Nobody would be shocked to hear that technology can change radically in twenty-five years. All you have to do is look back twenty-five years from today to see the distance we have traveled. In 1989, cell phones were a new phenomenon and could only make static-filled voice calls, and almost no one bought anything online. In 2013, a billion smartphones were sold. They have become the door to a vast parallel landscape of commercial, personal, and societal life. Among millennials, these devices are used for only the occasional phone call.

Over the coming twenty-five years, from 2015 to 2040, technology will change more radically than ever before. Information technology is just beginning a historic transformation. We’re about to move to computers that work differently from the way computers have worked since the 1950s. They won’t be programmed—they’ll learn on their own, and through data they will have an awareness of the world and events around them. More important, they will enable a new wave of digital services that fill gaps in our perception, improve our judgment, and magnify our cognition.

Data is about to become a vital commodity, a raw material—the next natural resource.
We’re about to evolve beyond the information age to enter the age of insight. In 2040, value will come from readily surfacing connections, insights, and answers buried deep in a universe of data—natural and human-generated—as limitless, emergent, and difficult to penetrate as the physical universe.

The shift will gain momentum, so that by 2040 technology will play a role we can barely comprehend today. (Who in 1990 could have grasped that something called Twitter might help organize rebels in Libya so they could overthrow a dictator?) In the coming era, technology will allow us to see more clearly using data and to think completely differently about age-old problems both large and small.

The economic implications will be significant. Among them will be a changed purpose for the corporation. In the 1930s, the economist Ronald Coase famously argued that corporations existed to reduce transaction costs—the costs of putting together the right set of suppliers and people and facilities to produce something. That was the very reason for the rise of the large-scale, vertically integrated company. The rise of programmable computers supported that. Over the past twenty-five years, computing and networks have crushed transaction costs, speeding the dismantling of vertical companies and blowing up business models of entire industries.

If transaction costs no longer matter so much, why will next-generation enterprises exist?

The answer is answers. We will dramatically lower the cost and reduce the time to get answers or the insights that lead to them.

Already in 2015, we see a glimmer of this in the world of online search, social networks, and retailing. Their value comes from knowing things about customers and the market. It’s why Facebook’s insights are perceived as more valuable to the market than the massive operations of companies that produce more traditional products and services.

Over the next twenty-five years, this will no longer be the realm of the Internet start-up. Every industry and every profession will transform itself through data.

To explain the implications of the age of insight in a tangible way, I’ll go deeper into three other industries where change is afoot as answers and insights are becoming more affordable and readily available.
Health

In twenty-five years, it will seem almost medieval that you would take a pill or any treatment without already knowing exactly what it will do to your body. We will have systems that essentially build a virtual model of you, based on data ranging from your DNA, your family history, and medical records to your eating habits, sleep patterns, and work. Before taking any new medication, your doctor will be able to ask your virtual self: what will it do to you?

You will benefit from everything we know about the right treatment before you ingest anything—not after.

Education

In twenty-five years, it will seem mind-boggling that education was based mostly on time served—four years of college gets you a bachelor’s degree, another two gets you an M.B.A. Instead, technology will make it possible to build a lifelong education around every student individually. The technology will be smart, persistent, and aware enough to act as a tutor and adviser for every person, always knowing what the student is learning or not learning—and what he or she needs to study to reach specific goals.

A student will have personalized support, and educators will have a more nuanced view into whether a student is learning before that student takes any tests.

Management

In twenty-five years, no decision maker will walk into a meeting without his or her cognitive apps, or “cogs”—much as he or she would look to spreadsheets and reports today. Cogs will ingest millions of pages of documents related to the decision maker’s domain—more information than any human could ever get through in a lifetime. Cogs could act as an executive assistant or a sophisticated decision support system offering appropriate material, constructing scenarios, and helping model alternative arguments and evidence relevant to such complex questions as
whether a firm should acquire, organically develop, or partner to enter a new business segment.

Cogs will exist in the cloud, interact easily as in dialogue, and be accessible through anything from a mobile device to a smart tabletop or a tiny embedded earpiece.

They will help managers understand the possible outcomes of decisions before those decisions get made.

Cognitive Computers

Technology is approaching a breaking point.

Torrents of data are flooding systems. Billions of people around the world are on the Internet, making transactions, posting on social networks, listening to music, and doing work—all of it generating data. By 2013, there were more than six billion mobile phone subscriptions globally, more than two billion of them with broadband capability adding more data.\(^2\)

On top of that, the physical world is becoming packed with sensors that digitize life—water flow in rivers, traffic on city streets, and the movement of goods and people and animals. There are already about a trillion connected devices generating data.

