2

BUILDING LEARNING SYSTEMS

In November 2009, when IBM's Watson was under development for its showdown on *Jeopardy!*, the machine made one laughable mistake after another in test matches. In one particularly funny instance, it was prompted to identify what the "Al" in the company name Alcoa stands for. It fired back, "What is Al Capone?" Everybody in the room cracked up. The machine had confused the first two letters of aluminum with the name of a famous gangster.[1]

No harm was done. One of the main purposes of these training sessions was for Watson to make mistakes so that the scientists on the technology development team could tweak the algorithms used by the computer and prevent the same mistakes from happening again. It was a painstaking, manual process. But there was no way that the scientists could anticipate every dilemma that Watson might be confronted with and write a rule to deal with it, so, as they developed Watson, they realized they would need to invent a system that would enable Watson to learn on its own.

Their invention represented a major advance in the science of machine learning, a branch of artificial intelligence that focuses on building systems that learn from data. The field was first defined in 1959 by IBM scientist Arthur Samuels and has found plenty of uses over the years—including common applications such as optical character recognition and e-mail spam filters. Such systems are trained to recognize repeated patterns in words or shapes and to react in a certain way when they encounter them again.

Watson takes machine learning to a new level. In creating the technology for Watson, called DeepQA, which includes the learning capability, the developers provided the machine with a large corpus of unstructured information and the algorithms to extract knowledge from it. They trained the machine to identify categories of information, such as famous people, locations, and relationships, and also to parse language. Then they created a set of statistical methods for learning how words are used in different contexts. This combination of techniques allows Watson to learn from the data rather than just follow detailed instructions. In a sense, Watson learns the way people learn, by experiencing a lot of things and drawing inferences and lessons from those encounters.

Watson's learning abilities represent a major step forward in the new era of computing. Today, tech companies, corporations, and governments employ armies of software programmers to create specific applications to accomplish specific goals. Those programmers write millions of lines of detailed instructions telling computers what to do, following the linear and deterministic approach charted out by von Neumann at the dawn of the programmable computing era.

But this approach simply isn't sufficient anymore. We need to design and build a new class of computers that use Watson-style learning systems in order to help us deal with the complexities of the modern world. These systems will be able to learn from both structured and unstructured data, discover important correlations, create hypotheses for those correlations, and suggest actions that produce better outcomes. In addition, they'll have feedback loops where they measure the results of their work and modify the way they do things based on that feedback. Once people design and program them and train them, they'll do the rest—essentially reprogramming themselves as they interact and learn.

We're in the early stages of the evolution of learning systems, but it's possible to anticipate how things will develop—if not the exact timing and details of the technology. The first stage is static learning. The version of the Watson system that played on *Jeopardy!* is in this category, according to IBM Research executive Dario Gil, who headed a team that mapped out the progression. At this point, it's still necessary to assemble teams of technology and domain experts to provide a learning system with the rules and information it needs to perform a certain well-defined task—then train it to use the information. One example is the customer-service call center. Today, when consumers call companies for help with technology problems, representatives read from scripts and search rigid knowledge-management systems for answers. The process is too often frustrating for the consumer and expensive for the company. Now, companies are training Watson to magnify the capabilities of their representatives. The systems will ingest vast amounts of information that's useful in solving a specific set of problems, say, glitches involving smart phones. The cognitive system will learn from interactions, steadily improving the advice the representatives give to consumers. In addition, there will be a self-service version of Watson that helps consumers with their problems and concerns directly.

In a second stage, dynamic learning, machines will continually mine information from a variety of sources (e.g., textual, numerical, visual, sensory). Also, scientists will create a variety of modes that the machines can use to communicate, including spoken conversations. Dynamic learning systems will expand their hypothesis generation across multiple domains that frequently intersect with one another, such as economics, business, and social sciences.

