DOI:10.1145/2063176.2063177

Moshe Y. Vardi

Artificial Intelligence: Past and Future

Chess fans remember many dramatic chess matches in the 20th century. I recall being transfixed by the 1972 interminable match between challenger Bobby Fischer and defending champion Boris Spassky for the World Chess Championship. The most dramatic chess match of the 20th century was, in my opinion, the May 1997 rematch between the IBM supercomputer Deep Blue and world champion Garry Kasparov, which Deep Blue won 3½–2½.

I was invited by IBM to attend the rematch. I flew to New York City to watch the first game, which Kasparov won. I was swayed by Kasparov's confidence and decided to go back to Houston, missing the dramatic second game, in which Kasparaov lost—both the game and his confidence.

While this victory of machine over man was considered by many a triumph for artificial intelligence (AI), John McCarthy (Sept. 4, 1927–Oct. 24, 2011), who not only was one of the founding pioneers of AI but also coined the very name of the field, was rather dismissive of this accomplishment. "The fixation of most computer chess work on success in tournament play has come at scientific cost," he argued. McCarthy was disappointed by the fact that the key to Deep Blue's success was its sheer compute power rather than a deep understanding, exhibited by expert chess players, of the game itself.

AI's next major milestone occurred last February with IBM's Watson program winning a "Jeopardy!" match against Brad Rutter, the biggest all-time money winner, and Ken Jennings, the record holder for the longest championship streak. This achievement was also dismissed by some. "Watson doesn't know it won on "Jeopardy!"," argued the philosopher John Searle, asserting that "IBM invented an ingenious program, not a computer that can think."

In fact, AI has been controversial from its early days. Many of its early pioneers overpromised. "Machines will be capable, within 20 years, of doing any work a man can do," wrote Herbert Simon in 1965. At the same time, AI's accomplishments tended to be underappreciated. "As soon as it works, no one calls it AI anymore," complained McCarthy. Yet it is recent worries about AI that indicate, I believe, how far AI as come.

In April 2000, Bill Joy, the technologists' technologist, wrote a "heretic" article entitled "Why the Future Doesn't Need Us" for *Wired* magazine, "Our most powerful 21st-century technologies—robotics, genetic engineering, and nanotech—are threatening to make humans an endangered species," he wrote. Joy's article was mostly ignored, but in August 2011 Jaron Lanier, another widely respected technologist, wrote about the impact of AI on the job market. In the not-too-far future, he predicted, it would just be inconceivable to put a person behind the wheel of a truck or a cab. "What do all those people do?" he asked.

Slate magazine ran a series of articles in September 2011 titled "Will Robots Steal Your Job?" According to writer Farhad Manjoo, who detailed the many jobs we can expect to see taken over by computers and robots in the coming years, "You're highly educated. You make a lot of money. You should still be afraid."

In fact, worries about the impact of technology on the job market are not only about the far, but also the not too far future. In a recent book, R*ace Against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*, by Erik Brynjolfsson and Andrew McAfee, the authors argue that "technological progress is accelerating innovation even as it leaves many types of workers behind." Indeed, over the past 30 years, as we saw the personal computer morph into tablets, smartphones, and cloud computing, we also saw income inequality grow worldwide. While the loss of millions of jobs over the past few years has been attributed to the Great Recession, whose end is not yet in sight, it now seems that technology-driven productivity growth is at least a major factor.

The fundamental question, I believe, is whether Herbert Simon was right, even if his timing was off, when he said "Machines will be capable ... of doing any work a man can do." While AI has been proven to be much more difficult than early pioneers believed, its inexorable progress over the past 50 years suggests that Simon may have been right. Bill Joy's question, therefore, deserves not to be ignored. *Does the future need us?*

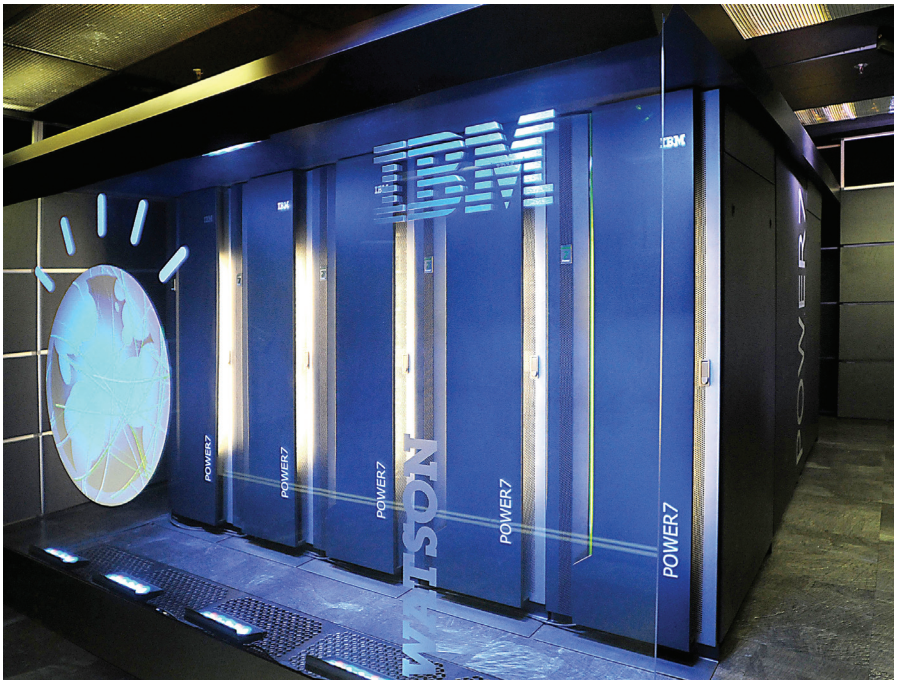
Moshe Y. Vardi, EDITOR-IN-CHIEF

Science | DOI:10.1145/1965724.1965730 Kirk L. Kroeker

Weighing Watson's Impact

Does IBM's Watson represent a distinct breakthrough in machine learning and natural language processing or is the 2,880-core wunderkind merely a solid feat of engineering?

In the history of speculative fiction, from the golden age of science fiction to the present, there are many examples of artificial intelligences engaging their interlocutors in dialogue that exhibits self-awareness, personality, and even empathy. Several fields in computer science, including machine learning and natural language processing, have been steadily approaching the point at which real-world systems will be able to approximate this kind of interaction. IBM's Watson computer, the latest example in a long series of efforts in this area, made a television appearance earlier this year in a widely promoted human-versus-machine "Jeopardy!" game show contest. To many observers, Watson's appearance on "Jeopardy!" marked a milestone on the path toward achieving the kind of sophisticated, knowledge-based interaction that has traditionally been relegated to the realm of fiction.



IBM's Watson soundly defeated the two most successful contestants in the history of the game show "Jeopardy!," Ken Jennings and Brad Rutter, in a three-day competition in February.

The "Jeopardy!" event, in which Watson competed against Ken Jennings and Brad Rutter, the two most successful contestants in the game show's history, created a wave of coverage across mainstream and social media. During the three-day contest in February, hints of what might be called Watson's quirky personality shone through, with the machine wagering oddly precise amounts, guessing at answers after wildly misinterpreting clues, but ultimately prevailing against its formidable human opponents.

Leading up to the million-dollar challenge, Watson played more than 50 practice matches against former "Jeopardy!" contestants, and was required to pass the same tests that humans must take to qualify for the show and compete against Jennings, who broke the "Jeopardy!" record for the most consecutive games played, resulting in winnings of more than $2.5 million, and Rutter, whose total winnings amounted to $3.25 million, the most money ever won by a single "Jeopardy!" player. At the end of the three-day event, Watson finished with $77,147, beating Jennings, who had $24,000, and Rutter, who had $21,600. The million-dollar prize money awarded to Watson went to charity.

Named after IBM founder Thomas J. Watson, the Watson system was built by a team of IBM scientists whose goal was to create a standalone platform that could rival a human's ability to answer questions posed in natural language. During the "Jeopardy!" challenge, Watson was not connected to the Internet or any external data sources. Instead, Watson operated as an independent system contained in several large floor units housing 90 IBM Power 750 servers with a total of 2,880 processing cores and 15 terabytes of memory. Watson's technology, developed by IBM and several contributing universities, was guided by principles described in the Open Advancement of Question-Answering (OAQA) framework, which is still operating today and facilitating ongoing input from outside institutions.