Computers won’t be able to keep up—at least not the programmable computers that have dominated for fifty to sixty years. Even if we can make those computers fast enough to crunch all that data, the software will be too inflexible to deal with the volume, speed, diversity, and uncertain veracity of it all. As a result we will often get answers after they’re no longer valuable. We are like drivers of a car in a snowstorm who only have a sense of what we’re not seeing through the windshield.

Fortunately, we’re entering a new technology era. Let me explain.

The first era of computing—the tabulating era—began in the nineteenth century and continued into the 1940s. Tabulating machines were made up of a series of mechanical switches and were essentially elaborate adding devices, helping tally up a national census or a company’s payroll. This was an enormous advance, and it made possible the rise of the modern bureaucratic nation-state and the modern corporation.

The second era emerged in the 1940s—the programmable computing era. At first these machines relied on vacuum tubes, then transistors,
which got smaller and smaller and packed into microprocessors. The ma-
chines operate on an architecture described by mathematician John von 
Neumann: software programs tell the machines to carry out calculations 
in a series of steps, pulling in data stored separately from the programs. 
This is how most computers work today. While computers have gotten 
faster and faster, the von Neumann architecture means that all the work 
has to be lined up in the right steps as it flows through a machine’s pro-
cessors.

This, too, was a major advance in technology, business, and society. 
It made possible the multinational corporation—providing a common 
platform for everything from back office accounting to supply chains, 
ATMs, and travel reservation systems.

When the data is highly structured in understandable formats, such 
as tables in a database, programmable architecture does the job very well. 
But today, structured data is becoming a smaller and smaller fraction 
of what we are capturing. Most of what is known as “big data” looks 
more like life itself—full of images, sounds, patterns, language, and un-
certainty.

The new era of computing, just emerging, will be one of cognitive sys-
tems. Instead of being programmed by humans, computers will “learn” 
from their interactions with data and people, and they will adapt, like 
the brain, to many parallel connections among data streams at once, gen-
erating patterns, observations, and insights. These systems will draw on 
unimaginable amounts of information and also take into account what 
they “know” about the people asking the question and the environment 
around them. The machines will, for the first time, interact with us in 
ways that are almost human, bringing order to the chaos of free-flowing 
and diverse information—and do so in something like real time.

IBM’s Watson represents a starting point for a cognitive system. It 
shows what this kind of technology can do today, which helps us im-
agine what it will do tomorrow.

The original Watson, built on software called DeepQA and running 
on a room–size computer, beat two human champions on the TV game 
show Jeopardy! in 2011. The system ingested hundreds of millions of 
pages’ worth of material, from poems to Wikipedia entries to textbooks. 
The algorithms in the software broke down the material into categories 
and learned how words are used in various contexts. In this way, Watson
started “learning” from the data rather than processing it through a set of instructions. In a sense, it became expert in a domain—the game of *Jeopardy!* It arrived at answers by coming up with numerous possibilities and ranking the certainty of each. Not only could Watson arrive at answers, it could leave a trail of breadcrumbs behind, showing how it came to the answers. And it became better through experience.

In the years since, Watson has been exploring applications in many domains from retail to financial services. But health care, where it has become impossible for physicians and even researchers to keep up with information, has become our grand challenge. Medical literature doubles in size every few years, yet doctors say they have little time to read medical journals each month. For physicians, incorporating hundreds of thousands of articles into their practice and applying them to patient care—together with seemingly unlimited patient-specific information, as we plumb the possibilities of genomics—is a significant challenge.

Today, the first of these systems are being “trained” with data in real environments, essentially advising professionals by bringing an immense store of knowledge to bear on a problem. These machines can “understand” spoken questions in natural language and logically arrive at possible answers. Watson in three seconds can sift through information equivalent to roughly two hundred million pages of data in natural language and analyze it to help a doctor identify the most likely diagnosis and treatment options in complex cases.

But there is much work still to be done in many domains, which will contribute to cognitive computing—from voice-based interaction with smartphones to deep exploration in “artificial intelligence.” At the Massachusetts Institute of Technology, for instance, Rosalind Picard is studying the role emotions play in making more efficient decisions, working toward eventually building helpful emotions into cognitive machines. And in Silicon Valley, Jeff Hawkins, who invented the Palm computer, runs Numenta, which is developing software that can learn the way the brain learns.

The pace of development is breathtaking, and it’s impossible to predict the future with confidence, based on the work we do today. But our experience from working with clients and universities at IBM Research tells me that by 2040, we will have exponentially greater computing power and a new cognitive software model that will be trained with data. The systems will gorge themselves on nearly infinite data.
Then the question is: what will we do with such magnificent technology?