How "smart" will learning systems get? It's possible that they'll be able to achieve a deep comprehension of natural language, images, and other sensory input. Already, programs exist that make it possible for machines to engage in scientific discovery, and, at the far end of the arc of time and imagination, they'll likely be able to take on complex problems that they identify themselves.[2] They'll read a host of scientific, social science, and technical publications; "watch" TV shows and read news reports; and perhaps even monitor environmental shifts like climate change. Years from now, when research scientists stroll into their offices they will greet systems that have access to vast knowledge resources and can use them to come up with new hypotheses.

JOURNEY OF DISCOVERY: THE PATH TO WATSON

The Watson project got its start in a surprising way. In the fall of 2004, IBM's head of computing systems software, Charles Lickel, traveled from his home in Tucson to spend the day with a small team he managed at an IBM facility in Poughkeepsie, New York. At the end of the workday, the team gathered at the nearby Sapore Steakhouse for dinner. They were bemused when, at seven p.m. sharp, many of the diners abruptly got up from their tables, rushed into the bar, and clustered excitedly around the TVs. One of Charles's guys explained that they were watching long-time champion Ken Jennings defend his title on *Jeopardy!*[3]

Charles hadn't followed *Jeopardy!* for years, but the scene made an impression on him. A few months later, research director Paul Horn asked his lieutenants to think up a high-profile project that the lab could take on that would demonstrate IBM's scientific and technological prowess. The company calls these its "grand challenges." The previous grand challenge had been a huge success: IBM's Deep Blue computer had beaten the world's top chess grand master in a highly publicized match in the mid-1990s. But a lot of time had passed since that victory.

During one of the brainstorming sessions aimed at picking the company's next grand challenge, Charles suggested building a computer that could compete on *Jeopardy!*

IBM has long used man-versus-machine games to motivate scientists, focus research, and engage the public. In the early 1960s, IBM researcher Arthur Samuel, the AI pioneer, created one of the first computer programs capable of learning when he wrote a checkers-playing program designed to run on the 701, IBM's first commercial computer. Samuel challenged one of the top U.S. checkers champions to a match—and won. IBM researcher Gerry Tesauro in the late 1980s developed a program called TD-Gammon, which used a technique called temporal difference learning to teach itself how to play backgammon. It was competitive in matches with some of the world's top backgammon players.[4]

Later, IBM turned the contests into true spectator sports. The 1997 match between Deep Blue and the reigning world chess champion, Garry Kasparov, was streamed live on a website, and millions of chess enthusiasts watched every move. The Deep Blue program didn't require much learning since chess is such a highly structured game, but, according to Murray Campbell, one of the researchers on the Deep Blue team, the match demonstrated that a combination of clever engineering and sophisticated algorithms could rival the performance of a top human expert in a specific domain of human achievement. "It made everybody understand in a clear way that problems previously considered too hard for a computer could now be tackled," Murray says.[5]

Of course, while IBM has played an important role in the evolution of artificial intelligence and learning systems, many other scientists and organizations have made major contributions. John McCarthy, a professor at Stanford University, coined the term "artificial intelligence" in 1955, and Marvin Minsky, a professor at MIT, has produced a long series of advances in the field since the 1950s. He's now focusing on giving machines the ability to perform humanlike commonsense reasoning.[6] Today, Andrew Ng of Stanford is leading a team of scientists in an attempt to create algorithms that can learn based on principles that the brain might also employ.

The field of artificial intelligence has advanced in starts and stops. Periods of soaring optimism have been followed by so-called AI winters, when seemingly promising avenues of research failed to produce the anticipated results. Put simply, this is hard stuff. So it's no surprise that when Charles Lickel proposed *Jeopardy!* as the next grand challenge, his suggestion was met initially with reactions ranging from skepticism to outright derision. But he quickly won Paul Horn's support. Paul thought the project could be very exciting—both to computer scientists and the public at large.[7] In mid-2006 Charles gave the go-ahead to researcher David Ferrucci, who was an enthusiastic evangelist for the project, to explore whether building a machine that could win on *Jeopardy!* was even plausible.