Judging by the sizeable coverage of the event, Watson piqued the interest of technology enthusiasts and the general public alike, earning "Jeopardy!" the highest viewer numbers it had achieved in several years and leading to analysts and other industry observers speculating about whether Watson represents a fundamental new idea in computer science or merely a solid feat of engineering. Richard Doherty, the research director at Envisioneering Group, a technology consulting firm based in Seaford, NY, was quoted in an Associated Press story as saying that Watson is "the most significant breakthrough of this century."

Doherty was not alone in making such claims, although the researchers on the IBM team responsible for designing Watson have been far more modest in their assessment of the technology they created. "Watson is a novel approach and a powerful architecture," says David Ferrucci, director of the IBM DeepQA research team that created Watson. Ferrucci does characterize Watson as a breakthrough in artificial intelligence, but he is careful to qualify this assertion by saying that the breakthrough is in the development of artificial-intelligence systems.

"The breakthrough is how we pulled everything together, how we integrated natural language processing, information retrieval, knowledge representation, machine learning, and a general reasoning paradigm," says Ferrucci. "I think this represents a breakthrough. We would have failed had we not invested in a rigorous scientific method and systems engineering. Both were needed to succeed."

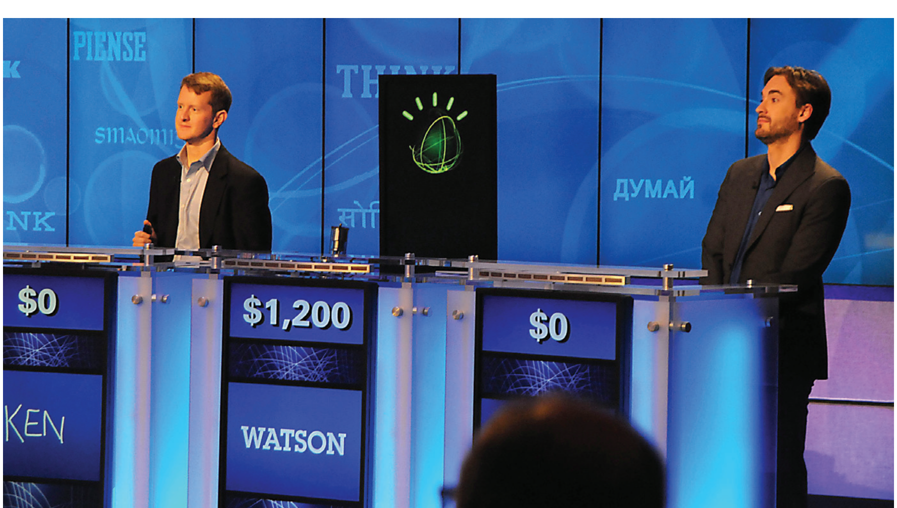
Contextual Evidence

The DeepQA team was inspired by several overarching design principles, with the core idea being that no single algorithm or formula would accurately understand or answer all questions, says Ferrucci. Rather, the idea was to build Watson's intelligence from a broad collection of algorithms that would probabilistically and imperfectly interpret language and score evidence from different perspectives. Watson's candidate answers, those answers in which Watson has the most confidence, are produced from hundreds of parallel hypotheses collected and scored from contextual evidence.

Ferrucci says this approach required innovation at the systems level so individual algorithms could be developed independently, then evaluated for their contribution to the system's overall performance. The approach allowed for loosely coupled interaction between algorithm components, which Ferrucci says ultimately reduced the need for team-wide agreement. "If every algorithm developer had to agree with every other or reach some sort of consensus, progress would have been slowed," he says. "The key was to let different members of the team develop diverse algorithms independently, but regularly perform rigorous integration testing to evaluate relative impact in the context of the whole system."

Ferrucci and the DeepQA team are expected to release more details later this year in a series of papers that will outline how they dealt with specific aspects of the Watson design. For now, only bits and pieces of the complete picture are being disclosed. Ferrucci says that, looking ahead, his team's research agenda is to focus on how Watson can understand, learn, and interact more effectively. "Natural language understanding remains a tremendously difficult challenge, and while Watson demonstrated a powerful approach, we have only scratched the surface," he says. "The challenge continues to be about how you build systems to accurately connect language to some representation, so the system can automatically learn from text and then reason to discover evidence and answers."

Lillian Lee, a professor in the computer science department at Cornell University, says the reactions about Watson's victory echo the reactions following Deep Blue's 1997 victory over chess champion Garry Kasparov, but with several important differences. Lee, whose research focus is natural



Watson's on-stage persona simulates the system's processing activity and relative answer confidence through moving lines and colors. Watson is shown here in a practice match with Ken Jennings, left, and Brad Rutter at IBM's Watson Research Center in January.

PHOTO COURTESY IBM

language processing, points out that some observers were dismissive about Deep Blue's victory, suggesting that the system's capability was due largely to brute-force reasoning rather than machine learning. The same criticism, she says, cannot be leveled at Watson because the overall system needed to determine how to assess and integrate diverse responses.

"Watson incorporates machine learning in several crucial stages of its processing pipeline," Lee says. "For example, reinforcement learning was used to enable Watson to engage in strategic game play, and the key problem of determining how confident to be in an answer was approached using machine-learning techniques, too."

Lee says that while there has been substantial research on the particular problems the "Jeopardy!" challenge involved for Watson, that prior work should not diminish the team's accomplishment in advancing the state of the art to Watson's championship performance. "The contest really showcased real-time, broad-domain question-answering, and provided as comparison points two extremely formidable contestants," she says. "Watson represents an absolutely extraordinary achievement."

Lee suggests that with language-processing technologies now maturing, with the most recent example of such maturation being Watson, the field appears to have passed through an important early stage. It now faces an unprecedented opportunity in helping sift through the massive amounts of user-generated content online, such as opinion-oriented information in product reviews or political analysis, according to Lee.

While natural-language processing is already used, with varying degrees of success, in search engines and other applications, it might be some time before Watson's unique question-answering capabilities will help sift through online reviews and other user-generated content. Even so, that day might not be too far off, as IBM has already begun work with Nuance Communications to commercialize the technology for medical applications. The idea is for Watson to assist physicians and nurses in finding information buried in medical tomes, prior cases, and the latest science journals. The first commercial offerings from the collaboration are expected to be available within two years.

"Natural language understanding remains a tremendously difficult challenge, and while Watson demonstrated a powerful approach, we have only scratched the surface," says David Ferrucci.

Beyond medicine, likely application areas for Watson's technology would be in law, education, or the financial industry. Of course, as with any technology, glitches and inconsistencies will have to be worked out for each new domain. Glitches notwithstanding, technology analysts say that Watson-like technologies will have a significant impact on computing in particular and human life in general. Ferrucci, for his part, says these new technologies likely will mean a demand for higher-density hardware and for tools to help developers understand and debug machine-learning systems more effectively. Ferrucci also says it's likely that user expectations will be raised, leading to systems that do a better job at interacting in natural language and sifting through unstructured content.

To this end, explains Ferrucci, the DeepQA team is moving away from attempting to squeeze ever-diminishing performance improvements out of Watson in terms of parsers and local components. Instead, they are focusing on how to use context and information to evaluate competing interpretations more effectively. "What we learned is that, for this approach to extend beyond one domain, you need to implement a positive feedback loop of extracting basic syntax and local semantics from language, learning from context, and then interacting with users and a broader community to acquire knowledge that is otherwise difficult to extract," he says. "The system must be able to bootstrap and learn from its own failing with the help of this loop."

In an ideal future, says Ferrucci, Watson will operate much like the ship computer on "Star Trek," where the input can be expressed in human terms and the output is accurate and understandable. Of course, the "Star Trek" ship computer was largely humorless and devoid of personality, responding to queries and commands with a consistently even tone. If the "Jeopardy!" challenge serves as a small glimpse of things to come for Watson—in particular, Watson's precise wagers, which produced laughter in the audience, and Watson's visualization component, which appeared to express the state of a contemplative mind through moving lines and colors—the DeepQA team's focus on active learning might also include a personality loop so Watson can accommodate subtle emotional cues and engage in dialogue with the kind of good humor reminiscent of the most personable artificial intelligences in fiction.

Further Reading

Baker, S.
Final Jeopardy: Man vs. Machine and the Quest to Know Everything. **Houghton Mifflin Harcourt, New York, NY, 2011.**

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.
Building Watson: An overview of the DeepQA project, *AI Magazine 59*, Fall 2010.

Ferrucci, D., et al.
Towards the Open Advancement of Question Answering Systems. *IBM Research Report RC24789 (W0904-093)*, April 2009.

