Cognition as a Service: An Industry Perspective

Jim Spohrer, Guruduth Banavar

■ *Recent advances in cognitive computing componentry combined with other factors are leading to commercially viable cognitive systems. From chips to smart phones to public and private clouds, industrial strength "cognition as a service" is beginning to appear at all scales in business and society. Furthermore, in the age of zettabytes on the way to yottabytes, the designers, engineers, and managers of future smart systems will depend on cognition as a service. Cognition as a service can help unlock the mysteries of big data and ultimately boost the creativity and productivity of professionals and their teams, the productive output of industries and organizations, as well as the GDP (gross domestic product) of regions and nations. In this and the next decade, cognition as a service will allow us to reimage work practices, augmenting and scaling expertise to transform professions, industries, and regions.*

From an industry perspective, the era of cognitive computing has dawned with the promise of human-centered cognitive prostheses, as just one of many benefits anticipated (Kelly and Hamm 2013). Gartner, the technology industry analyst firm, projects that by 2017, 10 percent of all computers will be learning (Plummer 2013). However, scholars familiar with the field of artificial intelligence and cognitive science are rightfully cautious, after witnessing a roller coaster of ups and downs over AI's relatively short 60-year history. Simply put, hard problems still remain.

So how can we know if this time is really different? After all, three of the key pillars of cognitive computing, namely machine learning (ML), natural language processing (NLP), and hypotheses generation with evidence-based explanation (EBE) capabilities, have existed for quite some time. The first author personally recalls programming hidden Markov model (HHM) learning algorithms for speech recognition in a

 ISSN 0738-4602

startup company in the late 1970s, as well as statistical natural language processing at the same company (for example, Verbex, from which Dragon Systems sprang) in the early 1980s. In graduate school at Yale a few years later, along with fellow graduate students, the first author built a number of explanation systems, including a master's thesis that provided expert and novice explanations of economic arguments after ingesting articles from the *Wall Street Journal*, as well as a dissertation that provided explanations of bugs by simulating novice programmers (Spohrer 1992, Kahney 1993). Frankly, while there was significant progress made in those earlier times, the context today is dramatically different and the commercial and scientific opportunities are vastly greater.

The most obvious differences between now and then are well documented (Brynjolfsson and McAfee 2014). First, there is no doubt that processor speeds, software algorithms, as well as storage capacity for data sets, both training and test data, have advanced by orders of magnitude, while costs have plummeted. Second, most of us are walking around with smart phones that not only provide easy access to the sum of crowdsourced human knowledge, but also provide easy access to a rapidly growing number of cloud-based global service organizations ready to compete for a share of our attention and wealth. Nevertheless, societal challenges, including the rising cost of continuously reskilling the work force and effectively addressing unemployment due to skills gaps even among recent college graduates, loom large in the minds of policy makers.

In this article, we provide an industry perspective on the topics of cognitive computing, cognitive systems, cognition as a service, and human-centered smart service systems. In conclusion, we describe a number of remaining research and commercialization challenges, and a new set of university programs aimed at addressing those challenges.

Cognitive Computing

Figure 1 highlights some of the recent growth of industrial strength cognitive computing componentry, which can be found in the diverse offering from many companies; every month there are more announcements, from small university startups to large company business unit formations, projects, products, and acquisitions (Harris 2013, Spivak 2013, Ardire and Roe 2014, Austin and LeHong 2014).

Cognitive computing is a term coined by industry leaders to refer to a third era in computing beyond the tabulator era and the programming era (Kelly and Hamm 2013). A proposed shared definition of cognitive computing by Ardire and Roe (2014) is narrowly focused on natural language processing, machine learning, and big data:

> Natural language processing of structured, unstructured, streaming-in big data, or smart data layers with machine learning for reasoning and learning to generate textual patterns and associations that enable humans to connect the dots faster and smarter for more informed decisions to drive better outcomes.

In practice, cognitive computing componentry builds on knowledge (concepts, algorithms, data sets) from artificial intelligence, cognitive science, neuroscience, and other related fields, in order to build industrial-strength platforms and applications that can provide cognition as a service to customers. The era of cognitive computing depends in part on the cloud, smart phones, and big data. Cognitive computing can be defined succinctly as:

> Cognitive computing is the set of industrial strength computational componentry required to deliver cognition as a service (including the three Ls — language, learning, and levels) to customers. Cognitive computing enables cognition as a service, and powers cognitive systems that augment and scale human expertise.

For the three Ls of cognitive computing: *language* refers to naturalistic input of multiple types, such as documents, audio, and video; *learning* implies the use of training data and test data sets for high performance on tasks with well-specified input-output pairs; and *levels* refers to levels of confidence in output hypotheses, with reasoning chains and explanations. If robotics is considered, we could add a fourth L, *limbs*. These four Ls are reflected to some degree in what Gartner has identified as seven machine smartness dimensions (Austin and LeHong 2014), including: appearing to understand and reflecting a well-scoped purpose (that is, language), learning actively and passively (that is, learning), dealing with complexity and making probabilistic predictions (that is, levels), acting autonomously (that is, limbs). Gartner is using these dimensions to evaluate commercial offerings and communicate capabilities to potential industry customers.

Figure 2 reminds us that in 2011, the Watson Jeopardy! DEEP-QA system performed well enough on the television game show *Jeopardy!* to defeat two of the all-time top champions, reflecting the three Ls in its highly tuned, real-time architecture for a well-specified task (Ferrucci et al. 2010). The system was trained on some 5.7 million training instances of input-output pairs, or a set of approximately 25,000 *Jeopardy!* questions with more than 200 correct/incorrect answers on average per question, with each training instance composed of 550 features (Gondek et al. 2012). The systems confidence estimation framework required more than 7000 experiments. The resulting win is now part of the history of AI.

Going well beyond that 2011 success, the IBM Watson Business Unit established in January 2014 describes cognitive computing as:

> . . . forging a new partnership between humans and computers that augments and scales human expertise, based on cognitive technology that can understand natural language, generate hypotheses based on evidence, and learn as it goes.

Company	Business Units, Projects, Acquisitions
Apple	SIRI
Facebook	Social Graph
Google	Dynamics, Bot and Dolly, DeepMind, DNNresearch Inc., Flutter, Holomni, Industrial Perception, Jetpac, Redwood Robotics, Meka Robotics, Metaweb, Nest, Now, SCHAFT, Inc., Viewdle, Wavii
IBM	Cognea, TrueNorth, Watson Business Unit
Intel	Ginger
Microsoft	Cortana
Qualcomm	Zeroth
Yahoo!	IQ Engine, LookFlow
Others: Adobe, Baidu, BHP Billiton, HRL Laboratories, LinkedIn, SRI International, Twitter, Wolfram Alpha	
Also: AlchemyAPI, Arria, Automated Insights, Cortica, Cognilytics, Cognitive Scale, Coherent Knowledge Systems, Deep Knowledge Ventures VITAL, Dropcam, Emerald Logics's FACET, Ersatz Labs, Gild, Knewton, Modernizing Medicine EMATM, Mohiomap, Narrative Science, N2Semantics, Next IT Alme for Healthcare, Numenta, Onlyboth, PARO Therapeutic Robot, Rethink Robotics Baxter, Rio Tinto, Saffron Technology, Semantria, Skytree, Sight Machine, Song Kick Existor, Talis, Vicarious, Viv Labs, Wise.io, and Yseop	

Figure 1. Commercialization of Cognitive Computing Componentry Circa 2014.

The AAAI-14 Workshop on Cognitive Computing and Augmenting Human Intelligence described cognitive computing as:

> . . . an emerging research topic inspired by a vision of how the unification [of machine learning and naturalistic input processing] could lead to a new generation of computing systems enabling genuine human-machine collaboration.

Cognitive computing componentry is also being developed at the chip level. Figure 3 highlights a recent IBM-led DARPA-funded advancement published in *Science* regarding the TrueNorth chip (Merolla et al. 2014). Figure 4 shows the array of cores that make up the chip and allow processing close to the data. This "neuromorphic chip" has been used in image-recognition tasks and other naturalistic input tasks. The chip contains 5.4 billion transistors connected to form an array of 1 million "digital neurons" linked by 256 million digital synapses. The chips are designed to be tiled into larger arrays, so a 10 x 10 array of chips would yield a 100 million neuron and 25.6 billion synapse cognitive computing module. A 32 x 32 array of such modules would in turn be approximately approaching estimates of a human brain's 100 billion neurons and 100 trillion synapses. TrueNorth uses a fraction of the power required by conventional von Neumann chips simulating digital neurons and synapses (Service 2014). A multiobject recognition task with input of 400 x 240 pixel video at 30 frames per second consumes just 63 milliwatts. TrueNorth's power density is 20 mW per cm^2, which is more than 1000 times less than a typ-

