in June 2013 when they were made public.

From the view of privacy advocates, the NSA mass surveillance activities were never authorized in law. From the NSA's point of view, the programs were just scaled-up versions of what the law does indeed allow. Shouldn't a security agency avail itself of the same modern tools that its adversaries are using to harm it?, goes the reasoning. The critics retort: get legal approval then, if you want those powers and believe the public will accept a dragnet.

Sadly, the American political system has yet to have a responsible and mature debate on these matters in order to find common ground. Although none of this analysis exonerates any activities, it perhaps takes a step forward in explaining them. Here again, we turn back the central motif of Big Data. More isn't just more. More is new. More is better. More is different.

No area of human endeavor or industrial sector will be immune from the incredible shakeup that is about to happen as Big Data ploughs through society, politics, and business. Man shapes his tools. And his tools shape him.

Part II: Different

The basis of commercial enterprise is information. That has not changed. Thus was it for Sumerian merchants many millennia ago, and so was it a mere century ago when Frederick Taylor performed his time-motion studies in American businesses.

Naysayers may feel that today's talk of Big Data is just a continuation of the past, but they are as wrong as if they were to claim that a tablet computer isn't fundamentally different from a stone tablet, or the web is just a continuation of the carrier pigeon, or an abacus similar to a supercomputer. It wouldn't be 100% wrong, but it would still be so preponderantly wrong as to be un-useful and a distraction.

The point of Big Data is that we can do novel things. One of the most promising ways the data is being put to use is in an area called "machine learning." It is a branch of artificial intelligence, which is a branch of computer science—but with a healthy dose of math. The idea, simply, is to throw a lot of data at a computer and have it identify patterns that humans wouldn't see, or make decisions based on probabilities at a scale that humans can do well but machines couldn't until now, or perhaps someday at a scale that humans can never attain. It's basically a way of getting a computer to do things

not by explicitly teaching it what to do, but having the machine figure things out for itself based on massive quantities of information.

Its origins are fairly recent. Though it was initially conceived in the 1950s, the technique didn't work very well for real-world applications. So people thought it was a failure. But an intellectual and technical revolution has taken place in just the past decade, as researchers have come up with lots of promising achievements using the technique. What had been missing before was that there wasn't enough data. Now that there is, the method works. Today, machine learning is the basis of everything from search engines, online product recommendations, computer language translation, and voice recognition, among many other things.

To understand what machine learning is, it is useful to appreciate how it came to be. In the 1950s a computer programmer at IBM named Arthur Samuel programmed a computer to play the board game checkers. But the game wouldn't be much fun. He'd win, because the machine only knew what a legal move was. Arthur Samuel knew strategy. So he wrote a clever subprogram that, at every move, scored the probability that a given board configuration would lead to a winning game versus a losing game.

Again, a match between man and machine wouldn't be very good—the system was too embryonic. But then Samuel left the machine to play itself. By playing itself, it was collecting more data. By collecting more data, it improved the accuracy of its predictions. Then Arthur Samuel played the computer, and lost. And lost. Man had created a machine that exceeded his own ability in the task that he had taught it.

So how do we have self-driving cars? Is the software industry any better at enshrining all the rules of the road into code? No. More computer memory? No. Faster processors? No. Smarter algorithms? No. Cheaper chips? No. All these things helped. But what really ushered in the innovation is that techies have changed the nature of the problem.

It's been turned into a data problem: instead of trying to teach the car how to drive—which is hard to do; the world is a complex place—the vehicle collects all the data around it, and tries to figure it out. It figures out that there is a traffic light; that the traffic light is red and not green; that this means the car must come to a stop. The vehicle might make a thousand predictions a second. The result is that it can drive itself. More data hasn't meant just more. More data produced different.

The idea of machine learning has led to some spooky findings that seem to challenge the primacy of human beings as the fount of understanding in the world. In a study in 2011, researchers at Stanford University² fed a machine-learning algorithm thousands of samples of cancerous breast cells and the patients' survival rates, and asked the computer to identify the telltale signs that best predict that a given biopsy will be severely cancerous.

And sure enough, the computer was able to come back with eleven traits that best predict that a biopsy of breast cells is highly cancerous. The nub? The medical literature only knew of *eight* of them. Three of the traits were ones that pathologists didn't know to look for.

Again, the researchers didn't tell the computer what to analyze. They simply gave the computer the cell samples, their general characteristics, and data on patient survival rates. (This one lived for another fifteen years; this one died eleven months later.) The computer found the obvious things. But it also spottedthe nonobvious things: disease signatures that people didn't see, because it was naked to the human eye. But it was spotted by an algorithm. Machine learning works because the computer is fed lots of data—more information than any human being could digest in a lifetime, or instantly remember.

In this instance, though, the computer outperformed the humans. It spotted signs that specialists did not. This allows for more accurate diagnoses. Moreover, because it is a computer, it can do these things at scale. So far, Big Data's "more" has not just been more of the same, it has been "better." But does this constitute "new" and "different" too? Yes.