Simmons, R.F.
Natural language question-answering systems, *Communications of the ACM 13*, 1, Jan. 1970.

Strzalkowski, T., and Harabagiu, S. (Eds.)
Advances in Open Domain Question Answering. **Springer-Verlag, Secaucus, NJ, 2006.**

Based in Los Angeles, **Kirk L. Kroeker** is a freelance editor and writer specializing in science and technology.

N news

Science | DOI:10.1145/2063176.2063182 Neil Savage

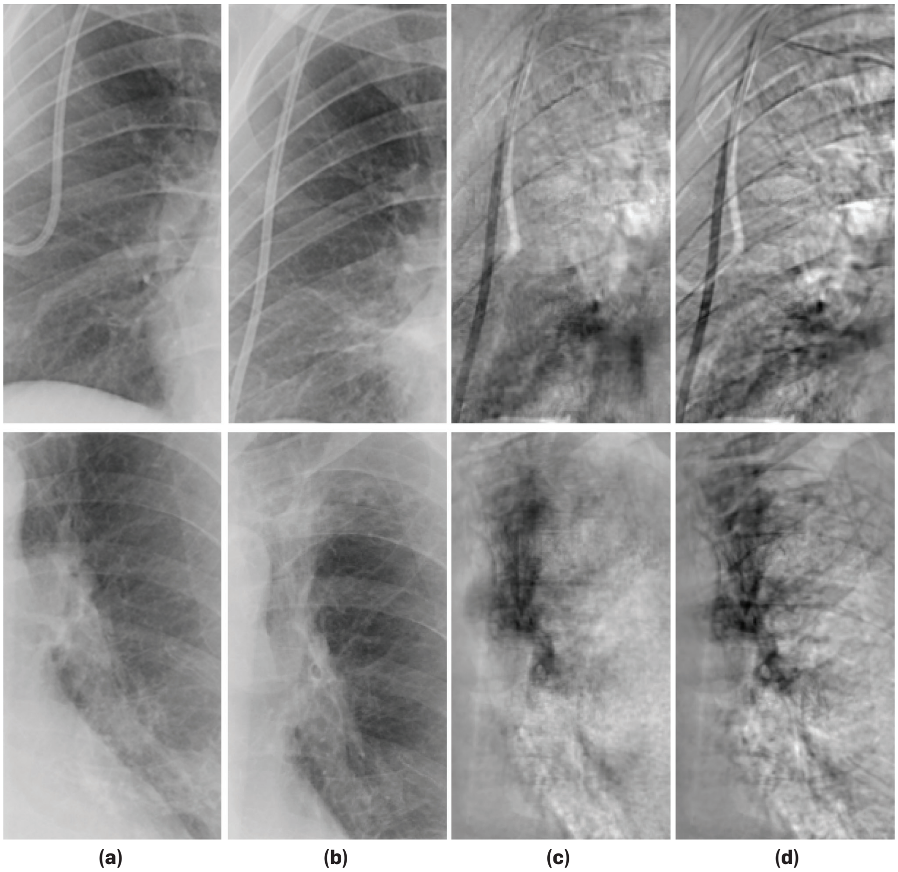
Better Medicine Through Machine Learning

Computers that tease out patterns from clinical data could improve patient diagnosis and care.

MEDICINE CAN BE as much art as science, a detective story in which doctors rely not only on lab tests and x-rays, but on their own experience and clues from a patient's history to develop diagnoses or predict future health problems. But all of those lab tests, blood pressure readings, magnetic resonance imaging (MRI) scans, electrocardiograms, and billing codes add up to reams of data, which before too long will be joined by individual gene sequencing. Computer scientists are increasingly applying machine learning techniques to all that data, searching for patterns that can aid diagnosis and improve clinical care.

"Machine learning plays, I think, an essential role in medical image analysis nowadays," says Kenji Suzuki, assistant professor of radiology and medical physics at the University of Chicago's Comprehensive Cancer Center. Suzuki has been working on automating the detection of cancerous lesions in images from x-rays or computed tomography scans. Considering that radiologists may miss 12%–30% of lung cancers in such scans, a machine learning tool offers great potential.

Since the mid-1980s, computer sci-



Kenji Suzuki and colleagues' comparison of their rib-suppressed temporal-subtraction (TS) images with conventional TS images: (a) previous chest radiographs, (b) current chest radiographs of the same patient, (c) rib-suppressed TS images with fewer rib artifacts, and (d) conventional TS images.

IMAGE COURTESY OF KENJI SUZUKI

entists have tried to improve on that performance using feature-based machine learning in which the computer would pick out morphological features, texture differences, and more to identify abnormal tissue. But such cataloging of features still misses some cancers that doctors are able to spot with their own eyes. Sometimes the features the computer is seeking can be subtle or overlap with normal anatomical structures such as bone, making them more difficult to spot. So Suzuki asks the machine to instead focus on the values, such as intensity, of individual pixels. "Because the computing power has increased dramatically in recent years, we can process the pixel values directly," he says. The resulting system is highly sensitive, achieving up to 97% accuracy.

But one concern is making the program so sensitive that it starts finding nonexistent lesions. In Suzuki's lung cancer tests, the feature-based algorithm falsely identified five lesions per patient, while the pixel-based method produced less than one false positive per patient and no false negatives. Suzuki says each method produces a different type of false positive, so combining the two approaches leads to the most accurate outcome. Suzuki is now working on expanding the technique to other types of cancer and other imaging methods, such as MRI and PET. "You just need to train the machine learning technique with new images," he says.

As the use of electronic medical records gains acceptance, machine learning is likely to play an even larger role in clinical medicine.

A Guide for Diagnosis

At IBM Almaden Research Center, the Advanced Analytics for Information Management (AALIM) project applies machine learning to a wide variety of data—readings of vital signs, tests such as echocardiograms, and demographic information—to chart the medical histories of patients over several years. By comparing the history of hundreds or thousands of people, the system can identify previous patients who are similar to a current patient, then apply collaborative filtering to suggest the best diagnosis and treatment options for the new patient.

"With a large number of pre-diagnosed patient datasets available in electronic health records, physicians can now benefit from the opinion of their peers on cases similar to their patients," says Tanveer Syeda-Mahmood, head of IBM's Multimodal Mining for Healthcare project. The hope is that by helping doctors base their decisions on quantitative information, the number of diagnostic errors can be reduced. AALIM, Syeda-Mahmood says, provides "a holistic view of the patient's condition," producing one-page summaries, long-term profiles of various measurements of health, and detailed comparisons showing diagnosis, treatment, and outcomes for similar patients.

The computer might, for instance, help a relatively new doctor decide that she needs to consult a specialist. Approximately 5% of cases, says Syeda-Mahmood, are ambiguous enough that even senior clinicians ask other doctors for their opinions, which AALIM easily provides. In emergency rooms, it can cut the time doctors spend flipping through charts by half. Although researchers have tested the system on patient data, it would likely require approval from the Food and Drug Administration before it could be used in a hospital.

Finding ways to use all this medical data often requires new developments in machine learning. For instance, Syeda-Mahmood wanted to give the

Education

Tech-Rich Learning Environments

What's a solution to classrooms plagued by students' low retention rates, high withdrawals, and failing grades? Toss out traditional lectures and create a technology-rich environment, say researchers at Rochester Institute of Technology (RIT).

Data from six years of research involving 500-plus undergraduates taking three different engineering courses revealed that 90% of the students said they learned and retained information better when a combination of tablet PCs, collaborative software, and multiple projection screens were used as teaching tools.

The study targeted three foundation courses for engineering-technology degree programs—pneumatics and hydraulics, applied dynamics, and applied fluid mechanics—all of which suffered from higher-than-average student withdrawals and low retention rates. In one class, for example, 27% of the students had previously received low or failing grades and had to repeat the course or withdraw from it.

"The students were just not getting the material," says Robert Garrick, associate professor at RIT's College of Applied Science and Technology. "They weren't understanding the intended learning objectives."

In a traditional laboratory setting, students work at different workstations while the instructor walks around, answers questions, and reviews circuit operations.

For the study, RIT researchers redesigned the courses to include a tablet PC and DyKnow collaborative software for each student. Classes took place in an interactive technology classroom that featured a multiscreen display connected to faculty and student tablet PCs. Class information and any notations were captured, recorded, and archived.

The students told the researchers they preferred the tablets for note taking, in-class work, test preparation, and classroom layout.

The combination of the technology resources improved the visual connection to the material, student-faculty interaction, and it enhanced the modeling of engineering problems, three areas seen as critical to retaining the technical information, explains RIT assistant professor Larry Villasmil.