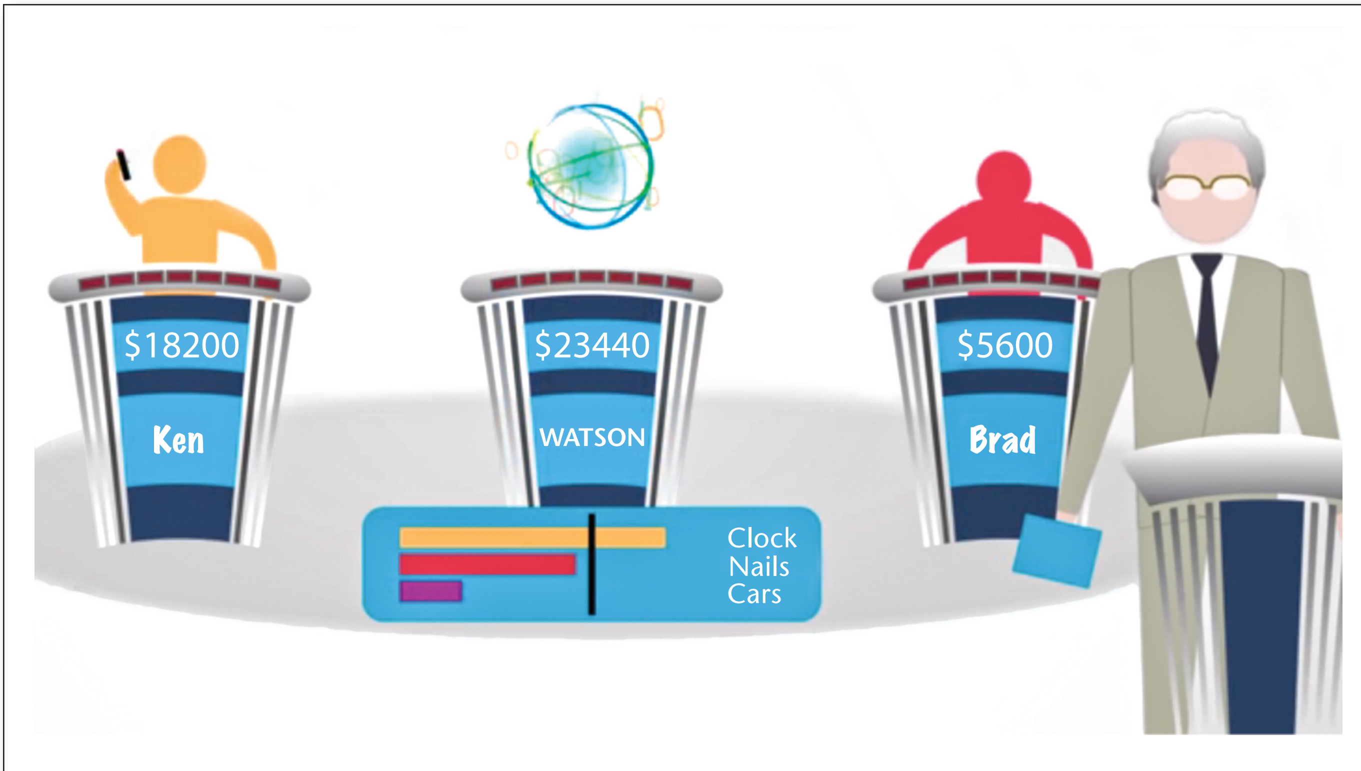


Figure 2. Watson Jeopardy! illustrated the 3 Ls of Language, Learning, and Levels.

Figure based on Brown (2013).

ical central processing unit in a laptop or server today.

The architectures of industrial strength cognitive computing componentry are becoming increasing diverse, at the hardware (von Neumann, neuromorphic, robotic), software (algorithms, knowledge graphs, training and test data sets), and network (cloud, Internet of Things) levels. As this componentry continues to evolve, key data sets and knowledge graphs could become standard commodities, and in some cases built directly into the hardware level. This is especially true of historical and episodic data and its manifestation in multiple languages and media formats. More recent data sets require more complexity to ensure security and privacy, as well as to meet the challenges of nascent business and regulatory models (Ng et al. 2013). As more data sets are built directly into the chip level, certain techniques for building chips may have advantages over other methods. The costs and benefits of self-assembled neuromorphic devices, for example, will need to be better understood (Avizienis et al. 2012). New programming models are often required to program neuromorphic devices. Figure 5 illustrates the Corelet programming language for the TrueNorth chip, which requires converting input corresponding to spiking patterns, processing the data through spiking pattern transformations, and then converting spiking patterns into output results (Amir et al. 2013).

Cognitive computing builds on and continues to benefit from the work of artificial intelligence, cognitive science, and neuroscience researchers as well as researchers from other related fields. Broad availability of industrial strength cognitive computing componentry will help researchers tackle even more ambitious challenges by enabling cognitive systems that boost creativity and productivity of researchers, and accelerate discoveries. The design, development, and delivery of cognitive systems will also be improved by these advancements, in a positive feedback loop powered in part by the growth of big data in multiple application domains.

Cognitive Systems for All Tasks and Professions

As cognitive computing componentry continues to mature, and along with curated data sets enables cognition as a service, cognitive systems that augment and scale human expertise can be more rapidly developed and deployed.

The opportunity to increase the productivity and creativity of professionals with the help of cognitive assistants is becoming a reality, profession by profes-

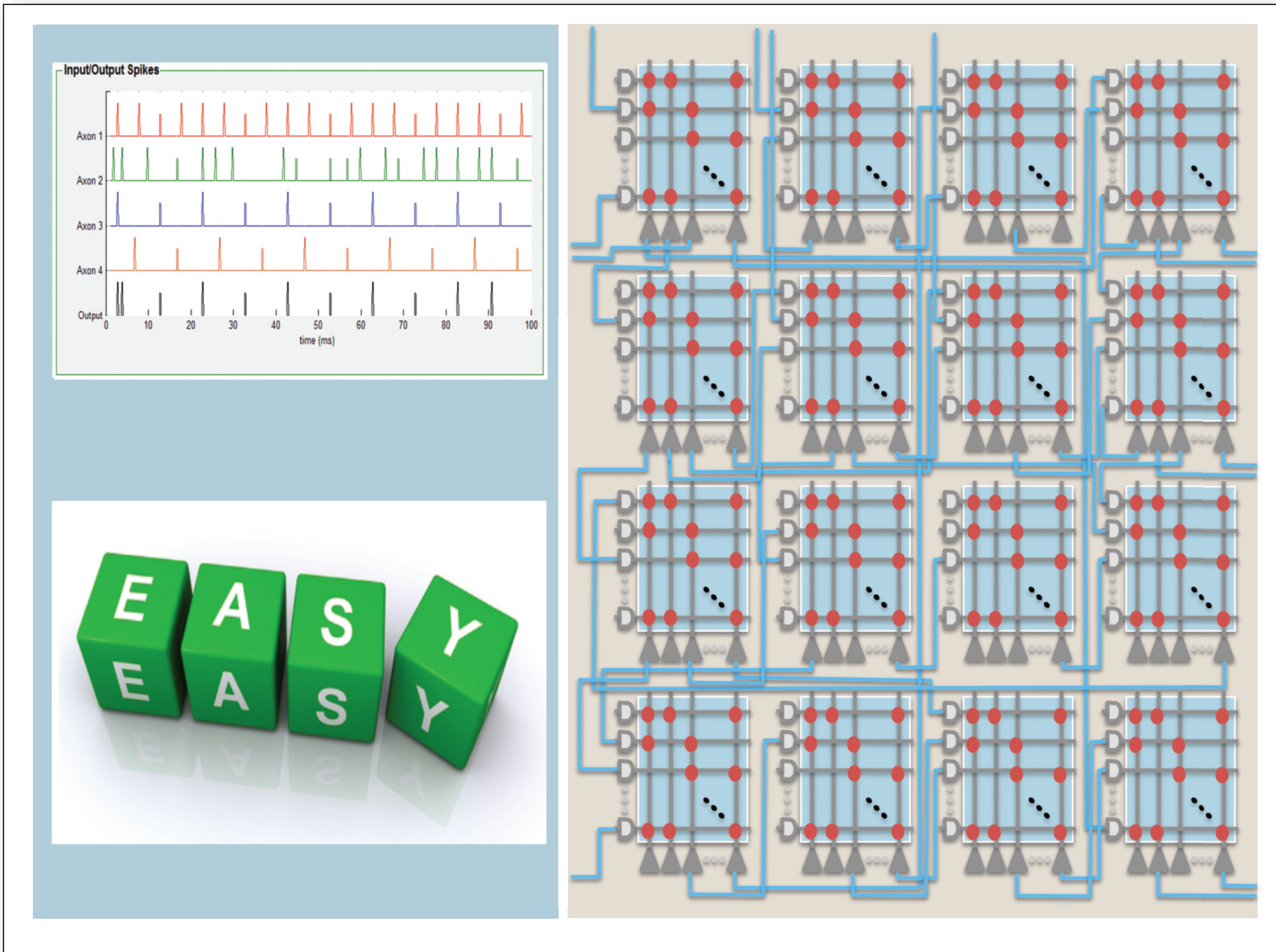


Figure 3. TrueNorth, a Network Execution Engine, Is Cognitive Computing Componentry on a Chip.

Programming TrueNorth requires specifying: (1) the dynamics of each neuron, (2) the configuration of each crossbar (defines local connectivity for each core), and (3) the routing of the network (which neurons are connected to which axon). Figure based on twitter feed by Dharmendra Modha.[2]

sion. Furthermore, thanks to the triple helix collaborations of university researchers, industry practitioners, and government funders, the pathways to creating more professional cognitive assistants are becoming more clearly defined every day.

In particular, consider the website O*Net OnLine,[3] which provides a description of about one thousand occupations. Each occupation described on O*Net has a description of the tasks that professionals commonly perform as practitioners within the occupation, as well as information about tools and technologies, knowledge, skills, abilities, work activities, work context, credentials, and other aspects that in part define that occupation. Many of these occuptions could be affected in some way by the availability of cognitive assistants. Figure 6, for example, shows that biochemical engineers (17-2199.01 — the number O*Net has assigned) is an occupation with a bright outlook with expected job growth.

One of the most important applications of cognitive systems will be to accelerate discoveries made by researchers sifting through mountains of data. For example, consider biochemist and biophysicists (19-1021.00) and biochemical engineers (17-2199.01) working to understand cancers by searching for kinases, or complex molecules that play a critical role in cell metabolism. Important new kinases are typically discovered at the rate of one per year; however, in one recent study, seven promising new kinases were discovered in a fraction of a year when biochemists used a cognitive system that had ingested 60,000 research papers focused on the p53 protein involved in cell growth (Simonite 2013).