Progress was extremely slow at first. For several years, Dave had been leading a group of researchers who had produced good results with question-answering technologies. However, their performance gains had plateaued. So

Dave and a small team cast around for the right technology and strategy with which to approach the *Jeopardy!* challenge. By mid-2007, though, Dave was convinced the job could be done, and Charles and Paul gave the project their blessing.[8] However, there was a danger that the project would never actually be realized as Paul retired from IBM soon after and became senior vice provost for research at New York University. His replacement as head of IBM Research was one of the authors of this book, John Kelly. Dave was terrified that John would kill the project, and his fears were not assuaged when, at their first meeting, John expressed deep reservations. At the time, the system, then called Blue J, was only able to answer 30 percent of previous *Jeopardy!* questions accurately. John told them: "Guys, we can't put the IBM brand on TV with 30 percent accuracy."

Some of the team members were discouraged. But Dave told them that he was convinced they could achieve their goals. He said he was so committed to the project that he was willing to risk failure and public humiliation. And he asked each of the twelve scientists who were then on the team to fully commit themselves, as well. "Are you ready for this?" he asked.[9]

They were. Over the following weeks and months they gradually improved the technology to the point where, even though the gap was still huge, John was convinced they could build a machine that could compete at the highest levels on *Jeopardy!* He saw that the effort could boost IBM's reputation as an innovator and that the science would be transformative, opening up massive new opportunities for computing. He said: "I'll give you whatever you need in the way of resources, but, if we're going to put the IBM brand on national TV, we must win. *We must win*." That launched the intense effort that led ultimately to Watson's victory.

On January 14, 2011, the *Jeopardy!* contest was conducted on stage in the IBM Research auditorium. Top IBM executives, research scientists, and guests packed the room. Emotions ran high. John told the audience that he did not know if Watson would win, but he believed that the contest represented an important moment in the history of computing. He said, "It's not a matter of *if* a computer will one day win at *Jeopardy!*, but *when*."

John was extremely nervous as the games began. He prayed that the system wouldn't crash and that Watson wouldn't make any embarrassing mistakes. During the final game, when it became clear to him that Watson would emerge victorious, he glanced across the room and caught Dave Ferrucci's eye. Both smiled. They knew it was done. They had bet big and won.

THE NEXT STEPS FOR WATSON

Watson was not designed just for playing *Jeopardy!* From the start, the goal was to create a technology platform that could be adapted to a wide variety of uses—a practical tool with the potential to transform business and society. Rather than trying to create a monolithic set of rules for analyzing data, the development team used many simpler algorithms that could be added to or mixed and matched depending on the task before Watson. They created an analytics program for weighing and evaluating evidence and conclusions, no matter what domain. And they made it possible to have experts in particular domains to contribute knowledge to the program. The result: “We created an architecture of discovery,” says Eric Brown, who heads IBM Research’s Watson team.[10]

Today, IBM is adapting Watson to a number of industries. A team of software developers, with considerable help from experts in various domains, is crafting Watson solutions for health care and other industries. They’re also creating applications for cross-industry uses, such as customer-service management.

WellPoint, one of the nation’s largest health-benefits providers, worked with IBM to produce capability aimed at improving the efficiency and effectiveness of health-care payer operations—focused initially on the process of treatment preapproval management. To expand Watson’s capabilities, nurse clinicians trained Watson to understand the U.S. medical treatment code system and WellPoint’s medical policies, clinical guidelines, and process for reviewing members’ requests for treatment authorization. As of early 2013, Watson had absorbed more than 25,000 test cases and 1,500 of WellPoint’s real-life cases and gained the ability to interpret the meaning of queries and analyze them. Then WellPoint and IBM co-developed two integrated applications, one for use by payers, such as WellPoint’s health plan subsidiaries, and the other for use by medical-care providers. In one continuous process, physicians can recommend treatments for specific patients and submit those recommendations to health-care payers, who can quickly approve the treatments. The system is now being tested by WellPoint's health plans and several hundred health-care providers.[11]