The next step for the researchers is to study how tech-rich learning environments can work for underrepresented groups, such as deaf and hard-of-hearing students, who make up about 10% of the students in engineering and technology courses.

—Paul Hyman

computer a doctor's ability to recognize some types of heart disease by the characteristic shape of waves produced by an electrocardiogram. She developed a new function, called a constrained non-rigid translation transform, which could identify the similarity between shapes in different ECG readouts.

The Power of Prediction

Although diagnosis and treatment are key aspects of medical care, prediction is also important, especially when it can lead to early interventions. We all know, for instance, that factors such as weight and blood pressure can give an idea of a person's risk of heart disease. But those sorts of risk scores are based on population-wide models, says Shyam Visweswaran, assistant professor of biomedical informatics at the University of Pittsburgh. "If you build a model from a group of people who are kind of similar to the current patient, you might do better," he says. Visweswaran has developed an algorithm that lets a computer use clinical data to learn a model tailored to one specific patient and predict outcomes for that person.

The computer takes all the data it has on the patient, such as age, blood pressure, and lab results, and then picks one variable and builds a model of all the patients in its database who share that variable. It could, for instance, compare everyone in the 50–55 age group.

"With a large number of pre-diagnosed patient datasets available in electronic health records, physicians can now benefit from the opinion of their peers on cases similar to their patients," says Tanveer Syeda-Mahmood.

It builds a model for each variable it can find, looks at which ones best fit the patient at hand, and then averages the best models to make a personalized prediction of that patient's outcome. Whereas a population model only uses a handful of variables considered to be the best—it could be a simple checklist of several risk factors, for instance—this approach can potentially use any of hundreds of variables. One additional advantage is the machine might identify some factor that is predictive, but that medical science was not previously aware of, opening up new areas for research, Visweswaran says.

As with Suzuki's pixel-based processing, this is another machine learning method that has benefited from the growth of processing power. As recently as five years ago it might have taken a half-hour to build all these models. Today it takes less than a minute, so the machine can guide diagnoses in real time during patient visits.

This approach could help predict an intensive-care patient's risk of an infection spreading to other organs, which is a notoriously difficult task, and lead to earlier or more aggressive treatment. It might help doctors decide, for instance, which pneumonia patients need to be admitted to the hospital and which patients could be sent home with antibiotics.

In the informatics program at Children's Hospital Boston, assistant professor Ben Reis and his colleagues are working on predicting a patient's future diagnoses years in advance. They have developed Bayesian models that they call Intelligent Histories, which comb through the standard diagnostic codes used for billing, to find patterns in a patient's history that predict risk. In their first application of the work, they discovered they could identify patients at risk of domestic abuse as much as two years before the doctors seeing those patients first discovered the problem.

Doctors are supposed to screen patients for domestic abuse, but often miss it until the problem becomes acute, says Reis. Not only can the computer aid in screening for known signs of abuse, it also picked up other diagnostic codes in the test that had not been thought of as predictive, such as infections, which might teach doctors something about domestic abuse.

The Children's Hospital Boston team is working to expand their modeling to other types of diagnoses. At the same time, they are refining the machine learning itself. For instance, "we're trying to quantify how the quality of the data that goes into the model affects the results that come out," says Reis.

As the world moves to a greater use of electronic medical records, machine learning is likely to play an even larger role in clinical medicine, researchers predict. Visweswaran says genetic data, in particular, is going to require complicated computational models if it is going to be of value. Soon, experts expect, the cost of gene sequencing will drop to the point that individual genomes will become part of people's medical records, and will be available to the same data mining and pattern recognition approaches being applied to other data.

Genetic data will be too complicated and too voluminous to be handled with old-fashioned charting systems. "You have to have computational tools to query this data," Visweswaran says. "There's no way it can be done on paper."

Further Reading

Gruhl, D., et al.
Aalim: A cardiac clinical decision support system powered by advanced multi-modal analytics, International Medical Informatics Conference, Cape Town, South Africa, Sept. 12–15, 2010.

Reis, B.Y., Kohane, I.S., and Mandl, K.D.
Longitudinal histories as predictors of future diagnoses of domestic abuse: modeling study, *British Medical Journal,* 339, Sept. 2009.

Syeda-Mahood, T., Beymer, D., and Wang, F.
Shape-based matching of ECG recordings, 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 23–27, 2007, Lyon, France.

Suzuki K., Zhang J., and Xu, J.
Massive-training artificial neural network coupled with Laplacian-eigen function-based dimensionality reduction for computer-aided detection of polyps in CT colonography, *IEEE Transactions on Medical Imaging 29*, 11, Nov. 2010.

Visweswaran, S., Angus, D.C., Hsieh, M., Weissfeld, L., Yealy, D., and Cooper, G.F.
Learning patient-specific predictive models from clinical data, *Journal of Biomedical Informatics 43*, 5, Oct. 2010.

Neil Savage is a science and technology writer based in Lowell, MA.

Technology | DOI:10.1145/1721654.1721662

Gary Anthes

Robots Gear Up for Disaster Response

After 15 years of research, robots for search and rescue may be nearing prime time.

ON JANUARY 17, 1995, an earthquake near Kobe, Japan killed approximately 6,434 people, caused the collapse of 200,000 buildings, and resulted in $102.5 billion in damages. Three months later a truck bomb exploded outside a federal government building in Oklahoma City, OK, and claimed 168 lives and damaged or destroyed 324 buildings within a 16-block radius. While it was an awfully memorable year for disasters, both of these tragedies set in motion a flurry of research in robotics that observers say could save countless lives in future disasters.

Indeed, developers of search and rescue robots say the technology—which spans such diverse disciplines as artificial intelligence, sensing, communications, materials, and mechanical engineering—is nearly ready for deployment. Applications could include search, reconnaissance and mapping, removing or shoring up rubble, delivery of supplies, medical treatment, and evacuation of casualties.

However, a host of technical challenges remain. Also, researchers are concerned that a lack of standards, scarce federal funding, and tepid interest from companies that don't yet see a big market for robotic rescuers stand in the way of the miniaturization, device hardening, and systems integration that are needed to make the technology mature.

Only one emergency response team in the U.S.—New Jersey Task Force One—so far owns a robot. And the robots tested by researchers in a handful of disasters in recent years have produced decidedly mixed performances. "We still don't know how to use these things," says Robin Murphy, a professor of computer science and engineering and director of the Center for Robot-Assisted Search and Rescue at Texas A&M University. "Real disasters are infrequent, and every one is different. The robots never get used exactly the way you think they will, and they keep uncovering new bottlenecks and problems. So it's an emerging technology."

A camera robot being inserted in a bore hole at Crandall Canyon Mine in Utah.

PHOTOGRAPH COURTESY OF THE CENTER FOR ROBOT-ASSISTED SEARCH AND RESCUE

Murphy says the devices are often tested in unrealistically robot-friendly labs or via simulations that don't quite duplicate the realities of real-life situations that involve dirt and sand, steep changes in elevation, or radio-blocking metal structures. Some of the most vexing problems seem simple yet remain frustratingly intractable. For example, at the Crandall Canyon Mine in Utah, where six miners and three rescue workers were killed in 2007, mud greatly hindered the effectiveness of the workers' camera robot. "We steered the robot to places where water was dripping and turned it face-up to rinse off some of the mud," Murphy says. However, the camera robot was eventually trapped by a rock slide, causing the robot's tether to snap and for it to be lost.

Howie Choset, associate professor of robotics at Carnegie Mellon University (CMU), specializes in snake robots, which are thin, legless devices with multiple joints. They can go places more traditional, track- or wheel-propelled robots can't, but the technology still needs work. "My last trial at a rubble pile in Texas didn't go so well," Choset notes. "They didn't get over little obstacles I thought they should have. Our control laws are still not well defined; we don't have good feedback; we don't have enough sensing in the robots; and their skins have to be better designed."

So while his mechanical snakes can perform remarkable feats such as crawling up the interior of a vertical pipe or swimming across a pool, Choset says a lack of funding stands in the way of making the snakes truly versatile and robust. Choset needs, for instance, to develop more mechanical snake gaits. He'd like his snakes to be able to change from a vertical undulating gait to a sidewinder gait on command or, better yet, to autonomously switch gaits to suit new conditions. And he'd like the mechanical snake to know how to execute one gait in its front segments and a different gait at the rear segments. "We have developed the greatest variety of snake gaits in the world," says Choset, "but a rubble pile has that many more situations than we can anticipate."