Next, consider the innovative chefs (chefs and head cooks 35-1011.00) in *Bon Appétit*'s test kitchen

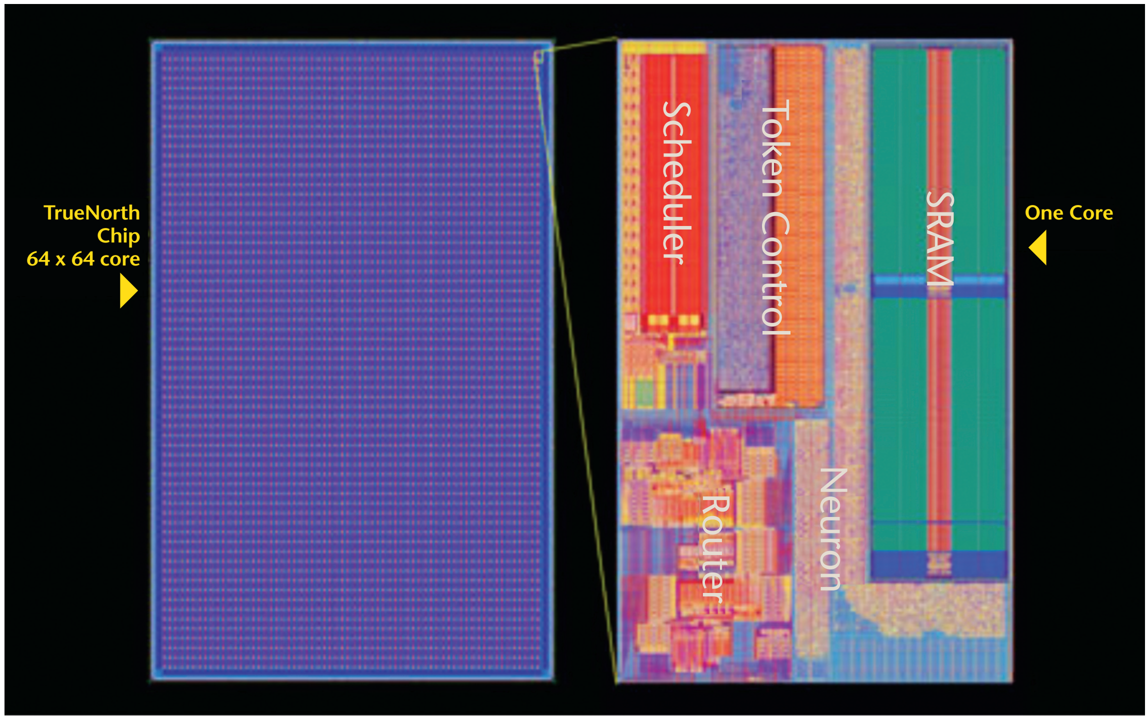


Figure 4. TrueNorth Chip Has an Array of Cores Architecture.

(Bilow 2014). Chef Watson "helps cooks discover and create original, totally unique recipes with the help of flavor compound algorithms." Training data consist in part of *Bon Appétit*'s database of nearly 10,000 recipes that include many ingredients in diverse combinations, and grouped into culinary categories. In addition, encyclopedic knowledge of complementary flavors compounds and a Bayesian surprise estimation method are used to inform search through quadrillions of possibilities to realize a cognitive assistant that helps chefs innovate (Varshney et al. 2013).

Often several occupations will have common tasks and knowledge. For example, climate change analysts (19-2041.01), political scientists (19-3094.00), and legislators (11-1031.00) must make legislative recommendations based on analysis of relevant scientific and economics reports. With the growth of scientific and economic reports on multiple sides of an issue, these professionals will benefit from cognitive assistants. Cognitive systems that can debate on many topics are beginning to appear (Cuthbertson 2014, Aharoni et al. 2014).

The impact of cognitive assistants can potentially be quite large, especially for occupations with many millions of workers globally. The number of call center agents in the world is estimated to be more than 2 million, including customer service representatives (43-4051.00), receptionists and information clerks (3-4171.00), telemarketers (41-9041.00). Organizations that provide this type of service often create enormous data sets that can become the basis for data curators to build training sets for cognitive assistants (Mishne et al. 2005). Also, improved customer satisfaction correlates with increased sales, justifying investment in customer engagement advisors (IBM 2013).

Health care is another area where cognitive assistants have enormous potential to improve performance, including physicians and surgeons (29-1069.00), registered nurses (29-1141.00), and home health aides (31-1011.00). The potential demand for health-care cognitive assistants is enormous; for example, the medical staff at Memorial Sloan Kettering treats more than 30,000 patients with cancer every year (Basset 2014). These cognitive assistants will need to have the ability to interpret diverse types of medical data sets, including medical images (Ting et al. 2013). Figures 7 and 8 illustrate the Watson Pathways interface designed to make evidence-based explanations easier for people to understand.

Within the corporate world, chief executives (11-1011.00), business intelligence analysts (15-1199.08), market research analysts (13-1161.00), investment fund managers (11-9199.03), and the like, will also benefit from cognitive assistants. After ingesting a company's strategy documents, competitive analysis

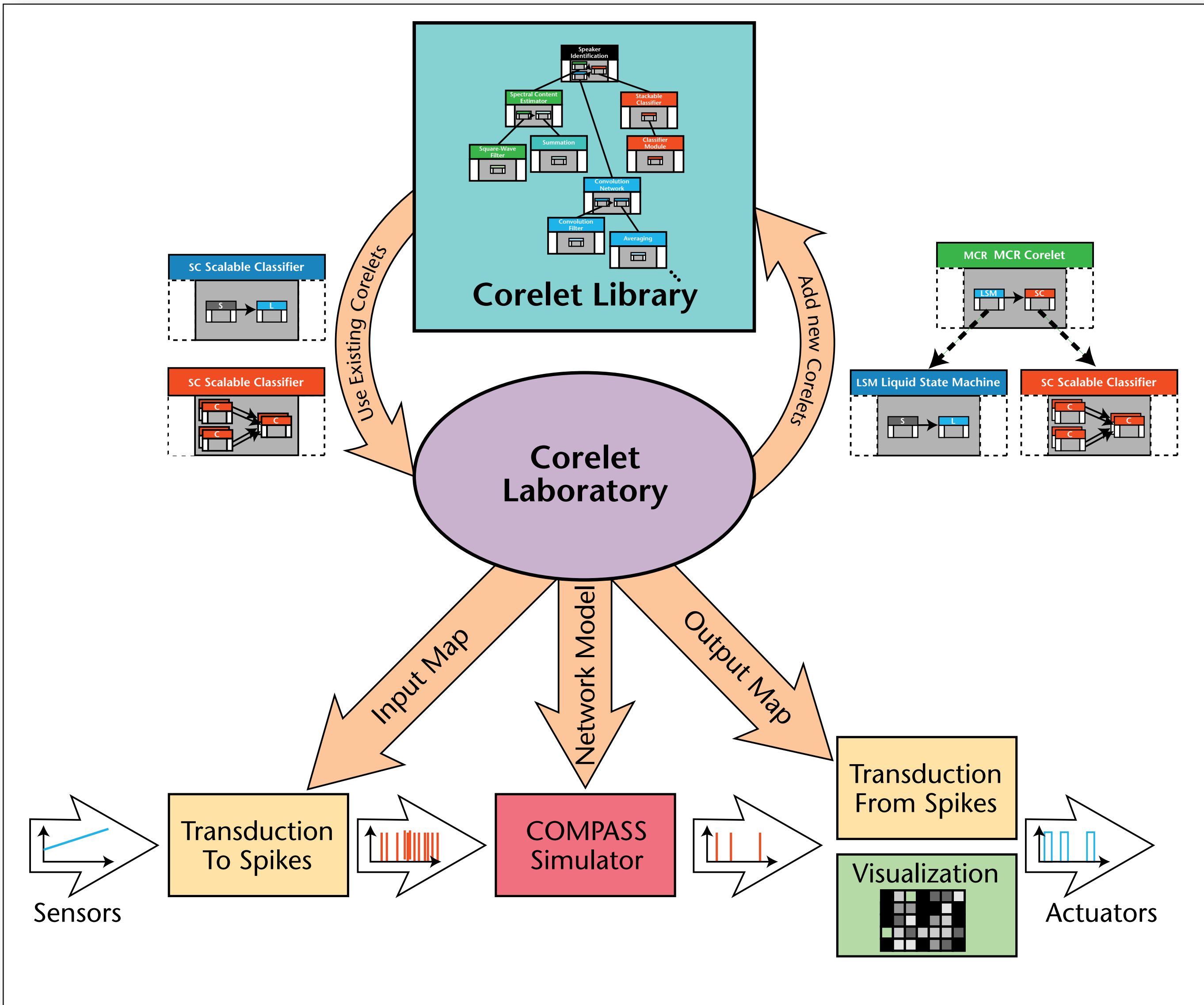


Figure 5. The Corelet Programming Language for Programming TrueNorth Chips Spiking Patterns.

documents, and market segment analysis documents into a cognitive system, teams of C-suite executives can brainstorm strategies in a decision support environment (Simonite 2014). The workplace of the future will log conversations to encourage all team members to contribute, check facts and assumptions to highlight mistakes, boost the collective IQ of teams, and retrieve relevant documents on the fly (Popescu-Belis et al. 2008, Pentland 2014). Cognitive assistants will also boost the productivity and creativity of executive administrative assistants (43-6011.00), and this type of cognitive assistant will benefit from deep knowledge of the social graphs of organizations and industry networks (Erickson et al. 2008). For example, the e-Merridy project in IBM Research uncovered issues concerning situational awareness, knowledge management, and prioritizations of event streams that arise when one entity assists another; such findings are proving useful in the design of cognitive assistants in general.