Modeling a Human

Soon after computers were invented, scientists realized the machines could be used to simulate aspects of the physical world as a way to get faster or more cost-effective answers to difficult questions. During World War II, von Neumann and another mathematician, Stanislaw Ulam, developed a simulation method—the Monte Carlo method—that helped them understand the behavior of neutrons. That was pretty impressive, considering the day’s computers could do only about four hundred calculations per second. The chip in a modern digital watch would put that to shame.

By the 1950s, scientists in the United States and Sweden began using computers to model, simulate, and predict weather, giving birth to the sophisticated meteorological predictions behind every TV weather reporter’s forecasts today. Modeling exploded in the 1960s, applied to everything from inventory control to industrial systems design to the Apollo moon missions. In the 2010s, the world couldn’t operate without simulations. Computer models test nuclear bombs without exploding any real ones, govern the traffic in the skies, and even keep Segway scooters from tipping over.

The Segway helps illustrate an interesting turning point in modeling. Modeling uses data to represent something that happens in the real world. Usually, it requires a lot of data, which means the simulation requires complex computations that, typically, take much longer than the actual event. When simulating a nuclear explosion, for instance, the simulation doesn’t happen as fast as a real explosion. And that’s OK for research. There’s enough time to run the computations and study the answers.

However, magic happens when you can compute a model at the speed of nature. Inside a Segway, a small computer runs a model of the physics involved in balancing the scooter on its two wheels. It can do this as fast as the physics happen in nature, constantly predicting exactly what the Segway’s machinery has to do to counter weight shifts and keep the scooter upright. The time to right answers is crucial—a split-second too late and the Segway’s rider would be flipped onto the pavement.
As the next twenty-five years progress, technology will increasingly be able to compute accurate models of complex events in real time. One of the more stunning outcomes is that we’ll all be able to access a real-time model of our own bodies.

Scientists at Lawrence Livermore National Laboratory are working with IBM to start down that path. The project, called Cardoid, is aimed at modeling the human heart. If we can accurately simulate the behavior of a heart down to the cellular level, we’ll be able to test surgical procedures and medications on computers instead of on people. But the heart is so complex that until recently it took forty-five minutes of computation to simulate a single heartbeat. At that rate, the time to test a single drug would be years or decades—not a helpful span of time to get to a right answer.

Livermore tackled the problem by using a supercomputer with 1.5 million computing cores that can operate in parallel. By 2013, it could replicate nine heartbeats in one minute of calculations. It could model an hour of heart activity in about seven hours. That’s an astounding improvement.

The pace of development has taken off over just a couple of years, and it will not be long before a supercomputer will be able to model a heart in real time. Watching such a computer model react to a drug or surgical procedure will then be the same as watching a real heart. And you can see where this is heading. A heart model that today runs on a supercomputer in a research lab will soon be delivered as a service accessed through the cloud on a tablet or smartphone.

While the cost and computation time go down, the complexity, customization, and accuracy will go up. In ten to fifteen years, your doctor will be able to access a model of your heart. You won’t wait until you have heart trouble to use it. You might even access your model to experiment with the impact of a daily two-mile run or a low-fat diet.

Advanced computer science and medical science are merging in myriad ways. We’ll be able to marry such high-performance computing capabilities with the knowledge and learning power of cognitive systems like Watson to assist doctors in providing quick, detailed assessments of a procedure or a drug for a specific individual.

In parallel, DNA sequencing is getting faster and cheaper. Over the next decade, knowing your genetic makeup will become as common as knowing your blood type, another valuable form of data. Today organi-
izations such as the New York Genomic Center are already exploring the potential of cognitive systems to improve outcomes for the most aggressive cancers—those that spread because they mutate rapidly.

This is all about the time value of insight. Understanding a patient at the genomic level and finding precise treatments among thousands of possible combinations will need to occur in a window of maybe a few weeks or a month, otherwise there is no value to the patient. We should be able to do such analysis in hours or minutes.

By 2040, all this varied work will have begun to come together. Researchers will soon be able to economically and expediently explore the entire human body, collecting data that can inform simulations of particular brains, nervous systems, muscles, cells, proteins, and other body systems. We’ll have enough data and computing power to start to model the entire human body as a system.

And as with search engines and cell phones today, we will wonder how anyone lived without this capability.