Asked if further animal study would help, Choset replies, "It's true you are inspired by biology, but snakes have 200 bones and the snake robot has just 15 links. Snakes have a material called muscles, but we are not going to be making muscles any time soon." And, he adds, snakes have marvelous sensors for heat and pressure in their skins, something else technology has yet to easily match.

Miniaturization, device hardening, and systems integration are all needed for the maturing of search and rescue robots.

Three Levels of Challenges

Users of search and rescue robots face challenges at three levels, says Sanjiv Singh, a research professor at CMU's Robotics Institute. At the lowest level lies information processing—getting and managing information about the environment. At the next level comes mobility—getting the robot to where it is needed. And at the highest level comes manipulation—enabling the robot to perform the appropriate physical task once it is in place. Singh's Ember project, partially funded by the U.S. National Science Foundation, seeks to aid first responders at the first two levels, and in situations that are dynamic, chaotic, and often providing poor visibility.

Inside a burning building, for example, it is unlikely that the structure's communication systems will remain working, and first responders won't have a map or plan for the building. Singh's group has developed technology whereby a firefighter or a robot can scatter smart radio beacons inside the building. Some beacons are stationary and some are attached to a human or robot. These nodes begin talking to each other and autonomously organize themselves into an ad hoc sensor network. The radios measure distances to each other and, using algorithms developed by Singh's group, construct a map of their physical layout—or a map of conditions such as temperature—and track the movement of robots or people.

"Imagine there is a commander standing outside the building, and he looks at a screen and he can see where all his people are inside the building," Singh says. "And they can do this without any prior survey of the building, and without any power or prior communications infrastructure inside the building."

Constructing spatial maps from distance-only data has been feasible for some time. It's possible to draw a map showing the positions of cities in the U.S. solely from the intercity mileage table at the back of an atlas, Singh says. But his innovation was the development of algorithms—based on Kalman filtering, Markhov methods, and Monte Carlo localization—that can do the job with a sparsely populated distance table.

Singh has also made progress at the second level of the hierarchy, the one dealing with robot mobility. He has developed a suite of search algorithms for teams of robots to use in spaces humans can't or don't want to go. Some are suited to looking for an immobile person, while others are geared to looking for moving people, such as an intruder. In the latter case, the robots might post "guards" at various locations in a building to spot the intruder's movement.

In addition, Singh's algorithms can be classified as "efficient" (find a target in the lowest expected time); "guaranteed" (clear the environment so capture is assured); or "constrained" (maintain robot positions that ensure network connectivity or line-of-sight communication). "You could combine these algorithms if you have a team of robots or a team of robots and humans," he says. Combining an efficient search with a guaranteed search would tend to minimize search time while still making sure the search ultimately succeeds.

Murphy, who has become a kind of evangelist for the search and rescue robotics community in the U.S., says the technical problems associated with the devices will be solved in due course. But she says strong government funding and support is needed if search and rescue robots are to see widespread use in fewer than 10 years. The standards being developed now at the National Institute of Standards and Technology will also be a big help, Murphy predicts.

Brilliant robotic technology exists, says Murphy, but it needs to be integrated into complete, robust systems, and sensors and other components must be made smaller, stronger, and cheaper. All of this requires corporate effort, she notes. "We are just inches away," Murphy says. "A lot of the software is just waiting for the hardware to catch up." ◼

Further Reading

Murphy, R., Tadokoro, S., Nardi, D., Jacoff, A., Fiorini, P., Choset, H., Erkmen, A.
Search and rescue robotics. *Springer Handbook of Robotics*, Siciliano, B. and Khatib, O. (eds.). Springer Science and Business Media, Secaucus, NJ, 2008.

Kumar, V., Rus, D., Singh, S.
Robot and sensor networks for first responders. *IEEE Pervasive Computing 3*, 4, October-December, 2004.

Shammas, E., Choset, H., Rizzi, A.
Geometric motion planning analysis for two classes of underactuated mechanical systems. *International Journal of Robotics Research 26*, 10, October 2007.

International Rescue System Institute
http://www.rescuesystem.org/IRSweb/en/IRSU.html

Biorobotics Laboratory, Carnegie Mellon University
http://www.cs.cmu.edu/~biorobotics/

Dekker, S.
***The Field Guide to Understanding Human Error.* Ashgate Publishing, Farnham, Surrey, U.K., 2006.**

Gary Anthes is a technology writer and editor based in Arlington, VA.

Technology | DOI:10.1145/1941487.1941494

Gregory Goth

I, Domestic Robot

With recent advances in laser rangefinders, faster algorithms, and open source robotic operating systems, researchers are increasing domestic robots' semantic and situational awareness.

INDUSTRIAL ROBOTS, FIXED-LOCATION and single-function machines, have long been staples of advanced manufacturing settings. Medical robots, which can help surgeons operate with smaller incisions and cause less blood loss than traditional surgical methods, are making fast inroads in metropolitan and suburban hospitals. Rescue robots, included wheeled and snake-like robots, are increasingly common, and were deployed in the search for survivors in the aftermath of the earthquake and tsunami that recently struck Japan. On the other hand, the promise of multipurpose domestic assistance robots, capable of a wide range of tasks, has been a distant goal.

However, recent advances in hardware such as laser rangefinders, open source robotic operating systems, and faster algorithms have emboldened researchers. Robots are now capable of folding laundry, discerning where to place an object on cluttered surfaces, and detecting the presence of people in a typical room setting.

"It's easy for me to be optimistic, but if robots aren't actually being useful and fairly widespread in 10 years, then I will be fairly disappointed," says Charles Kemp, assistant professor of biomedical engineering at Georgia Tech University.

Sensors Enable Awareness

In recent months, numerous research teams have published papers detailing advances in robots' perceptual capabilities. These perceptual advances enable the robots' mechanical components to complete domestic tasks hitherto impossible.

Kemp and his research team have pioneered semantic and situational awareness in robots through several methods, including the creation of radio frequency identification (RFID) semantic tags on common objects such as light switches, and by combining sensor data taken from both two-dimensional camera data and three-dimensional point clouds gathered by laser rangefinders.

Willow Garage's PR2, an open source robotics research and development platform.

University of Bonn researchers Jörg Stückler and Sven Behnke also demonstrated success, using a combination of 2D laser and camera sensors. They programmed a mobile service robot to combine laser rangefinder data that hypothesizes the presence of a person's legs and torso with 2D frontal and profile images of the detected face.

Stückler and Behnke also modeled the semantic probability of detecting a person's presence in different locations of a room—high probability in a chair and low probability on a bookshelf, for instance—and supplied the robot with that knowledge. The prior knowledge of the room semantics and precalculated range of likely valid facial height helps the Bonn researchers discern false positive returns.

Steve Cousins, CEO of Willow Garage, which manufactures the open platform general-purpose PR2 robot, says further advances in perceptual capabilities may be even more likely with the recent debut of sensing technology that enables a computer to analyze an area in three dimensions and then to create what the technology's manufacturer, PrimeSense, calls a synchronized depth image. The technology sells for less than 1/20th of the de facto standard research rangefinder, which costs about $5,000. Both Cousins and Kemp believe the low cost of the PrimeSense sensor (it is a key component of Microsoft's Kinect gaming system) may lead to a surge in situational and semantic robotic research. Kemp says his team recently installed one of the new sensors to its PR2.

In essence, Kemp says its real-time technology greatly simplifies a robot's data-gathering process.

Prior to installing the new sensor, on projects such as the work on making the robot discern clutter, he says "we had to tilt the laser rangefinder up and down, then snap a picture and relate those two things. That's a pretty slow process and really expensive."

A Semantic Database

Kemp says there are two distinct research areas for similar problem sets in domestic robotics: those related to perceptual problem sets, and those related to mechanical awareness. For example, a roving robot meant to help a person with basic housekeeping chores must not only know how to differentiate a refrigerator door handle from a light switch, but it must also be able to calculate which approach its arms must take, and how firmly it must grip the respective levers.

In the experiment using RFID tags, Kemp created a semantic database the robot could refer to after identifying an object. The database contains instructions on how the robot should act upon an object. For example, under "actions," after a robot identifies and contacts a light switch, the commands are "off: push bottom" and "on: push top." Each of these actions is further sub-programmed with a force threshold the robot should not exceed.

Kemp is also investigating another approach to providing robots with such situational awareness that entails equipping human subjects with touch sensors. The sensors are held during the completion of common tasks such as opening refrigerators and cabinet doors in multiple settings. The information on the kinematics and forces of such actions is then entered into a database a service robot can access when it approaches one of these objects en route to performing a task.