Hard questions remain to be solved in order to accelerate the development of cognitive assistants for job tasks and professions. For example, can textbooks or professional guidebooks be used as a basis for bootstrapping cognitive assistants? How best can human performance be measured in ways that allow comparison to machine capabilities? For use in complex sociotechnical systems with teams of people and their cognitive assistants, how best can interactions be structured? As governments, foundations, and

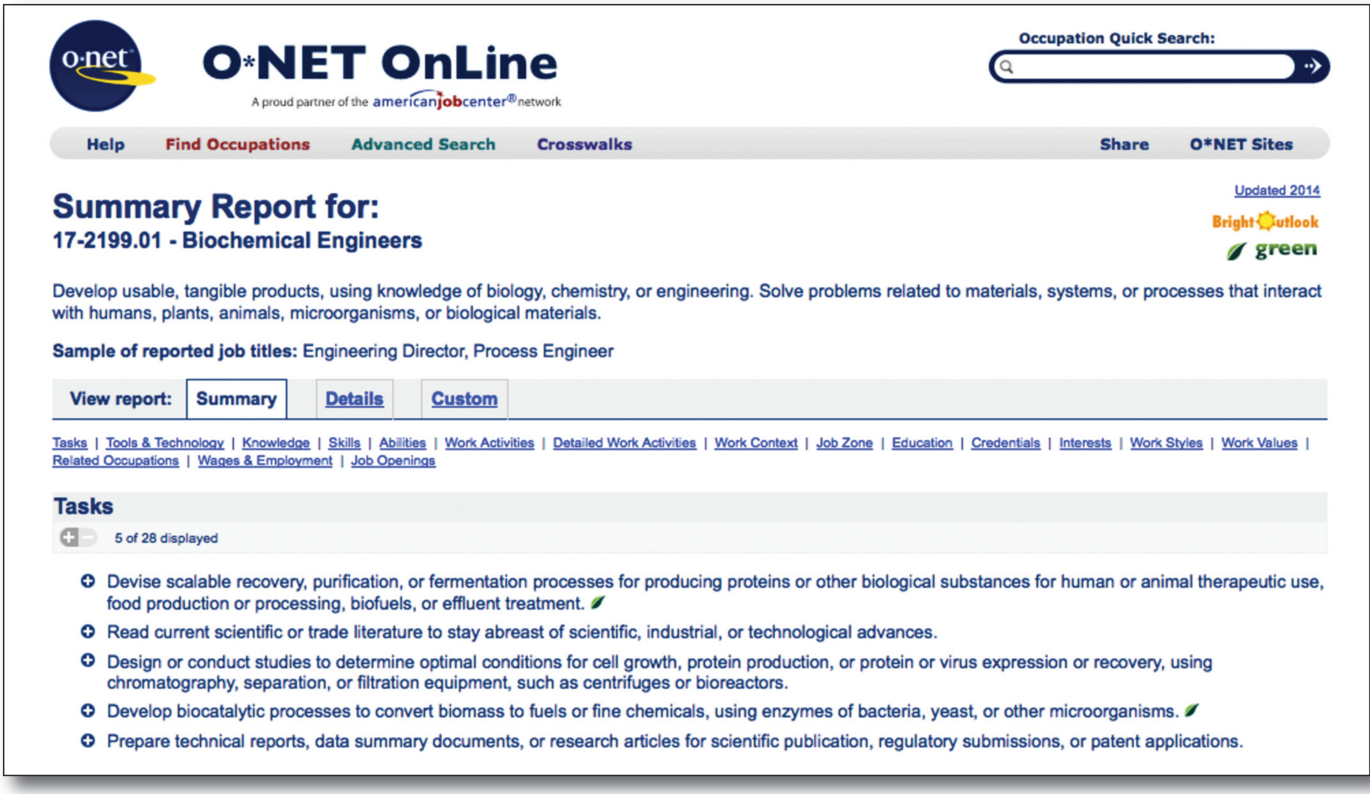


*Figure 6. O*Net OnLine Tasks Descriptions for Hundreds of Occupations.*

regional economic development groups look for high return infrastructure investments, public-private partnerships that link university researchers with regional industry ecosystems to develop and deploy cognitive assistants will become a priority. Augmenting and scaling the expertise of key professions region by region will likely build on existing regional advantages and centers of excellence (Florida 2010).

Human Centered Smart Service Systems

Augmenting and scaling human expertise through cognitive assistants is likely to emerge as the key to sustainable advantage of organizations and regions, because productivity and innovativeness of human capital are both improved (Lewis 2005). The National Science Foundation has begun to establish funding programs aimed at human-centered smart service systems to build needed innovation capacity (NSF 2013). Human-centered smart service systems can be defined as sociotechnical systems that involve a technology platform, such as cognition as a service, large numbers of customers who share information to realize benefits, such as providing their cognitive assistants with detailed user models, and an integration of knowledge and technologies from a range of disciplines, often including engineering, computer science, social sciences, behavioral sciences, and cognitive science, paired with market knowledge to increase its social benefit. Human-centered smart service systems are where people, technology, organizations, and information come together to co-create value, both social and economic (Maglio, Kieliszewski, and Spohrer 2010). From an industry perspective, a variety of capabilities — cognitive, cloud, analytics, mobile, social, secure, system of systems — combine to enable the development of smarter systems at all scales in business and society.

Spivak (2013) describes a world in which "…vast cognitive capabilities of global CaaS [cognition as a service] providers will be cheap and available via APIs to every device from the nano scale up to the giant global applications and services … a world where everything is smart probably will be quite normal to our grandkids. In their world cognition as a service could be as ubiquitous as electricity is for us today."

But how can we get there? Hard research and development problems remain. The design loop for sociotechnical systems, especially human-centered smart service system that augment and scale expertise through cognitive assistants, can only be improved through both improved multidisciplinary collaborations at universities and better collaboration among academia, industry, and government (Kline 1995).

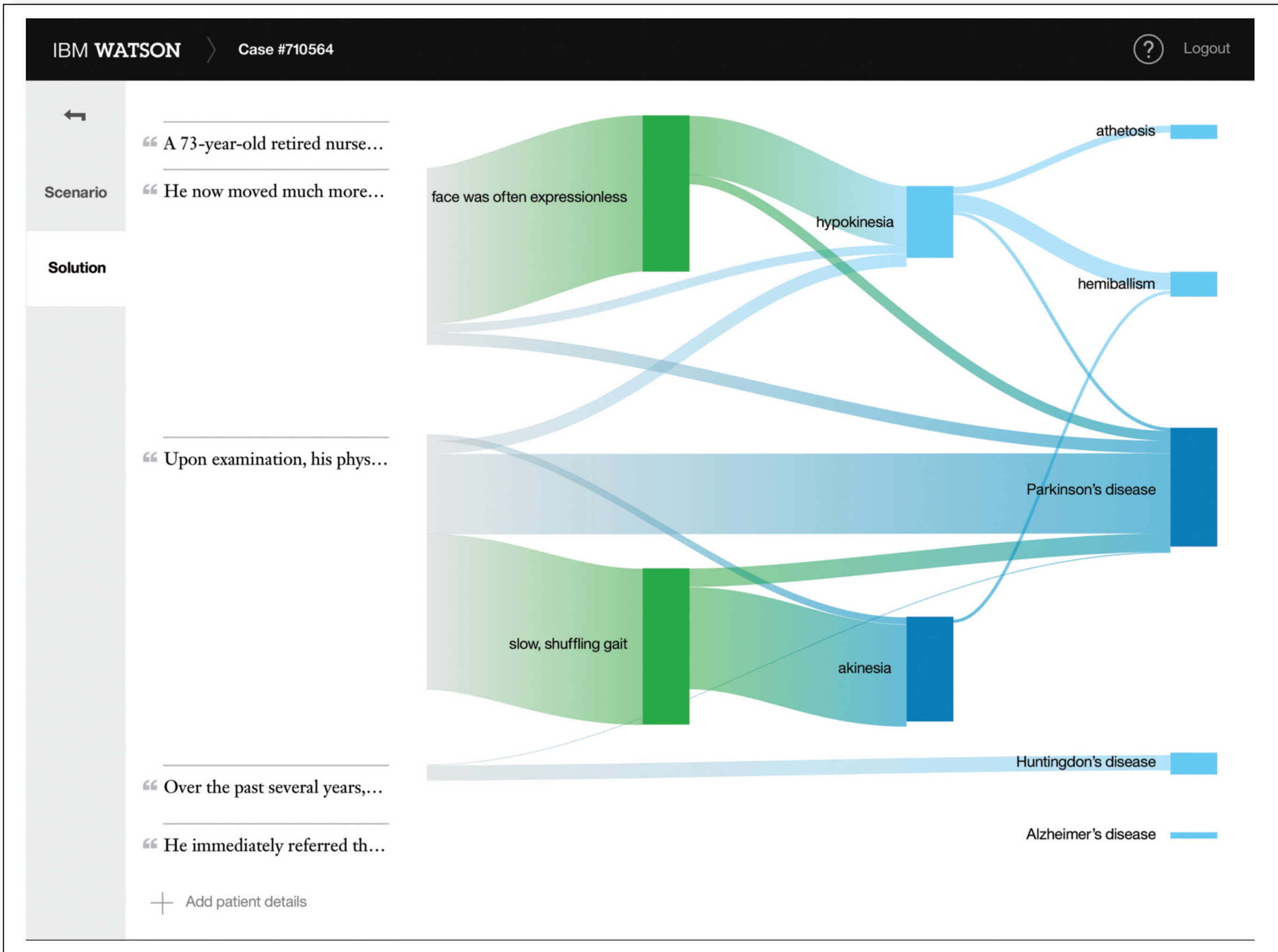


Figure 7. Watson Pathways for Health Care Generates Evidence-Based Explanations.