Consider: by employing this approach at scale, we might be able to read biopsies once a day, every day, on an entire population—not just once or several times in a lifetime. In so doing, we may be able to spot what cancer looks like at its earliest stages, so we can treat it with the simplest, most effective, and least expensive intervention—a win for the patient, a win for society, and a win for government healthcare budgets that pay for it.

How is it new? Keep in mind, the computer did not just improve the accuracy of the diagnoses by adding new signals. It also in effect made a scientific discovery. (In this case, the three traits of severe cancer previously unknown

were the relationships among cells in cellular material called stroma, not just features within the cells themselves.) The computer produced a finding that eluded people, and which advances the state of human understanding.

What does it mean to have more data? A powerful example comes from Manolis Kellis, a genetic researcher at the Broad Institute in Cambridge, Massachusetts. As a White House report on Big Data in May 2014 noted: "A large number of genetic datasets makes the critical difference in

Big Data's "more"
has not just been
more of the same,
it has been "better."
But does this
constitute "new"
and "different" too?

identifying the meaningful genetic variant for a disease. In this research, a genetic variant related to schizophrenia was not detectable when analyzed in 3,500 cases, and was only weakly identifiable using 10,000 cases, but was suddenly statistically significant with 35,000 cases." As Kellis explained: "There is an inflection point at which everything changes."

The medical industry offers another powerful example of how Big Data is poised to reshape business. Healthcare makes for rich examples because it already has a lot of data, yet it is rather behind the times in using it relative to its great potential. So some of the most impressive wins have begun to happen in the area of healthcare, even though restrictive privacy laws risk hindering progress.

Consider the issue of how to spot an adverse drug interaction; that is, a case when a person takes two different drugs that are safe and effective on their own, but when taken together produce a dangerous side-effect. With tens of

The medical industry
offers a powerful
example of how Big Data
is poised to reshape
business, even though
restrictive privacy laws risk
hindering progress

thousands of drugs on the market, it is a hard problem to tackle since it is impossible to test all drugs together. In 2013 Microsoft Research and several US universities came up with an ingenious approach to identify these instances: by analyzing search queries.⁴

The researchers produced a list of eighty terms associated with symptoms for a known ailment, hyperglycemia (such as "high blood

sugar" or "blurry vision"). Then, they analyzed whether people searched for one drug paroxetine (an antidepressant) and/or another drug, pravastatin (which lowers cholesterol). After analyzing a staggering 82 million searches over several months in 2010, the researchers struck gold.

Searches for only the symptoms but neither of the drugs were extremely low, less than 1%; background noise. People who searched for the symptoms and one drug alone came to 4%; the symptoms and the other drug alone was 5%. But people who searched for the symptoms and both drugs came to a startling 10%. In other words, people were more than twice as likely to be typing certain medical symptoms into a search engine if they were also looking for both drugs than for just one or the other.

The finding is powerful. But it is not a smoking gun. The police cannot storm the pharmaceutical executives' homes and haul them away. It is just a correlation; it says nothing about causation. However, the results are significant, with profound meaning for business and corporate value. This adverse drug interaction wasn't known before; it wasn't on the label. It hadn't been part of the medical study or its approval process. It was uncovered by analyzing old search queries—again, some 82 million of them.

The value of this data is immense. If you are a patient, you need to know this information. If you are a doctor, you want this information. If you are a health insurance provider, you especially want it. And if you are a drug regulator, you absolutely want it. And if you are Microsoft, perhaps you should

think about establishing a division to license the data as a way to develop a new revenue stream, not just earn income from the ads next to the search results.

This new world of data, and how companies can harness it, bumps up against two areas of public policy and regulation. The first is employment. At the outset, business leaders see the need for new sorts of workers in the labor force—the great age of the data scientist. Management consultants issue dire warnings about a shortage. Universities are gearing up to fill that demand. But all this is very myopic thinking. Over the medium to long term, Big Data is going to steal our jobs. We can expect a wave of structural unemployment to spring from the technology.

This is because Big Data and algorithms challenge white-collar knowledge workers in the twenty-first century in the same way that factory automation and the assembly line eroded blue-collar labor in the nineteenth and twentieth centuries. Then it was muscle that was seen as a commodity and machines could perform better than people. In the future, it will be our minds that are shown to be weaker than the machine. A study by researchers at Oxford University⁵ predicts that as much as 47% of work that is done today in the United States is at risk of being taken over by computerization.

Consider the example of the pathologist who is no longer needed because a machine-learning algorithm can read cancer biopsies more accurately, faster, and more cheaply. Pathologists typically have medical degrees. They buy houses. They pay taxes. They vote. They coach their children's football teams on the weekends. In short, they are stakeholders in society. And they—and a whole class of professionals like them—are going to see their jobs completely transformed or perhaps utterly eliminated.