"If the robot knows it is a refrigerator, it doesn't have to have worked with that specific refrigerator before," he says. "If the semantic class is 'refrigerator' it can know what to expect and be more intelligent about its manipulation. This can make it more robust and introduces this notion of physically grounded common sense about things like how hard you should pull when opening a door."

Offboard computation akin to the kinematic database is also being done to improve already successful robotic tasks. A team of researchers led by Pieter Abbeel, an assistant professor of computer science at the University of California, Berkeley, programmed a general-purpose Willow Garage PR2 robot to fold towels randomly laid down on a tabletop by using a dense optical flow algorithm and high-resolution stereo perception of the towels' edges and likely corners. Abbeel's experiment yielded a perfect 50-out-of-50-attempt success rate; the robot was able to recalculate failures in the 22 instances that were not initially successful by dropping the towel, regrasping a corner, and carrying on until the task was completed.

Abbeel says his team has been able to greatly reduce the amount of time necessary to fold each towel in subsequent experiments, from 25 minutes to approximately four minutes, by utilizing a new approach: rather than rely heavily upon onboard perceptual data, Abbeel has performed parallel computations on the Amazon cloud on mesh models. Those models, he says, are "triangles essentially put together like people using computer graphics or physics-based simulations. Once you have that mesh model, you can do a simulation of how this article of clothing would behave depending on where you pick it up."

The new approach, he says, relies on observations that the bottommost point of any hanging article is usually a corner. Two consecutive grasps of a towel, he says, will be highly likely to yield two diagonally opposed corners. For t-shirts, he says, likely consecutive grasps will be at the end of two sleeves for a long-sleeved shirt or the end of one sleeve and diagonally across at the hip for a short-sleeved shirt.

"There are a few of these configurations you are very likely to end up in, then all you need to do perception-wise is to differentiate between these very few possibilities," Abbeel says.

ROS is Boss

Another hallmark advance of the domestic robot community is the growth of an open-source ecosystem, built around the BSD-licensed Robot Operating System (ROS), largely maintained by Willow Garage and Stanford University.

"Our goal has basically been to set the foundation for a new industry to start," Cousins says. "We want two people to be able to get together in a garage and get a robotics business off the ground really quickly. If you have to build software as well as hardware from scratch, it's nearly impossible to do that."

Abbeel says the ROS ecosystem may go a long way to taking the robots out of the lab and into real-world locations.

"In order for these robots to make their way into houses and become commercially viable, there will need to be some sort of bootstrapping," Abbeel says. "It will be very important for people to do some applications extremely well, and there has to be more than one. So I hope what may be happening, with robots in different places, is that different schools will develop a true sensibility for the robot, and these things could potentially bootstrap the process and bring the price down. A single app won't be enough."

Cousins says the combination of falling hardware prices for devices such as the PrimeSense sensor, and the blooming ROS ecosystem might be analogous to the personal computer research of the early 1970s, specifically comparing the PR2 to the iconic Xerox Alto desktop computer. List price on the PR2 is $400,000.

"Right now the PR2 is the platform to work on if you want to do mobile manipulation research," Cousins says. "It's a little expensive, but in today's dollars it's about the same as the Alto. It's not going to be the robot you put into your grandmother's home, but the software we develop on the PR2 will likely be a key component of the market. I think ROS is going to be driving those future personal robots." C

Further Reading

Stückler, J. and Behnke, S.
Improving people awareness of service robots by semantic scene knowledge, *Proceedings of RoboCup International Symposium,* Singapore, June 25, 2010.

Maitin-Shepard, J., Cusumano-Towner, M., Lei, J., and Abbeel, P.
Cloth grasp point detection based on multiple-view geometric cues with application to robot towel folding, 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, May 3–8, 2010.

Schuster, M.J., Okerman, J., Nguyen, H., Rehg, J.M., and Kemp, C.C.
Perceiving clutter and surfaces for object placement in indoor environments, 2010 IEEE-RAS International Conference on Humanoid Robots, Nashville, TN, Dec. 6–8, 2010.

Yamazaki, A., Yamazaki, K., Burdelski, M., Kuno, Y., and Fukushima, M.
Coordination of verbal and non-verbal actions in human–robot interaction at museums and exhibitions, *Journal of Pragmatics 42*, 9, Sept. 2010.

Attamimi, M., Mizutani, A., Nakamura, T., Sugiura, K., Nagai, T., Iwahashi, N., Okada, H., and Omori, T.
Learning novel objects using out-of-vocabulary word segmentation and object extraction for home assistant robots, 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, May 3–8, 2010.

Gregory Goth is an Oakville, CT-based writer who specializes in science and technology.

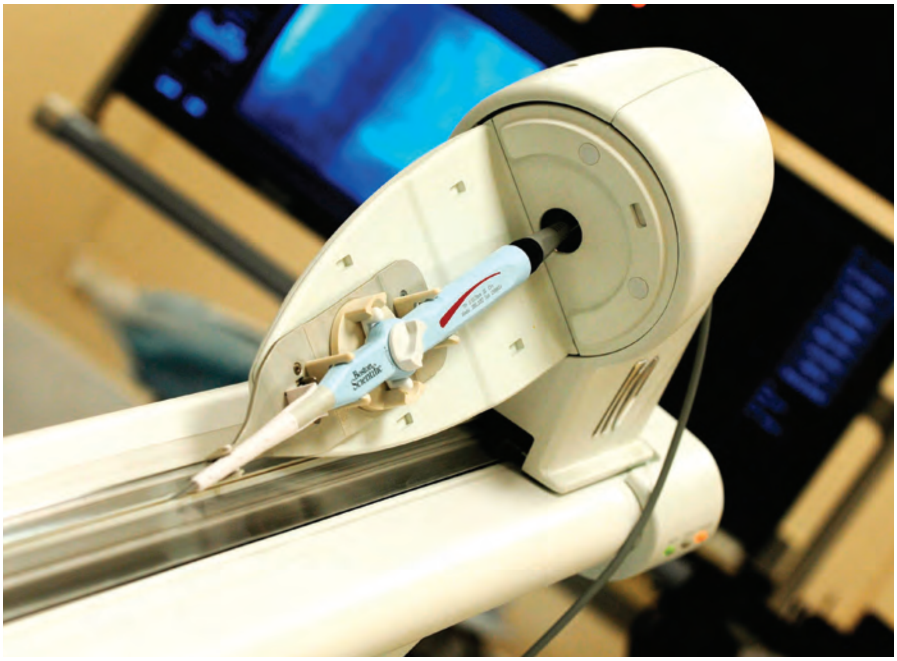
DOI:10.1145/1785414.1785428

Jason Borenstein

Computing Ethics
Work Life in the Robotic Age

Technological change results in changes in expectations, in this case affecting the workplace.

Robots are being designed to perform a broader array of work-related tasks. Global economic hardships may be (temporarily) causing the demand for industrial robots to decline,[4] but improvements in artificial intelligence and the drive for efficiency will likely encourage companies to develop and use increasing amounts of robotic workers. Though the justification for automation is often couched in the language of liberation, this oversimplifies the complexities associated with technological change. Merely because technology is well designed from an engineering perspective, it does not follow that society's problems are solved. This is not to say that efforts to create robotic workers must stop, but the robotics community must be diligent in dealing with emerging ethical issues. Design pathways must be selected that either mitigate or prevent the negative consequences of using robots in the workplace. Otherwise, troubling historical occurrences, such as the decimation of certain segments of the work force, might be repeated.



A view of a robot arm used in world's first remote heart operation performed at Glenfield Hospital in Leichester, U.K., on April 28, 2010.

PHOTOGRAPH BY RUI VIEIRA/AP PHOTO

With each significant technological change, visions of how improved and efficient our lives will become are typically offered. To some degree, the promise that we will be "liberated" from performing repetitive and mundane tasks has held true. Most of us do not mourn the passing of having to wash clothes or dishes by hand. Yet expectations in both our personal and professional lives tend to shift correspondingly, which in many ways counterbalances the "liberating" features that technology offers. Ruth Schwartz Cowan recognized years ago that the introduction of electronic devices into the home did not free women from the burden of doing household chores. As Cowan states, "What a *strange paradox* that in the face of so many labor-saving devices, little labor appears to have been saved!"[1] In short, increasing expectations absorbed all of the extra time that was supposed to be freed up.