To improve the design loop for sociotechnical systems, the cognitive capabilities must be feasible — being developed by AI researchers; desirable — evaluated in the context of meaningful human tasks where human performance measures are available on specific standardized data sets; and viable — related to occupations where improved performance generates significant social and/or economic benefits.

For example, consider the socially significant work of childcare workers (39-9011.00) and child, family, and school social workers (21-1021.00). Augmenting and scaling this type of human expertise through cognitive assistants and smart environments can provide enormous social benefits, even if just a small fraction of developmentally challenged children can get appropriate interventions to help their language and social skills develop. Economic returns of between nearly 2 times and over 10 times are well documented for early childhood interventions (Karoly, Kilburn, and Cannon 2005). Early language learning depends critically on joint attention between child and caregiver, and assessing joint attention can be complex (Pusiol et al. 2014).

Image understanding and video stream analysis capabilities will be critical to develop cognitive assistants for many occupations. In just the past four years on one data set that has grown to 14 million images, cognitive capabilities are showing large (fourfold) performance improvements, due to factors including improved algorithms and larger data sets for training (Markoff 2014, Ramanathan et al. 2014). Working with governments, foundations, and industry to develop cognitive assistants for professionals, including childcare workers, university researchers can play a key role in improving the design loop for sociotechnical systems and human-centered smart service systems.

To further improve the design loop, augmenting professions that deal with large data sets will be critical, including data warehousing specialists (15-1199.07), clinical data managers (15-2041.02), and curators (25-4012.00). To achieve artificial general

Big Data and the Cognitive Computing Era

Big data is now a fact of life. Over the last decade, new kinds of unstructured data from social networks, streaming data, and online publications, as well as massive data emitted from sensors from the physical world have outpaced traditional forms of structured data. And this will continue to grow exponentially. The insights embedded in this massive amount of data can provide unprecedented opportunities for business and social value. Data has indeed become one of our most precious resources, and with its accelerated pace of evolution, its growing abundance will surely factor heavily into the future trajectory of business and society.

From a macroeconomic viewpoint, another major trend over the last century is the rapid growth of the service sector, which contributes more than 80 percent of the GDP of modern economies. Knowledge-based service offerings are the biggest proportion of this growth. Knowledge work, which is at the core of knowledge-based service growth, will change profoundly in the face of big data. The value of future enterprises will be determined more and more by how well they enable knowledge workers to mine the most important insights from big data.

Knowledge workers who have the ability to exploit the insights embedded in big data have the potential to be hugely more effective than they have been in the past. These knowledge workers will need new tools to extract insights out of modern big data, which is unstructured, noisy, and unreliable. These new tools cannot be built from the same techniques — manually specified rule-based symbolic computing — that helped us exploit clean, structured data of the past. These new tools will use machine learning and interact more naturally with us, providing evidence-based explanations of candidate insights. Explicitly represented knowledge will itself become big data, and mining it will reveal commonsense knowledge as well as uncommon insights.

We call this emerging class of systems that learn and interact naturally with us to perform knowledge work *cognitive systems*. For example, IBM's Watson Business Unit is working with customers to develop a family of systems that are capable of learning and interacting naturally in a variety of domains. The future of cognitive systems goes beyond question answering to support discovery of insights hidden in big data, such as in huge repositories of scientific literature, reasoning with evidence to support or refute topics of discussion, and to go beyond textual data to images and videos.

In a few years, cognitive systems will help us perform complex tasks in almost every domain: from health care to education to business strategy. Knowledge workers in nearly every domain will have tools — or *cognitive assistants* — to help them penetrate and interpret huge amounts of data, solve complex problems, and create new ideas. For example, a physician can connect information about an individual's genome, to that patient's clinical history, to the vast body of experimental literature, to diagnose and get better treatment options, with the help of cognitive systems. An educator can accelerate learning by tying the content to an individual's needs and goals compared to the current precanned, monolithic systems today. A business leader can extract insights like demand patterns, product acceptance, and competitive differentiators within markets to inform tactical and strategic decisions. Similar arguments can be made for most knowledge-based professions.

As we move to this new era, concerns will emerge related to cybersecurity, privacy, and other important societal dimensions. Since malicious events can happen very fast, how to react to prevent disastrous consequences from such attacks is a major challenge. Also, there is a growing gap between technology-enabled capability and the social and economic value that nations are able to derive from that capability. Simply put, their educational and training institutions cannot supply the market quickly enough with the skills required to take immediate advantage of new technologies. In large measure, this is because the development cycle of technology, and its diffusion throughout the economy, has accelerated to a matter of months as opposed to years.

Every new era of technology, from the industrial era to the Internet era, has had a massive impact on the world. Productivity goes up, professions are redefined, new professions are created, and certain professions become obsolete. Cognitive computing certainly has the potential to have such a massive impact on the entire segment of knowledge-based professions.

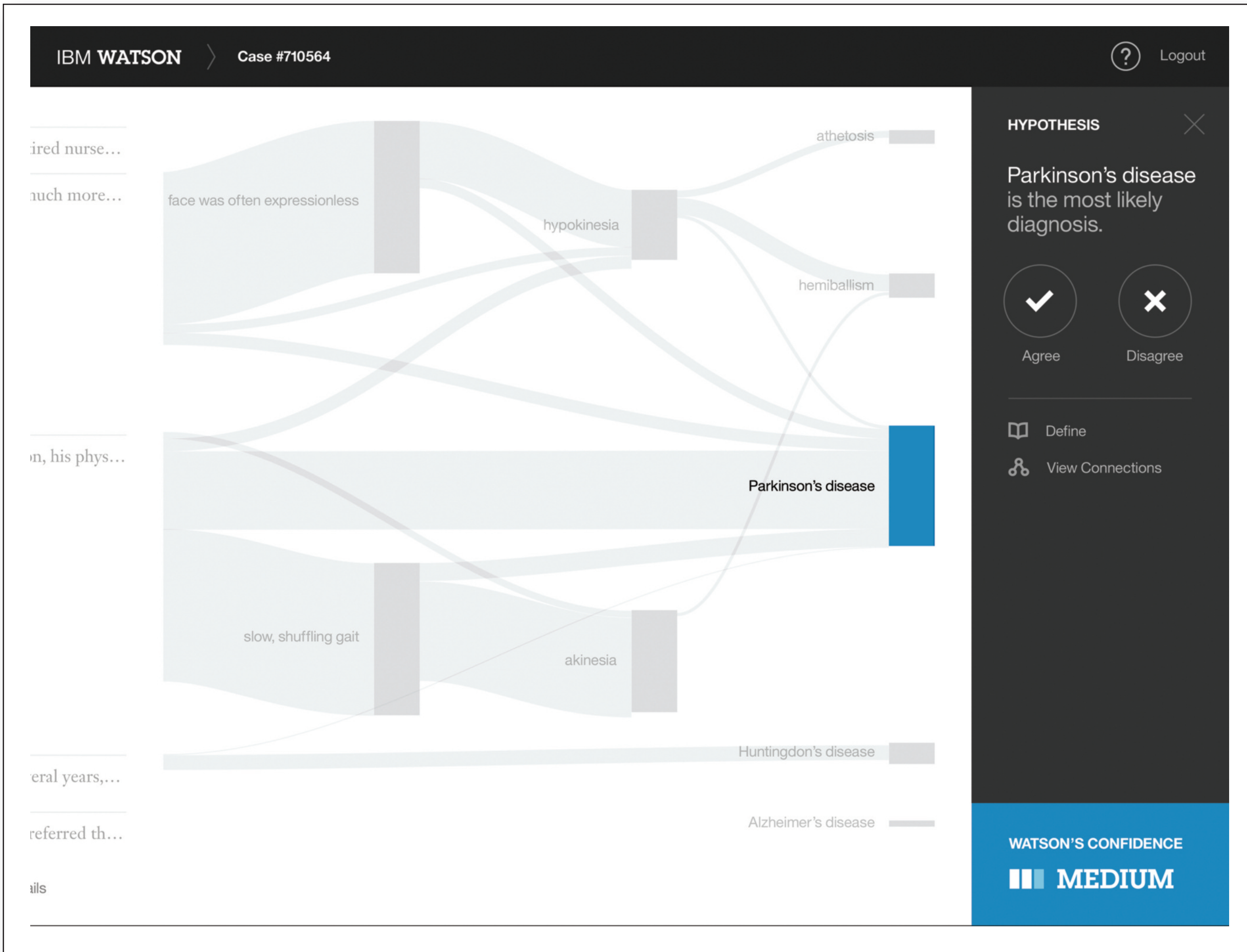


Figure 8. Medium Level of Confidence in Hypothesis.

intelligence, large child development data sets or some equivalent could be required to boot up a brain/mind with necessary capabilities to interact with physical and social environments. The MIT Media Labs Human Speechome Project provides an example of the enormous amounts of data (for example, 300,000 hours of video and audio) that can be captured from early childhood, and the tools needed to process and analyze it (Roy et al. 2006, Roy 2011).