A study by researchers at Oxford University predicts that as much as 47% of work that is done today in the United States is at risk of being taken over by computerization

The benefit is that Big Data will bring about great things in society. The risk is that we all become yoga instructors and baristas to a small group of millionaire computer-scientists. We like to think that technology leads to job creation, even if it comes after a temporary period of dislocation. And that was certainly true for the disruption that took place in our frame of reference, the Industrial Revolution. Then, it was machines that replaced artisanal labor. Factories sprung up in cities and poor, uneducated farm hands could—once labor laws and public education emerged—improve their lives and enjoy social mobility. To be sure, it was a devastating period of dislocation, but it eventually led to better livelihoods.

Yet this optimistic outlook ignores the fact that there are some jobs that go away and simply never come back. As the American Nobel Prize-winning economist Wassily Leontief observed, the Industrial Revolution wasn't very good if you were a horse.⁶ That is to say, once tractors were introduced in

farming and automobiles replaced carriages, the need for horses in the economy basically ended. One sees the traces of that shift today, in the former stables throughout London's posh West End that have been converted into fancy mews houses.

The upheavals of the Industrial Revolution created political revolutions and gave rise to entirely new economic philosophies and political movements like Marxism. It is not too much of an intellectual stretch to predict that there will be new political philosophies and social movements built up around Big Data,

Big Data will change business, and business will change society. The hope is that the benefits outweigh the drawbacks

robots, computers, and the internet, and their effect on the economy and representative democracy. Recent debates over income inequality and the occupy movement seem to point in that direction.

The second policy area is privacy. Of course, privacy was a problem in a "small data" era. It will be a problem in the Big Data era too. At first glance, it may not fundamentally look like a different problem, but only the same problem at a greater scale. But here too, more is different. The nature of securing personal information changes when the potential privacy harm does not happen once a day or once an hour but a thousand times a second. Or, when the act of collecting data does not happen by overt, active means but invisibly and passively, as a byproduct of another service.

For example, websites in Europe are compelled to inform web visitors that they collect "cookies" used to identify people visiting the sites. Such a requirement sounds reasonable on the surface. But what happens when every light fixture in a building is identifying if there is a person in the room on the grounds of security and protection (i.e., in a fire, rescuers know where to go). And the software, at near-zero marginal cost, is sophisticated enough to identify who those people are, based on their image, gait, or perhaps pulse. It is hard to imagine how classic privacy law would handle that world; how a person who feels wronged would take action—or even be aware of the situation.

It gets worse. A basis of privacy law around the world is the principle, enshrined by the OECD privacy guidelines, that an entity discards the data once its primary purpose has been fulfilled. But the whole point of Big Data is that one ought to save the data forever since one can never know today all the valuable uses to which the data can be put tomorrow. Were Microsoft to have deleted its old search queries from 2010, it never would have been able to identify the adverse drug interaction between paroxetine and pravastatin in 2013.

So just as a theme of Big Data is that more isn't just more, but more is new, better, and different, so too modern businesses will need regulators who understand that the rules that govern Big Data cannot just be more—more of the same. In fact, the rules today do a poor job of protecting privacy, so simply heading forward with more of a mediocre policy makes little sense. Instead, Big Data businesses cry out for regulations that are new, better, and different.

Big Data will change business, and business will change society. The hope is that the benefits outweigh the drawbacks, but that is mostly a hope. The reality is that all this is very new, and we as a society are not very good at handling all the data that we can now collect. It was only as recently as the 1893 Chicago World's Fair that a gold medal was won by the invention of the vertical filing cabinet, a then brilliant solution to the problem of the storage and retrieval of paper documents—an era when the stream of information swamped business; the "beta version" of Big Data in corporate life.

What is clear is that we cannot extrapolate to foresee the future. Technology surprises us, just as it would an ancient man with an abacus looking upon an iPhone. What is certain is that more will not be more. It will be different.

Notes

1

V. Mayer-Schönberger, *Delete:* The Virtue of Forgetting in the Digital Age (Princeton: Princeton University Press, 2009).

2

A.H. Beck et al., "Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival," *Science Translational Medicine*, 3.108 (2011) http://stm.sciencemag.org/content/3/108/108ra113.full.pdf

3

"Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President of the United States, May 2014 http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

4

R.W. White et al., "Web-scale Pharmacovigilance: Listening to Signals from the Crowd," Journal of the American Medical Informatics Association, 20.3 (May 2013), 404-8 http://www.ncbi.nlm.nih.gov/pubmed/23467469

5

C.B. Frey and M.A. Osborne, "The Future of Employment: How Susceptible Are Jobs to Computerisation?" Oxford University, September 17, 2013 http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

6



OpenMind Channel



Article

- Big Data and the Future of Business

About the Author

Kenneth Cukier

Related Articles

- From Deconstruction to Big Data: How Technology is Reshaping the Corporation
 - Who Owns Big Data?
 - The Internet and Business

Read the Full Book

- Reinventing the Company in the Digital Age

Other Books













Share