A major cancer research center in the southwest is working with IBM to use Watson technologies as a medical-research application to help its physicians choose the best treatments for patients who have unusual conditions. There's a fairly large population of cancer patients for whom reliable treatments have not yet been found. Here, researchers for the cancer center are collaborating with IBM engineers to feed into Watson vast quantities of disease, treatment, and outcome data and analyze it to find hidden patterns and correlations. Do people with a combination of genetic dispositions and health problems react better to a certain treatment? Would a drug that's used now for one cancer be useful for another? Or would a combination of drugs help out? The hope is that the system will give physicians a huge head start in identifying effective treatments, testing them in clinical trials, publishing the results, and quickly getting solutions into the hands of physicians all over the world.

ANZ, a financial-services firm headquartered in Australia and with operations in thirty-three countries, is working exclusively with IBM to explore the possibilities of using the Watson technology in the financial-services industry. The goal is for ANZ to provide its wealth managers with a cognitive assistant to help them better advise the bank's 2 million wealth-management clients. Here's a scenario the bank has worked up: Imagine that you're talking to your advisor about retirement planning. Amid all of the uncertainty in today's markets, you're not sure how to coordinate your investments, retirement savings, and insurance coverage in a way that will result in a secure, well-funded retirement. The advisor, aided by Watson, will be able to quickly develop a custom-tailored plan to fit your needs.[12]

While this work is going on, the IBM Research team is developing new capabilities that will be useful no matter what the specific application is. The goal is to make Watson better at learning and at interacting with humans. On *Jeopardy!*, Watson could respond to a clue with the answer it was most confident in. In the next-generation Watson technology, people will be able to engage Watson in ongoing dialogues aimed at finding the most useful insights. After receiving an initial query, Watson will be able to ask for additional information to help it understand more precisely what the human being wants to know. The learning capability includes a series of steps: hypothesis generation from evidence, hypothesis ranking, question generation, and answer acquisition. In addition to getting answers from humans it's interacting with, Watson will be able to tap a directory of experts for advice or even learn through crowd-sourcing. As it acquires answers, it will build a collection of learned axioms that strengthen its command of given domains.

Other improvements to Watson have come. People are now able to view the logic and evidence upon which

Watson presents options. Watson is now able to digest not just textual information but also structured statistical data, such as electronic medical records. A different group at IBM is working on natural-language-processing technology that will allow people to engage in spoken conversations with Watson. At the highest level, many of the changes are aimed at moving Watson from answering specific questions to dealing with complex and incomplete problem scenarios—the way humans experience things.

In fact, as people in particular professions and industries experiment with Watson, they find that the basic question-and-answer capabilities, while useful, are not the most valuable aspects of the systems. Rather, they see Watson as a discovery tool that doesn't just make them better at what they already do; through human-machine interactions, it helps them think differently and work differently. David Nahamoo, who heads up IBM Research's speech group, says it this way: "Traditional computing systems are systems of record. The computers of the future will be systems of engagement. They're about dynamic knowledge, not static knowledge."[13]

Language is the key that makes it possible for machines to help us better understand our world and to engage with us. But don't think of language only in the most obvious sense—English, Chinese, Spanish, and the like. Every domain of knowledge and expertise has it's own specialized language, such as chemical symbols, medical images, or the specific meaning embedded in arcane legal terms.

And that poses particular challenges for computers. As we ask computers to understand and reason more like we do, we must teach them not just the structure and vocabulary of spoken and written languages but the words and concepts that are particular to professional and business domains. When they have that kind of capability, they can begin to offer us not just answers to questions but fresh insights that might not have occurred to us on our own.