Challenges and Opportunities

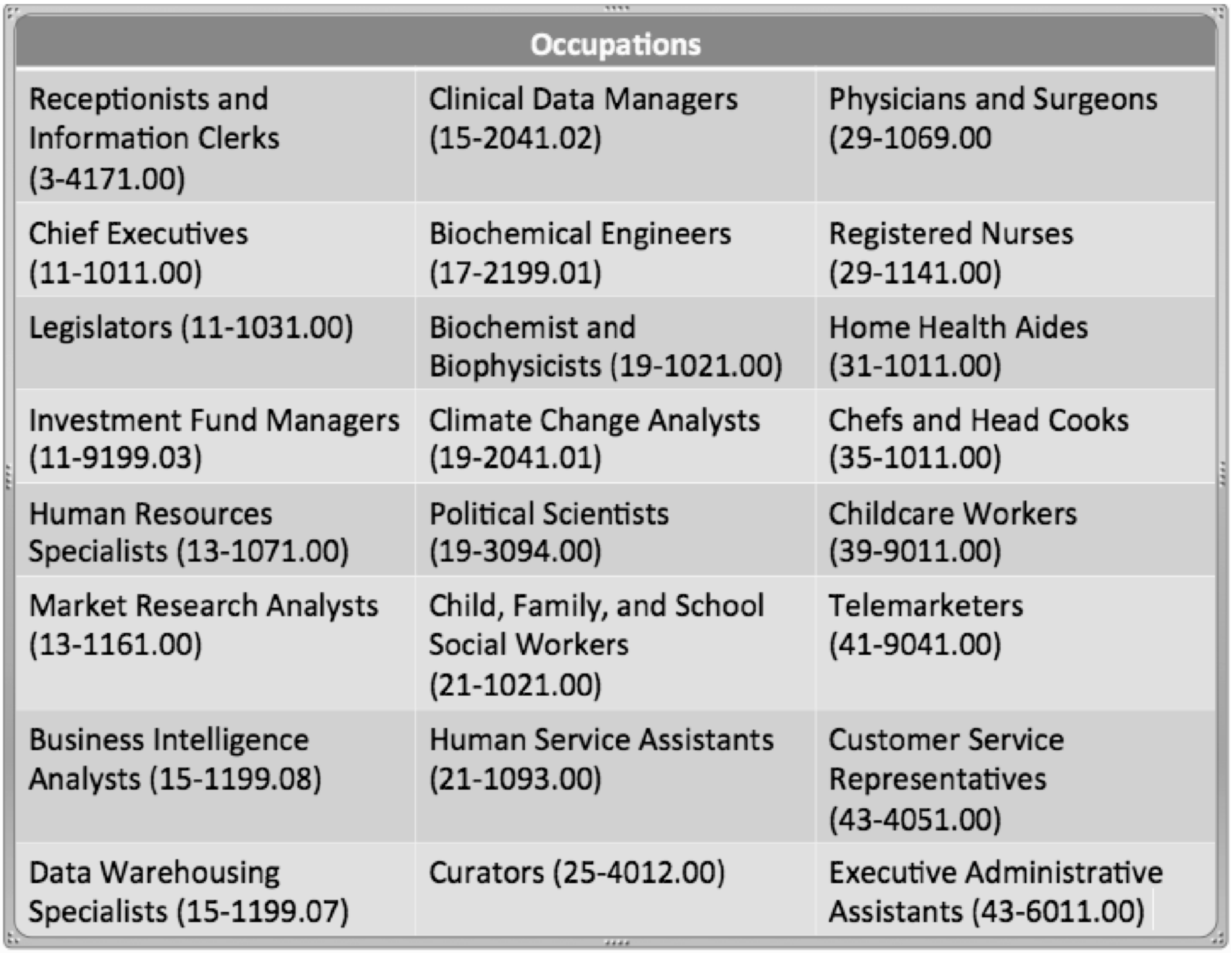
Many challenges remain — in fact, defining better grand challenges is one of the challenges. While the Turing test has inspired multiple generations of AI researchers, it has also been criticized as driving shallow deception capabilities more than deep perception and reasoning capabilities (Marcus 2014). Watching a video (for example, a news show, comedy movie, classroom lecture, and others) and answering questions, at least as well as people of specific ages and cultural/language groups perform, would be a better grand challenge of genuine human-level intelligence. Several organizations have proposed grade-level reading comprehension, as well as video-watching deep question-answering systems as a better grand challenge than the Turing test to drive the development of useful cognitive computing componentry. However, designing a better grand challenge is not easy, in part because almost any challenge can be gamed by sufficiently clever people, as discovered by those who proposed, "a TED talk given by a robot, which has ingested all previous TED talks, and is worthy of a standing ovation" (Poladian 2014). Challenges are being sought that both advance a deeper understanding of human cognition (science) and give rise to useful cognitive computing componentry (engineering).

A different type of grand challenge is dealing with public perception and formulating public policy in the era of cognitive computing. In a recent survey by Pew Research (Smith and Anderson 2014), most respondents agreed "robotics and artificial intelligence will permeate wide segments of daily life by 2025," but 48 percent of the responses voiced concerns that more jobs will be displaced than created, at least in the short term, and "this will lead to vast increases in income inequality, masses of people who are effectively unemployable, and breakdowns in the social order."

Perhaps in part because of views about accelerating technological change, two of the occupations with a bright outlook according to O*NET OnLine are social and human service assistants (21-1093.00) and human resources specialists (13-1071.00). With an estimated $100 billion spent on unemployment benefits each year, cognitive systems have been proposed as one way to augment and scale the expertise of human service assistants to disrupt unemployment with innovation (Nordfors 2014). Others have argued the merits of a basic income guarantee to deal with these concerns (Walker 2014). Laws and rules matter as much as technology and innovations in sociotechnical systems. Therefore, as cognitive systems accelerate the augmentation and scaling of human expertise, the design loop for sociotechnical systems (Kline 1995) will become more explicit, forcing more consideration of the rules for smart service system operations as well as the role of universities and other institutions driving new knowledge creation and "progress" (Spohrer, Piciocchi, and Bassano 2012; Spohrer et al. 2013). In a world where a single cognitive system can contain billions of phrasal and linguistic variations of hundreds of millions of concepts, and hundreds of billions to trillions of relationships between those concepts, based on training data that may exceed what an individual person could experience in a lifetime, a better understanding of knowledge in people in networks is needed (Hidalgo 2011).

Another type of grand challenge relates to definitions and measurements. If cognitive science is the study of cognitive systems with different capabilities evolved or built on different substrates, and artificial intelligence is the study and construction of cognitive systems with different capabilities built on different substrates, how can we measure these capabilities? The term *cognitive systems* is tricky to define, because it includes instances that are natural biological systems (evolved) and artificial technological systems (built), as well as organizational aggregations in networks (evolved and built). More than mere networks, we live in a rapidly evolving ecology of nested, networked cognitive systems entities. Measurement requires comparison, and ultimately to measure advances in cognitive science and artificial intelligence, we need to make comparisons between the capabilities of specific entities, as well as networks of entities. The relationship between specific physical-symbol system entities, cognitive system entities, and service system entities is an area for future exploration (Simon 1980, Simon 1996, Spohrer and Maglio 2010).

At the same time that new grand challenges are being defined, the opportunities to rapidly prototype and deploy sophisticated cognitive systems is also on the rise. Engineering a variety of cognitive systems is getting easier due to the emergence of cognition as a service that provides access to cognitive computing componentry that can be composed and configured through higher-level APIs. For example, a question-answering service exposes an API that accepts a question as natural language text and returns a set of candidate answers with corresponding confidence levels. A machine translation service translates a natural language sentence from one language to another. An image-recognition service provides a taxonomy of recognized objects within an image. One can imagine an application that combines the above three example services to (1) recognize objects in a tourist's image, (2) find descriptive information about those objects by asking questions, and then (3) translate the description into a different language. This fairly sophisticated application can be constructed with a few lines of code using a cloud platform. For example, the IBM Watson Services on Bluemix[1] is an example of such a platform that provides services for natural language processing, image and video processing, user modeling, knowledge management, visualization, dialog management, and a variety of other cognitive capabilities. Training the above example services to new domains is a major technical challenge with cognitive systems. For example, question answering in a new domain, for example, history of America, using current-generation supervised learning techniques requires detailed annotations on a corpus of ground truth data (larger the better), and painstaking training cycles for the underlying machine-learning algorithms until the desired accuracy is achieved. Within a cognitive cloud platform, training APIs and tools for handling new domains is an ongoing area of design and optimization. Research is under way to apply transfer learning and active learning techniques to improve the productivity of the training methods. Future systems may apply large-scale learning techniques to reduce the manual supervision required for such services. While the cognitive services described above can be thought of as traditional passive software components that react to API invocation, there is another class of proactive cognitive services that may initiate various workflows based upon internal state changes — these have been called cognitive agents. An example is a planning agent that can proactively notify interested parties (people or other cognitive components) about a newly discovered better path to a goal, such as a travel destination, a shopping objective, or a business outcome.



Occupations		
Receptionists and Information Clerks (3-4171.00)	Clinical Data Managers (15-2041.02)	Physicians and Surgeons (29-1069.00
Chief Executives (11-1011.00)	Biochemical Engineers (17-2199.01)	Registered Nurses (29-1141.00)
Legislators (11-1031.00)	Biochemist and Biophysicists (19-1021.00)	Home Health Aides (31-1011.00)
Investment Fund Managers (11-9199.03)	Climate Change Analysts (19-2041.01)	Chefs and Head Cooks (35-1011.00)
Human Resources Specialists (13-1071.00)	Political Scientists (19-3094.00)	Childcare Workers (39-9011.00)
Market Research Analysts (13-1161.00)	Child, Family, and School Social Workers (21-1021.00)	Telemarketers (41-9041.00)
Business Intelligence Analysts (15-1199.08)	Human Service Assistants (21-1093.00)	Customer Service Representatives (43-4051.00)
Data Warehousing Specialists (15-1199.07)	Curators (25-4012.00)	Executive Administrative Assistants (43-6011.00)

*Figure 9. O*NET OnLine Occupations Mentioned Here.*

Compositions of such proactive agents can be constructed to support a variety of scenarios from personal assistants to professional advisors.