**Personal Lifelong Tutor**

In the 1930s, in the worst of the Great Depression—at the same time that Ronald Coase was developing the ideas for which he would later win the Nobel Prize in Economics—a Columbia University professor named Ben Wood asked for a meeting with IBM’s longtime CEO, Thomas Watson Sr. Wood had been pioneering standardized testing as a way to measure students across schools and geographies. But scoring the tests by hand was a long and tedious process. Wood convinced Watson to donate IBM’s tabulating machines so Wood could greatly speed the test-scoring process and make standardized testing a reality. For the first time, machines could know what a student had learned.

Over the following eighty years, computer analysis of standardized tests has surfaced information that has been valuable for educational institutions. But because the information is not part of a true system—at least not one that senses and responds in real time—it has not mattered much to individual students’ actual learning or growth. It takes weeks or months to identify a student veering off track. Too often these insights are not fully grasped by the students or their support system when they have the most value.
The insights generated by a cognitive system could tell us not just what a student is learning but how he or she is best able to learn, and how fast. Such a system might then help the teacher tailor lessons and support specifically for that student. In fact, a smart system could become a lifelong educational adviser, guiding each student to courses, subjects, colleges, even jobs, that are a good match.

In 2015, a couple of developments are merging to help create this education technology. A very important one is the rise of massive open online courses (MOOCs)—digitized educational content widely available to anyone with an Internet connection. As more students use MOOCs, often as a supplement to classroom learning, the courses generate a tremendous amount of data about student behavior, such as how long students stay engaged with certain kinds of content or where and when they access content. This is the beginning of developing a more real-time data model of the way an individual student learns.

This is quite a flip in the world of education. Until now, students had to learn the courses; now we see courses beginning to “learn” the students.

In 2013, schools in Gwinnett County, Georgia—a district of 140,000 students—started working to analyze digital learning of individual students. The schools began to find, for instance, that the data can identify a student who is at risk of doing poorly before he or she takes a test. By intervening and helping, a student who might have failed a test and become discouraged would have a chance to correct course, get better grades, and have more positive feelings about school.

By 2040, this shift will have a more fundamental impact, altering the way we think about educational achievement in general. Learning today is mostly time based—students get a diploma for completing twelve years of school, then a degree for completing another four, and so on. By 2040, measures of achievement will be competency based. Students will get degrees or certifications that reflect not how long they have studied but what they know.

This, in turn, will change the role of the teacher. Instead of being a one-way delivery vehicle for information—or, even worse, a test proctor or disciplinarian—the teacher will be more like a tutor or mentor, valued as a source of wisdom and as a provider of individual counseling. The teacher will be able to draw on the system, which will “know” what the student knows.
Finally, because the classroom is no longer an arena of regimented socialization or one-way information transfer, it will become a more collaborative space where teachers help and advise individual students or small groups and where the group itself creates positive feedback loops of mutual knowledge and support. In 2040, the one-size-fits-all classroom lecture will seem as archaic as the twentieth-century television broadcast tower.

And because the system will “learn” along with you, it will not abandon you after college. If you want to go after a particular skill set or change careers, the digital assistant will help sustain focus on your objectives, the core requirements, and alternative pathways to proceed. It will act as a guide—an adviser and advocate for skills and learning. In fact, it will become the embodiment of “lifelong learning.” By 2040, students will be surprised to find out learning wasn’t always lifelong.

Cognitive Business Apps

At IBM’s Thomas J. Watson Research Center, we built what we call the Cognitive Environments Lab. It’s a way for us to work on the interactive environments of the future and to create cognitive applications that will help people make better decisions.

The entire room is lined with ultra-high-resolution screens. With the wave of a wand, a user can move data around the room. A natural language system recognizes spoken questions. This is the beginning of embedding cognitive computing in the environments we inhabit (whether our offices, our meeting rooms, our cars, or our homes) and utilizing such capability in our daily lives. These new cognitive environments will improve the work flow of experts and teams engaged in complex decision making.

By 2040, a manager may walk into a cognitive room at his or her company. The manager’s personal cognitive agent will interface with the room, allowing him or her to plug into and interact with all the elements of the system. The personal agent will connect with other agents and people and help the manager generate hypotheses, gather evidence for and against, and generate possible answers when they are most valuable—before taking action.

Since the tabulating era of computers, machines have helped managers make decisions. They’ve tallied up financial numbers, analyzed per-
formance, identified potential customers, and looked for trends in social media. Spreadsheets of old have given way to sophisticated analytics that can look for patterns no human would ever find. Hedge funds run detailed models of the global economy to spot risk and to try to know where to invest as early as possible. Computers have constantly marched toward providing enterprises with better answers and more detailed scenario models, in shorter time frames.