Think, for example, about a challenge confronting a corporate marketing department. Time and again, as the marketing team tries out new techniques in a world that's quickly being transformed by digital media, social networking, and mobile communications, it finds that its campaigns are not effective. But what if a cognitive assistant could help? The system could examine past campaigns in detail, looking for common patterns that might help the team spot what it's doing wrong. At the same time, the system can crawl the world of information for case studies of successful digital-marketing campaigns and spot the patterns that point to success. In order to do this kind of work, the system will need to comprehend the language and incorporate knowledge from a variety of specialized domains, from classic marketing to search-engine optimization and from behavioral psychology to demographics and statistics. Everything that has a meaning within a profession or domain must have the same meaning to the system if it is to collaborate with people at a deep level.

One of the most promising domains for this kind of human-machine interaction is the health-care field. That's where the IBM partnerships with Memorial Sloan-Kettering Cancer Center and Cleveland Clinic come in.

Eric Brown's team at IBM Research is working with Cleveland Clinic on two Watson projects. First, they're embedding Watson into the curriculum of the clinic's Lerner College of Medicine. Physicians, faculty, and medical students there are helping to adapt Watson for use as a collaborative learning tool to assist with medical information exploration and training in clinical reasoning. In addition, Cleveland Clinic is using Watson to help unlock hidden insights in the unstructured data in anonymized electronic medical records, with the goal of helping physicians make more informed decisions about patient care.

Leaders at Lerner are among those who believe that medicine must be taught differently—through problem-based learning and simulation laboratories where small groups of students solve sample cases. The medical school's students use a manual visualization technique to work their way through chains of evidence, inferences, and beliefs about what's wrong with a patient and what should be done to help. They draw circles and arrows with pencil and paper to represent pieces of evidence and potential conclusions.

The IBM Research team adapted the technique to create a user interface called WatsonPaths, an interactive program for teaching Watson how physicians solve problems. Now, working with WatsonPaths, faculty and students will be able to review the evidence the computer presents and the inferences it draws. They will type in additional information and insights. As the faculty members and students interact with the machine, the arrows that represent the strongest connections will become thicker and the circles that represent important evidence and strong conclusions will become more pronounced. In the background, Watson will learn how to interpret the way physicians talk and write about their craft. The students will train the machine and, at the same time, the machine will help train the students.

Just as developing a version of Watson for *Jeopardy!* helped focus the early research efforts on the technology, the team has set a new goal that's even more challenging: becoming capable of passing the online section of the U.S. Medical Licensing Exam, the professional exam that physicians with an MD are required to pass before they're permitted to practice medicine. It includes sections on anatomy, pharmacology, pathology, behavior science, genetics, and other medical disciplines. The goal is for Watson to be capable of passing the exam as a measure of progress, though the system will not actually take it. Professors and students at Lerner are helping Watson achieve this capability. "Problem-based learning techniques for teaching medical students align with the way doctors think. In addition, medical students and physicians need help in harnessing the volumes of evidence now available," says Dr. James Young, executive dean of the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University. "We are helping to train Watson so that eventually Watson will be a partner in the delivery of healthcare for our patients."

Meanwhile, Memorial Sloan-Kettering Cancer Center, WellPoint, and IBM are developing a Watson-powered product called Interactive Care Insights for Oncology to help oncologists access the latest information on treatments.

MSKCC is helping provision Watson with the body of knowledge it will need to serve as a valuable resource to help physicians make evidence-based decisions. The corpus includes everything from case histories and vast libraries of medical literature to clinical guidelines, best practices from top physicians, and reports from drug trials. Watson has by now ingested 600,000 pages of medical evidence, 2 million pages of text from 42 medical journals and clinical trials, and several thousand case histories. The physicians are using all that data to train Watson to make treatment recommendations that are aligned with best practices. Once oncologists from MSKCC and community-based oncology centers begin testing the product, additional learning will take place. Think of it as on-the-job training for Watson. The more physicians interact with the machine, the smarter it will get.