Conclusions

The IBM Cognitive Systems Institute Group is being established to improve industry, academic, and government focus on the opportunities to augment and scale human expertise by advancing the state of the art in cognitive computing, connecting artificial intelligence, cognitive science, and other multidisciplinary researchers (Johnson 2013). The goals of this new virtual organization include piloting new industry-university programs that (1) advance progress on next-generation grand challenges and provide access to cognition as a service platform to build next-generations skills, (2) encourage more and better grant proposals to government and foundation funding agencies aimed at understanding and prototyping cognitive assistants for important job tasks and occupations, (3) improve the design loop on human-centered smart service systems by embedding industry researchers in residence at universities, and (4) compile point-of-view (POV) documents on the potential impacts of cognitive systems on business and society to encourage more multidisciplinary collaboration on envisioning the future and reimaging work to transform professions, industries, and regions. We can ask the question "how much valuable work does not get done because of the lack of access to expertise within an organization or region?" By augmenting and scaling human expertise, industrial-strength cognition as a service will boost the creativity and productivity of people allowing more valuable work to get done — including the work of artificial intelligence and cognitive science researchers.

Of all the occupations mentioned in this article (figure 9), which is most important? The Internet pioneer Douglas E. Engelbart, who invented the computer mouse and aspects of interactive computing, had a perspective on this question. He advocated that improving our ability to improve ("double-loop learning") was often neglected by organizations. For example, industrial researchers work to improve product and service offerings of their companies, but which occupation improves the capabilities of those researchers? Cognitive science, artificial intelligence, neuroscience, and related researchers are unlocking our ability to improve improvement through cognitive assistants in the cognitive era. Unlike when the authors of this article where students of artificial intelligence, today there is a whole new level of industry-tested cognitive computing componentry on which to build. When Engelbart passed away in 2013, the world lost more than an important inventor, but a pioneer and visionary whose quest to use computers to improve improvement is in many ways just now becoming a reality. Engelbart's vision to augment human intellect to address complex and urgent problems is now at hand (Engelbart 1962, Engelbart 1995).

Acknowledgements

Thomas Dalton, Paul Hofmann, and the editors provided helpful suggestions, and any errors that remain are those of the authors. Thanks to the many IBMers, faculty, researchers, and others whose works are referenced in or influenced this article. You are helping us to dream bigger.

Notes

1. See Watson Services for Bluemix: Rapidly Prototype and Build Powerful Cognitive Apps in the Cloud. (ace.ng.bluemix.net/#/solutions/solution=watson).

2. See Brain-inspired Chips: Neuroscience Provides a Path Towards Energy Efficient Computing #IBM #IBMResearch. August 7, 2014 (twitter.com/DharmendraModha/status/497483075719147524).

3. www.onetonline.org

References

Aharoni, E.; Polnarov, A.; Lavee, T.; Heshcovich, D.; Levy, R.; Rinott, R.; Gutfreund, D.; and Slonim, N. 2014. A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics. Paper presented at the First Workshop on Argumentation Mining, 52nd Annual Meeting of the Association for Computational Linguistics (ACL), June 26, Baltimore, MD. dx.doi.org/10.3115/v1/w14-2109

Amir, A.; Datta, P.; Risk, W. P.; Cassidy, A. S.; Kusnitz, J. A.; Esser, S. K.; Andreopoulos, A.; Wong, T. M.; Flickner, M.; Alvarez-Icaza, R.; McQuinn, E.; Shaw, B.; Pass, N; and Modha, D. S. 2013. Cognitive Computing Programming Paradigm: A Corelet Language for Composing Networks of Neurosynaptic Cores. In *Neural Networks (IJCNN), The 2013 International Joint Conference,* 1–10. Piscataway, NJ: Institute of Electrical and Electronics Engineers. dx.doi.org/10.1109/ijcnn.2013.6707078

Ardire, S., and Roe, C. 2014. Cognitive Computing: An Emerging Hub in IT Ecosystems — Data Management's New Imperative. *DATAVERSITY*. August 18, 2014. (content.dataversity.net/rs/wilshireconferences/images/CC_paper.pdf)

Austin, T., and LeHong, H. 2014. Digital Business Innovation with Smart Machines. Gartner Report G00265758. August 7, 2014. Stamford, CT: Gartner Inc. (www.gartner.com/doc/2818821/digital-business-innovation-smart-machines)

Avizienis, A. V.; Sillin, H. O.; Martin-Olmos, C.; Shieh, H. H.; Aono, M.; Stieg, A. Z.; and Gimzewski, J. K. 2012. Neuromorphic Atomic Switch Networks. *PloS One* 7(8), e42772. dx.doi.org/10.1371/journal.pone.0042772

Bassett, J. 2014. Memorial Sloan Kettering Trains IBM Watson to Help Doctors Make Better Cancer Treatment Choices. April 11, 2014. New York: Memorial Sloan Kettering Cancer Center. (www.mskcc.org/blog/msk-trains-ibm-watson-help-doctors-make-better-treatment-choices)

Bilow, R. 2014. How IBM's Chef Watson Actually Works. *Bon Appetit*. June 30, 2014. (www.bonappetit.com/entertaining-style/trends-news/article/how-ibm-chef-watson-works)

Brown, E. 2013. TED: Cognitive Computing. YouTube Video. August 13, 2013. San Bruno, CA: YouTube Inc. (www.youtube.com/watch?v=np1sJ08Q7lw)

Brynjolfsson, E., and McAfee, A. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: WW Norton and Company.

Cuthbertson, A. 2014. IBM Supercomputer Watson Programmed to Debate Moral Issues. *International Business Times* May 6, 2014. (http://www.ibtimes.co.uk/ibm-supercomputer-watson-programmed-debate-moral-issues-1447413)

Engelbart, D. C. 1962. Conceptual Framework for Augmenting Human Intellect. Original SRI Summary Report AFOSR-3223. August 19, 2014. Menlo Park, CA: SRI International. (www.dougengelbart.org/pubs/augment-3906.html)

Engelbart, D. C. 1995. Toward Augmenting the Human Intellect and Boosting Our Collective IQ. *Communications of the ACM* 38(8): 30–32. dx.doi.org/10.1145/208344.208352

Erickson, T.; Danis, C. M.; Kellogg, W. A.; and Helander, M. E. 2008. Assistance: The Work Practices of Human Administrative Assistants and Their Implications for IT and Organizations. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*. 609–618. New York: Association for Computing Machinery. dx.doi.org/10.1145/1460563.1460658

Ferrucci, D.; Brown, E.; Chu-Carroll, J.; Fan, J.; Gondek, D.; Kalyanpur, A. A.; Lally, A.; Murdock, J. W.; Nyberg, E.; Prager, J.; Schlaefer, N.; and Welty, C. 2010. Building Watson: An Overview of the DeepQA Project. *AI Magazine* 31(3): 59–79.

Florida, R. 2010. *Who's Your City?: How the Creative Economy Is Making Where to Live the Most Important Decision of Your Life*. New York: Random House LLC.

Gondek, D. C.; Lally, A.; Kalyanpur, A.; Murdock, J. W.; Ariel Duboue, P.; Zhang, L.; Pan, Y.; Qiu, Z. M.; and Welty, C. 2012. A Framework for Merging and Ranking of Answers in DeepQA. *IBM Journal of Research and Development* 56(3.4): 14:1–14:12.

Harris, D. 2013. The Gigaom Guide to Deep Learning: Who's Doing It, and Why It Matters. Gigaom Research. November 1, 2013. Houston, TX: Knowingly Corp. (gigaom.com/2013/11/01/the-gigaom-guide-to-deep-learning-whos-doing-it-and-why-it-matters)

Hidalgo, C. 2011. Networks Understanding Networks: Knowledge in People in Networks @ MIT Media Lab. Video Talk (minutes 6:00-19:00). October 17, 2011. San Bruno, CA: YouTube Inc. (www.youtube.com/watch?v=mwIjcv7OWMo)

IBM. 2013. IBM Watson Engagement Advisor: Transforming the Way People and Organizations Interact over the Lifetime of Their Relationships. IBM Watson Solutions Brochure. Somers, NY: IBM Press.

Johnson, R. C. 2013. IBM Unveils Cognitive Systems Institute. *EETimes*. October 3, 2013. (www.eetimes.com/document.asp?doc_id=1319693

Kahney, H. 1993. MARCEL: Simulating the Novice Programmer, by James C. Spohrer (Book Review). *International Journal of Man-Machine Studies* 38(6): 1049–1055.

Karoly, L. A.; Kilburn, M. R.; and Cannon, J. S. 2005. Early Childhood Interventions: Proven Results, Future Promise. Report MG-341-PNC. Santa Monica, CA: Rand Corporation. September 14, 2014. (www.rand.org/publications/MG/MG341)

Kelly III, J.; and Hamm, S. 2013. *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. New York: Columbia University Press.

Kline, S. J. 1995. *Conceptual Foundations for Multidisciplinary Thinking*. Palo Alto, CA: Stanford University Press.

Lewis, W.W. 2005. *The Power of Productivity: Wealth, Poverty, and the Threat to Global Stability*. Chicago, IL: University of Chicago Press.

Maglio, P. P.; Kieliszewski, C. A.; and Spohrer, J. C. 2010. *Handbook of Service Science*. Berlin: Springer. dx.doi.org/10.1007/978-1-4419-1628-0

Marcus, G. 2014. What Comes After the Turing Test? *The New Yorker,* June 9, 2014. (www.newyorker.com/tech/elements/what-comes-after-the-turing-test)

Markoff, J. 2014. Computer Eyesight Gets a Lot More Accurate. Bits Blog, *New York Times*, August 18, 2014. (bits.blogs.nytimes.com/2014/08/18/computer-eyesight-gets-a-lot-more-accurate)

Merolla, P. A.; Arthur, J. V.; Alvarez-Icaza, R.; Cassidy, A. S.; Sawada, J.; Akopyan, F.; Jackson, B. L.; Iman, N.; Guo, C.; Nakamura, Y.; Brezz, B.; Vo, I.; Esser, S. K.; Appuswamy, R.; Taba, B.; Amir, A.; Flickner, M. D.; Risk, W. P.; Manohar, R.; and Modha, D. S. (2014) A Million Spiking-Neuron Integrated Circuit with a Scalable Communication Network and Interface. *Science* 345(6197): 668–673. dx.doi.org/10.1126/science.1254642

Mishne, G.; Carmel, D.; Hoory, R.; Roytman, A.; and Soffer, A. 2005. Automatic Analysis of Call-Center Conversations. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management,* 453–459. New York: Association for Computing Machinery. dx.doi.org/10.1145/1099554.1099684

Ng, I. C. L.; Maull, R.; Parry, G.; Crowcroft, J.; Scharf, K.; Rodden, T.; and Speed, C. 2013. Making Value Creating Context Visible for New Economic and Business Models: Home Hub-of-all-Things (H.A.T) as Platform for Multi-Sided Market Powered by Internet-of-Things. Panel Session at The Future of Value Creation in Complex Service Systems Minitrack of Hawaii International Conference on Systems Science (HICSS), January 7–10, Honolulu, HI.