The cognitive computing era will take decision support to an entirely new level. The kinds of cognitive systems that will build models of the human body could certainly build models of a lot of things businesses care about. By 2040, a company might build a model of a target market and test a product before manufacturing it. A company could even build a model of itself, constantly learning from all the company’s trial and error, success and failure, generating ideas and solutions unbiased by any individual human perspective or groupthink and giving the CEO a way to try out decisions before implementing them.

As in education, a company’s cognitive system will be able to get to know individual employees, understanding their strengths and weaknesses and work patterns, intervening before failure, and helping guide their careers. By 2040, an employee’s human resources file will likely be a sophisticated model of that employee’s work and knowledge.

Ultimately, by 2040 we believe an executive will have a readily available, naturally interactive set of tools to model his or her business—a digital model always evolving as new information is drawn in. We call this emerging idea cognitive business apps, or “cogs.”

A personal cog will “learn” from our experience and even anticipate what we need. It will “learn” through interacting with other people, other cogs, and other data sources, becoming more helpful, capable, and precise over time. It will “understand” and model its user. The personal cog will interact with many other domain-specific cogs that specialize in market information, news feeds, supply chains, patents, and regulations, and in the organization itself. It will interact with these other cogs as needed to answer questions, explore strategies, and anticipate issues.

We picture a businessperson in 2040 walking into a meeting space in which cogs that reside on a phone or tablet can be deployed in the room. Meetings will have another dimension, whether in a cognitive room or out for coffee with a smartphone. Businesspeople will have interactive
access to data and expertise, along with a better understanding of the alternatives.

This will no doubt lead to new kinds of corporate structures and strategies, just as lowering transaction costs led to the development of the vertically integrated company in the mid-twentieth century.

It will also give rise to a new kind of executive. In a world of cogs and sophisticated computer models, the most successful decision makers will always be exploring. They will differentiate themselves not by knowing all the answers, but by knowing the right questions to ask. It will be easy and economical to bring in new perspectives, to challenge intuition and raise evidence. You can imagine that the economics and speed of insights and answers will drive a new kind of creativity in questioning, a bias toward agile approaches and openness to continual transformation.

2040

It remains daunting to make predictions about technology twenty-five years out. But, extrapolating from what we know now, I can be quite sure that by 2040 we’ll be well into the era of cognitive systems that will fuel incredible innovation, just as the first computers and then the Internet did before them.

The technology will impact almost every facet of life and work. We’re quite sure the technology will completely change the way we approach health care, education, and management. We’ve already begun working on applying cognitive technology to those fields.

The real-time nature of these cognitive models will play out even more in purely digital domains. One such area is a very current concern, cybersecurity. In protecting people and assets, these systems watch over networks that operate at speeds limited often only by physics. In 2040, these systems will be targeted and even personal, adept at detecting patterns buried in billions of network log files and emerging models, developing, in a sense, “gut instincts,” so they can react to a threat in a split second—the way a human instinctively ducks when an object comes at him or her.

When scientists in our labs at IBM think about the coming era of answers, we can think of a wide variety of other ways cognitive computing could be applied. This is not about “perfect” knowledge—whatever that
might be. It’s about good-enough confidence with appropriate timeliness. This is what we mean by “the right answer”: the medical treatment that can halt a cancer before it metastasizes; the educational intervention that can set a young child’s learning on the most individually tailored path at a point when the child’s brain is still being formed; the business decision about an emergent market space that can anticipate and avoid “the innovator’s dilemma.”

But how exactly this will play out is, frankly, a bit of a challenge for a technologist. Perhaps it is better suited to the imaginations of science fiction writers, particularly as the application of technology becomes less and less linear. This era of systems will always be questioning, adapting, and changing, such that no model will be frozen in time. In fact, time will be a rich dimension in cognitive systems, as it is in the natural world. And the nature of prediction itself will change.

How will such cognitive systems and data models emerge in our lives? How will they be monetized? How will such a rich world of readily available insight and answers change the economy, society, and standards of living? Will it change how we actually think? Will it open a door to more fluid, imaginative, emergent—more fully intelligent—habits of mind? This goes beyond the familiar idea that technology throughout human history has freed us from lower-level physical or routine work, has freed us to apply ourselves to higher-order thinking. This suggests a change in what we mean by “higher-order thinking.”

Technology always affects the global economy and society. Cognitive technology will be no different. From where I sit, the time value of answers and insights will drive the most profound changes. We’re entering an age when raw information will be a resource, a commodity. The most valuable thing on earth will be knowing what to ask.

Notes

3. The Watson technology is named for IBM’s founder, Thomas J. Watson Sr., as well as being an allusion to Sherlock Holmes’s fictional sidekick, Dr. John Watson.