Similarly, we need to seriously consider how the increased use of robots will alter workplace expectations. For instance, if robots can help surgical procedures to be completed more rapidly, will demands on surgeons increase so they will have to perform more procedures per day? Expectations in terms of what it means to be a "good" professional are also likely to change, especially if a robot's error rate is lower than a human's. Briefly put, we should be wary of predictions that robots will be our liberators considering how the typical workweek does not seem to be getting shorter or less demanding in the digital age.

The U.S. military is enjoying the benefits of robots since they can complete "dull, dirty, or dangerous" tasks, and their labor is very useful in the civilian realm as well. Yet automation can eliminate job opportunities and usually causes the demographics of the work force to be significantly altered in a relatively short amount of time. Employers find robots to be rather enticing since they do not receive benefits or request vacation time. Through the design choices they make, scientists and engineers play a key role in determining the kinds of employment practices that can and will transpire.

Employment Impacts and Implications

If categories of jobs do indeed vanish as a result of robots, will the relevant skills of displaced workers be transferred to another application or will those skills be rendered obsolete? This concern is not unique to robots. But what may be a new variation now is that the jobs available to humans may be drastically reduced as computers, the Internet, and robots replace humans in employment sectors that used to be thought of as immune to automation. At present, it is fairly difficult for people to find work that is not connected in some way to these technologies. This development might not be conducive to the flourishing of each person's respective talents, and robots are likely to exacerbate this situation. Also, the type of skills that will be in demand if and when the robotic age takes hold might be obvious in some ways but not so apparent in others.[a]

The impact of robotic workers can and will extend beyond the elimination of labor-intensive jobs, which captures a key reason why the ethical dimensions of robots seem to be drawing increased attention. It is not only possible to eliminate "dangerous" and "boring" work but at least some jobs requiring specialized expertise, such as being a surgeon, may start to disappear. A decade ago, Bill Joy, the cofounder of Sun Microsystems, famously warned against this.[2] Even if we don't share Joy's apprehension about the future of robotics, we can still appreciate the perils of trying to replace "uniquely human" abilities such as critical thinking and intuition.

Scientists and engineers should reflect on their ethical responsibilities to communicate with the public about a robot's capabilities and limitations.

To illustrate this point, we can look at the robots being created to assist with the health care needs of elderly populations. An outgrowth of this effort is that it could subtly or perhaps dramatically change how nursing homes function. In principle, robots could free up the time of nursing home staff; for example, a robotic assistant can provide medication reminders or warnings if a resident is in danger. Such a robotic counterpart might enable human workers to be more caring and productive. However, nursing homes and other care facilities will be tempted to downsize their human staff when a robot is "hired" instead of freeing up human staff to give more time to residents.[3] Since many nursing home residents in the U.S. and elsewhere already do not get enough care and individualized attention, this is a very troubling possibility. Theoretically, an increased emphasis on in-home care could for example lead to the creation of other types of jobs but we should be skeptical about this. Financial considerations, the drive for efficiency, and overconfidence in technology are strong driving forces that can push humans "out of the loop."

On a related note, reliance on automation may exacerbate a common human tendency to shift our attention to a different task when we believe (perhaps falsely) that we can trust someone or something else to deal with the task at hand.[b] Returning to the issue of health care, will nursing home staff be less attentive if a robotic assistant is placed in a resident's room? The more reliable we think automated systems are, the more likely it is our attention will stray. What complicates matters is that this type of behavioral shift might not be consciously detected. Hence, it would be wise to temper the confidence that users place in robots and other automated systems, especially when people could be significantly harmed. This could be accomplished in part by ensuring that risks are transparently presented to users. To that end, scientists and engineers should reflect on their ethical responsibilities to communicate with the public about a robot's capabilities and limitations, and not merely leave it to marketers, sales departments, and others to fill this role.

Conclusion

Ethical concerns about integrating robots into the workplace are becoming increasingly pronounced. Again, the intention here is not to stop innovation. Rather, the hope is to inform the design process. Ideally, the robotics community will select design pathways that mitigate the associated concerns and thereby enhance the public's lives.

a For example, in *Wired for War*, P.W. Singer discusses how cooks might have more job security than military pilots because they can prepare food in creative ways. In the civilian realm, he reassures hairstylists by suggesting their specific abilities may keep them employed; Penguin Press, NY, 2009, 130–132.

b Placing too much confidence in technology, often at the expense of other sources of information, seems to be a growing problem with GPS in automobiles; see for example, Is your GPS navigator a friend or foe?" *The Sydney Morning Herald*, (Jan. 12, 2010); http://www.smh.com.au/executive-style/gadgets/is-your-gps-navigator-a-friend-or-foe-20100112-m4ei.html

References

1. Cowan, R.S. *More Work For Mother: The Ironies Of Household Technology From The Open Hearth To The Microwave.* Basic Books, 1983, 44.
2. Joy, B. Why the future doesn't need us. *Wired 8*, 4 (Apr. 2000).
3. Sparrow, R. and Sparrow, L. In the hands of machines? The future of aged care. *Minds and Machines 16*, 2 (May 2006), 141–161.
4. Tabuchi, H. In Japan, machines for work and play are idle. *The New York Times* (July 12, 2009); http://www.nytimes.com/2009/07/13/technology/13robot.html

Jason Borenstein (borenstein@gatech.edu) is the director of Graduate Research Ethics Programs in the School of Public Policy at Georgia Tech in Atlanta, GA.

The author would like to thank Rachelle Hollander, Keith W. Miller, and the anonymous reviewers for their helpful insights and guidance.

DOI:10.1145/1787234.1787248

Samir Chopra

Viewpoint
Rights for Autonomous Artificial Agents?

The growing role of artificial agents necessitates modifying legal frameworks to better address human interests.

It is a commonplace occurrence today that computer programs, which arise from the area of research in artificial intelligence known as intelligent agents, function autonomously and competently;[1] they work without human supervision, learn, and, while remaining 'just programmed entities', are capable of doing things that might not be anticipated by their creators or users.

In short, leaving philosophical debates about the true meaning of 'autonomy' aside, they are worthy of being termed 'autonomous artificial agents'.[a] And on present trends, we, along with our current social and economic institutions, will increasingly interact with them. They will buy goods for us, possibly after carrying out negotiations with other artificial agents, process our applications for credit cards or visas, and even make decisions on our behalf (in smarter versions of governmental systems such as TIERS[2] and in the ever-increasing array of systems supporting legal decision-making[3]). As we interact with these artificial agents in unsupervised settings with no human mediators, their increasingly sophisticated functionality and behavior create awkward questions. If it is a reasonable assumption that the degree of their autonomy will increase, how should we come to treat these entities?

The artificial agent is better understood as the means by which the contract offer is constituted.

Societal norms and the legal system constrain our interactions with other human beings (our fellow citizens or people of other nations), other legal persons (corporations and public bodies), or animal entities. There are, in parallel, rich philosophical discussions of the normative aspects of these interactions in social, political, and moral philosophy, and in epistemology and metaphysics. The law, taking its cues from these traditions, strives to provide structure to these interactions. It answers questions such as: What rights do our fellow citizens have? How do we judge them liable for their actions? When do we attribute knowledge to them? What sorts of responsibilities can (or should) be assigned to them? It is becoming increasingly clear these questions must be addressed with respect to artificial agents.[4] So, what place within our legal system should these entities occupy so that we may do justice to the present system of socio-economic-legal arrangements, while continuing to safeguard our interests?

The Contracting Problem

Discussing rights and responsibilities for programs tends to trigger thoughts of civil rights for robots, or taking them to trial for having committed a crime or something else similarly fanciful. This is the stuff of good, bad, and simplistic science fiction. But the legal problems created by the increasing use of artificial agents *today* are many and varied. Consider one problem, present in e-commerce: If two programs negotiate a deal (that is, my shopping bot makes a purchase for me at a Web site), does that mean a legally binding contract is formed between their legal principals (the company and me)?

A traditional statement of the requirements of a legally valid contract is that "there must be two or more separate and definite parties to the contract; those parties must be in agreement i.e., there must be a consensus ad idem; those parties must intend to create legal relations in the sense the promises of each side are to be enforceable simply because they are contractual promises; the promises of each party must be supported by consideration i.e., something valuable given in return for the promise."[5]

a Jim Cunningham has pointed out that a certain degree of autonomy is present in all programs; consider Web servers or email daemons for instance. One might think of intelligent agents as a move toward one end of the spectrum of autonomy.