Nordfors, D. 2014. How to Disrupt Unemployment. *Huffington Post*. July 25. (www.huffingtonpost.com/david-nordfors/how-innovation-can-disrup-unemployment_b_5616562.html)

NSF. 2013. Partnerships for Innovation: Building Innovation Capacity (PFI:BIC). Program Solicitation NSF 13-587. Directorate for Engineering, Industrial Innovation, and Partnerships. December 31, 2013. Washington, DC: National Science Foundation. (www.nsf.gov/pubs/2013/nsf13587/nsf13587.htm)

Pentland, A. 2014. *Social Physics: How Good Ideas Spread — The Lessons from a New Science*. London: Penguin.

Plummer, D. 2013. Top 10 Strategic Predictions: Gartner Predicts a Disruptive and Constructive Future for IT. Speech at Gartner Symposium/ITxpo, October 8, 2013, Orlando, FL. (www.gartner.com/newsroom/id/2603215)

Poladian, C. 2014. A.I. XPrize: Can a Robot Deliver a TED Talk Worthy of a Standing Ovation? *International Business Times,* March 23, 2014. (www.ibtimes.com/ai-xprize-can-robot-deliver-ted-talk-worthy-standing-ovation-1563109)

Popescu-Belis, A.; Boertjes, E.; Kilgour, J.; Poller, P.; Castronovo, S.; Wilson, T.; Jaimes, A.; and Carletta, J. 2008. The AMIDA Automatic Content Linking Device: Just-in-Time Document Retrieval in Meetings. In *Machine Learning for Multimodal Interaction,* Lecture Notes in Computer Science Volume 5237, 272–283. Berlin: Springer. dx.doi.org/10.1007/978-3-540-85853-9_25

Pusiol, G.; Soriano, L.; Fei-Fei, L.; and Frank, M. C. 2014. Discovering the Signatures of Joint Attention in Child-Caregiver Interaction. In *Proceedings of the 36th Annual Cognitive Science Conference*. Wheat Ridge, CO: Cognitive Science Society, Inc. (vision.stanford.edu/pdf/guido14.pdf)

Ramanathan, V.; Joulin, A.; Liang, P.; and Fei-Fei, L. 2014. Linking People in Videos with "Their" Names Using Coreference Resolution. In *Computer Vision: Proceedings of the European Conference on Computer Vision 2014,* 95–110. Berlin: Springer. dx.doi.org/10.1007/978-3-319-10590-1_7

Roy, D. 2011. Why I Taped My Son's Childhood. CNN Opinion: TED Talk Tuesday. CNN. March 13, 2011. Atlanta, GA: Cable News Newwork LP, LLLP. (www.cnn.com/2011/OPINION/03/13/roy.tapes.childhood) dx.doi.org/10.1007/11880172_15

Roy, D.; Patel, R.; DeCamp, P.; Kubat, R.; Fleischman, M.; Roy, B.; Mavridis, N.; Tellex, S.; Salata, A.; Guinness, J.; Levit, M.; and Gorniak, P. 2006. The Human Speechome Project. In *Proceedings of the 28th Annual Cognitive Science Conference,* 192–196. Wheat Ridge, CO: Cognitive Science Society, Inc.

Service, R. F. 2014. The Brain Chip. *Science* 345(6197): 614–616. dx.doi.org/10.1126/science.345.6197.614

Simon, H. A. 1980. Cognitive Science: The Newest Science of the Artificial *Cognitive Science* 4(1): 33–46.

Simon, H. A. 1996. *The Sciences of the Artificial,* 3rd edition. Cambridge, MA: MIT Press.

Simonite, T. 2014. A Room Where Executives Go to Get Help from IBM's Watson. *MIT Technology Review*. August 4, 2014. (www.technologyreview.com/news/529606/a-room-where-executives-go-to-get-help-from-ibms-watson)

Simonite, T. 2013. Software Mines Science Papers to Make New Discoveries. *MIT Technology Review,* November 25,

AAAI to Cosponsor the 25th International Joint Conference on Artificial Intelligence (IJCAI-16)!

Join Us in New York City!

July 9–16, 2016
ijcai-16.org

Join us at the New York Hilton Midtown Hotel for IJCAI-16. The Hilton is just a ten-minute walk from Central Park, where the conference banquet will be held, and half a block from the world-famous Museum of Modern Art, where the opening reception will kick off the main technical conference on July 11. Located in the heart of New York City in Midtown Manhattan, this hotel is within walking distance from New York's premier attractions such as Times Square, Radio City Music Hall, Fifth Avenue shopping, the Broadway Theatre District, and many more iconic New York landmarks — an extra reason to join us and to add a couple of days to your visit.

Important Upcoming Dates

Abstract deadline: 1/27
Paper deadline: 2/2
Rebuttal period: 3/10-11
Notification: 4/4
Camera Ready: 4/20

Conference Overview

Saturday 7/9 – Monday 7/11 Workshops and Tutorials
Monday 7/11 Reception (Museum of Modern Art)
Tuesday 7/12 - Friday 7/15 Technical Program
Wednesday 7/13 Banquet (Central Park Zoo)

2013. (m.technologyreview.com/news/520461/software-mines-science-papers-to-make-new-discoveries)

Smith, A., and Anderson, J. 2014. AI, Robotics, and the Future of Jobs. Pew Research Internet Project. August 6, 2014. Washington, DC: Pew Research Center. (www.pewinternet.org/2014/08/06/future-of-jobs)

Spivak, N. 2013. Why Cognition as a Service Is the Next Operating System Battlefield. Gigaom Research. December 7, 2013. Houston, TX: Knowingly Corp. (gigaom.com/2013/12/07/why-cognition-as-a-service-is-the-next-operating-system-battlefield)

Spohrer, J. C. 1992. *MARCEL: Simulating the Novice Programmer*. Bristol, UK: Intellect Books. dx.doi.org/10.1007/978-1-4419-1628-0_9

Spohrer, J. C., and Maglio, P. P. 2010. Toward a Science of Service Systems. In *Handbook of Service Science,* ed. P. Maglio, C. Kieliszewski, and J. Spohrer, 157–194. Berlin: Springer. dx.doi.org/10.1287/serv.1120.0012

Spohrer, J.; Giuiusa, A.; Demirkan, H.; and Ing, D. 2013. Service Science: Reframing Progress with Universities. *Systems Research and Behavioral Science* 30(5): 561–569.

Spohrer, J.; Piciocchi, P.; and Bassano, C. 2012. Three Frameworks for Service Research: Exploring Multilevel Governance in Nested, Networked Systems. *Service Science* 4(2): 147–160. dx.doi.org/10.1002/sres.2213

Ting, C.; Kumar, R.; Axel-Troianowski, G.; and Syeda-Mahmood, T. 2013. PSAR: Predictive Space Aggregated Regression and Its Application in Valvular Heart Disease Classification. In *IEEE International Symposium on Biomedical Imaging,* 1122–1125. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Varshney, L. R.; Pinel, F.; Varshney, K. R.; Schorgendorfer, A.; and Chee, Y. M. 2013. Cognition as a Part of Computational Creativity. In *Proceedings of the 12th IEEE Internatinal Conference on Cognitive Informatics and Cognitive Computting,* 36–42. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Walker, M. 2014. BIG and Technological Unemployment: Chicken Little Versus the Economists. *Journal of Evolution and Technology* 24(1): 5–25. (jetpress.org/v24/walker.htm)

Jim Spohrer is director of IBM Global University Programs and leads IBM's Cognitive Systems Institute Group. He received his Ph.D. in computer science and artificial intelligence from Yale University and a BS in physics from the Massachusetss Institute of Technology. Spohrer was founding CTO of IBM's Venture Capital Relations Group in Silicon Valley, cofounded IBM's first Service Research group and the International Society of Service Innovation Professionals (ISSIP.org). While at Apple Computer in the 1990s, he was recognized as a distinguished engineer scientist and technologist for his work on next-generation learning platforms. Today, his research priorities include service science, including human-centered smart service systems in which teams of T-shaped professionals augmented by next-generation cognitive systems innovate. With more than 90 publications and nine patents, he is also a PICMET Fellow.

Guruduth Banavar is vice president of cognitive computing at IBM Research, leading a worldwide team responsible for creating the next generation of cognitive systems in the Watson family. Previously, as CTO of IBM's Global Public Sector business, Banavar focused on the Smarter Cities initiative. Earlier, Banavar was the director of IBM Research in India, which he helped establish as a preeminent center for Services Research and Mobile Computing. He and his team received a National Innovation Award by the President of India in 2009 for the Spoken Web project. His early work was on distributed systems and programming models at IBM's TJ Watson Research Center in New York, which he joined in 1995 after his Ph.D. in computer science. His work has been featured in *The New York Times, The Wall Street Journal, The Economist,* and other international media. Banavar is an elected member of the IBM Academy of Technology and has served on the boards of IBM's Industry Academy and those of client and academic institutions.