WellPoint has selected several community-based oncology centers to test the Interactive Care Insights for Oncology technology, including WESTMED Medical Group in Purchase, N.Y. By evaluating the system's performance against their existing patient records, they can see how well it works and provide feedback to make it more usable. Once the application is operational, oncologists at WESTMED and other beta sites will be able to use it to tap into the expertise and knowledge base of MSKCC and other sources, resources they would not have had such ready access to otherwise.

A key element of preparing Watson for this new role as a physician's helper is training it to interact with people the way that's most natural for the humans. Earlier attempts at creating support systems for physicians fell short partly because they forced the doctors to adapt to the ways of the computer. Watson's user interface is being designed to work on different types of computers including desktops, laptops, or even tablets that doctors can easily carry around with them. Its procedures are based on the way physicians like to work. Getting things done should be far simpler than is typical in today's human-machine interactions.

SCENARIO: PHYSICIAN'S ASSISTANT

Dr. Mark Kris, the chief of thoracic oncology at Memorial Sloan Kettering Cancer Center, has developed a detailed scenario describing how he thinks Watson could be used by oncologists:

> Preparation: The oncologist is preparing to treat a Japanese-American woman who has lung cancer. Ahead of their first meeting, he reviews her personal information and medical records, including diagnosis and test results—presented to him by the Watson interface on his tablet computer. He has already spoken to the referring physician. Armed with all of this information, he begins developing a hypothesis.
>
> But the oncologist wants a second opinion—from Watson. So he types in a request for Watson to use all of the information at its disposal to evaluate the treatment options for this patient. Then he taps the "submit" button. In a matter of seconds, Watson pulls together the

patient information, medical records, relevant medical literature, and treatment results based on MSKCC's experience. Watson points out that women from Japan who are nonsmokers are fairly likely to have a mutation in a gene, *EGFR*. If so, a drug, Erlotinib, shrinks the cancer 95 percent of the time. Watson suggests a test to see if the patient has the gene mutation.

Watson suggests other tests as well, and, after reviewing the options, the oncologist can tap a few buttons to submit a request to the patient's insurance carrier for preauthorization of the tests.

First Consultation: The physician meets with the patient in his office. He talks to her about her condition and asks her about her treatment preferences. She tells him that she would prefer to avoid a treatment that would result in loss of hair. It would be too disturbing to her children. After the meeting, the doctor types his consultation notes into Watson and orders the new tests.

Review of Test Results: A few days have passed since the first visit, and the oncologist is reviewing the test results in Watson. It turns out that the patient has one of the EGFR mutations, but it's a mutation for which Erlotinib will not work.

Second Consultation: During the second consultation with his patient, the physician explains his deliberations. He also asks the patient if she has noticed any changes in her condition since they last met. She says yes. Two days earlier, she began noticing blood in her phlegm. The physician asks Watson how that new piece of information affects its suggested treatment

> options. The program comes back with a very different set of options, ranked on its confidence level.
>
> Watson has 90 percent confidence in the top option on the list, a cocktail of two often-used chemotherapy drugs. In addition, Watson points out that the treatment is less likely than others to cause loss of hair.
>
> After discussing the treatment with his patient and securing her agreement, the doctor taps a button on the Watson interface to request preauthorization of the treatment from the insurance company. All of the evidence that he and Watson have gathered is forwarded to the insurance company to support the request.

The system works for all the parties involved. The physician gets access to a tremendous amount of up-to-date information that he might not have gathered and digested on his own. The patient gets a carefully chosen treatment that's likely to be the most beneficial and that may spare her unpleasant side effects. She also doesn't have to wait long to begin treatment. The system works for the insurance company, too. It gets the evidence it needs to confidently give the go-ahead for a treatment regimen. And society gets better health care that's more efficiently and affordably delivered. "We're making evidence-based decisions. Watson proves it," says Mark. "It's going to be much better for all of us."[14]