These requirements give rise to difficulties in accounting for contracts reached through artificial agents and have sparked a lively debate as to how the law should account for contracts that are concluded in this way. Most fundamentally, doctrinal difficulties stem from the requirement there be two parties involved in contracting: as artificial agents are not considered legal persons, they are not parties to the contract. Therefore, in a sale brought about by means of an artificial agent, only the buyer and seller can be the relevant parties to the contract. This entails difficulties in satisfying the requirement the two parties should be in agreement, since in many cases one party will be unaware of the terms of the particular contract entered into by its artificial agent. Furthermore, in relation to the requirement there should be an intention to form legal relations between the parties, if the agent's principal is not aware of the particular contract being concluded, how can the required intention be attributed?

Legal scholarship has suggested a variety of solutions,[6] ranging from the idea programs should be treated as "mere tools" of their principals to those suggesting programs be granted full legal personhood in order to grant legal efficacy to the deals entered into by them. Some of the suggested solutions struggle to solve this problem when: protocols between buyers and sellers (and their agents) are not specified in advance; the terms of use governing individual transactions are not specified; the terms of a contract are not finalized via human review; or when agents capable of determining the terms of contracts are employed. In these settings, agents might arrive at negotiated or reasoned decisions, which their principals might not have agreed to had they been given the opportunity to review the decision. Given this fact the agent cannot just be understood as a 'mere tool' or 'means of communication' of the principal; rather, the artificial agent is better understood as the means by which the contract offer is constituted.[b]

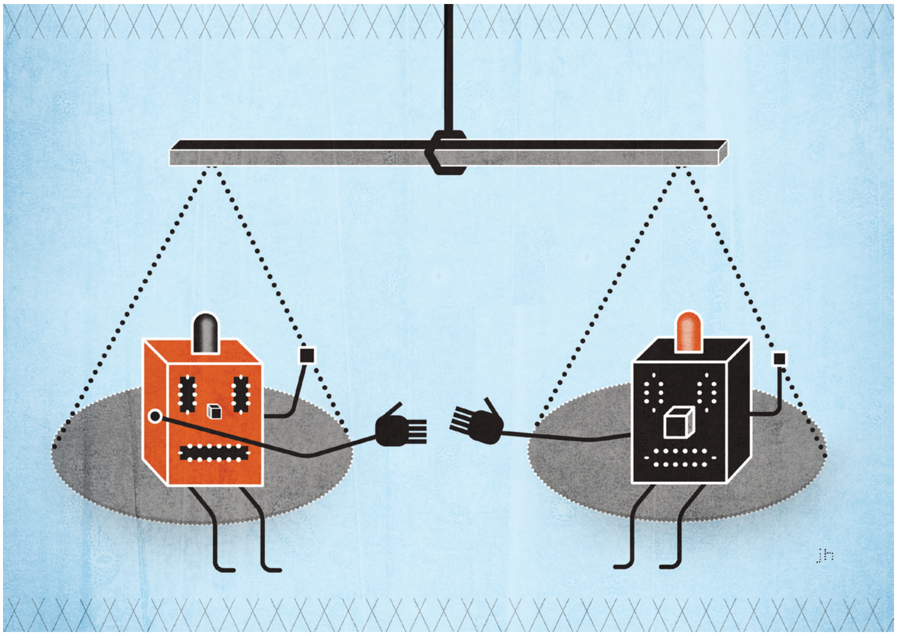
b This discussion is considerably oversimplified but I hope the outlines of the legal problem are clear.

ILLUSTRATION BY JOHN HERSEY

Artificial Agents as Legal Agents

One possible solution, which would require us to grant some legal standing to the programs themselves,[7] would be to treat programs as *legal agents* of their principals, empowered by law to engage in all those transactions covered by the scope of their authority. We would understand the program as having the authority to enter into contracts with customers, much as human agents do for a corporate principal. Some of its actions will be attributed to its corporate principal (for instance, the contracts it enters into), while those outside the scope of its authority will not. The 'knowledge' it acquires during transactions, such as customer information, can be attributed to the corporate principal, in the way that knowledge of human agents is. Lastly, the established theory of liability for principal-agent relationships can be applied to this situation. The details of this solution aside, the most important aspect here is that, unlike a car, a program is neither a thing nor a tool; rather, it is an entity with legal standing in our system.

In granting the status of a legal agent to a computer program, we are not so much granting rights to programs as protecting those that employ and interact with them. Understanding appropriately sophisticated programs as legal agents of their principals could be a crucial step to regulating their presence in our lives. It will enable us to draw upon a vast body of well-developed law that deals with the agent-principal relationship, and in a way that safeguards the rights of the principal user and all concerned third parties. Without this framework, neither third parties nor principals are adequately protected. Instead, we find ourselves in a situation where increasingly sophisticated entities determine the terms of transactions that affect others and place constraints on their actions, though with no well-defined legal standing of their own. Viewing a program as a legal agent of the employer could represent an economically efficient, doctrinally satisfying, and fair resolution that protects our interests, without in any way diminishing our sense of ourselves.



Rights and Legal Personhood for Artificial Agents

There are two ways to understand the granting of rights, such as legal agency, to artificial agents. Rights might be granted to artificial agents as a way of protecting the interests of others; and artificial agents might interact with, and impinge on, social, political, and legal institutions in such a way that the only coherent understanding of their social role emerges by modifying their status in our legal system—perhaps treating them as legal agents of their principals, or perhaps treating them as legal persons like we do corporations or other human beings. And when they enjoy such elevation, they

Artificial agents have a long way to go before we can countenance them as philosophical persons.

must conform to the standards expected of the other entities that enjoy standing in our legal system.

The question of legal personality suggests the candidate entity's presence in our networks of legal and social meanings has attained a level of significance that demands reclassification. An entity is a viable candidate for legal personality in this sense provided it fits within our networks of social, political, and economic relations in such a way that it can coherently be a subject of legal rulings. Thus, the real question is whether the scope and extent of artificial agent interactions have reached such a stage. Answers to this question will reveal what we take to be valuable and useful in *our* future society as well, for we will be engaged in determining what sorts of interactions artificial agents should be engaged in for us to be convinced that the question of legal personality has become a live issue.

While the idea of computer programs being legal persons might sound fanciful, it is worth noting the law has never considered humanity a necessary or sufficient condition for being a person. For example, in 19th century England, women were not full persons; and, in the modern era, the corporation has been granted legal personhood.[c] The decision to grant personhood to corporations is instructive because it shows that granting personhood is a pragmatic decision taken in order to best facilitate human commerce and interests. In so doing, we did not promote or elevate corporations; we attended to the interests of humans.

Artificial agents have a long way to go before we can countenance them as philosophical persons. But their roles in our society might grow to a point where the optimal strategy is to grant them some form of limited legal personhood. Until then, we should acknowledge their growing roles in our lives and make appropriate adjustments to our legal frameworks so that our interests are best addressed. Indeed, this area requires an international legal framework to address the ubiquity of artificial agents on the Internet, and their deployment across national borders.[d] I have merely scratched the surface of a huge, complex, multidisciplinary debate; in the years to come, we can only expect that more complexities and subtleties will arise.

c In his Max Weber Lecture, "Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law," Gunther Teubner notes that animals were often treated as legal actors including being brought to trial.

d The work being done on the "Alfabiite" project (at Imperial College London, NRCCL Oslo, CIRFID Bologna) may be of interest in providing guidance in this regard.

References and Further Reading

1. The fast-growing literature on agent technologies is truly gigantic; for introductions, see M.J. Wooldridge, *Reasoning about Rational Agents*, MIT Press, Cambridge, MA, 2000, and N.R. Jennings and M.J. Wooldridge, Eds., *Agent Technology: Foundations, Applications and Markets*, Springer Verlag, 1998.
2. See http://www.hhs.state.tx.us/consolidation/IE/TIERS.shtml
3. The *Proceedings of the International Conferences on Artificial Intelligence and Law* (http://www.iaail.org/past-icail-conferences/index.html), and the journal *Artificial Intelligence and Law* are rich sources of information on these systems.
4. A very good source of material on the legal issues generated by the increasing use of artificial agents may be found at the Law and Electronic Agents Workshops site: http://www.lea-online.net/pubs
5. *Halsbury's Laws of England* (4th edition) Vol. 9 paragraph 203; c.f. Restatement (Second) of Contracts, § 3
6. There is a large amount of literature in this area; some very good treatments of the contracting problem may be found in: T. Allan and R. Widdison, "Can computers make contracts?", *Harvard Journal of Law and Technology 9* (1996), 25–52,; A. Bellia Jr., "Contracting with electronic agents," Emory Law Journal 50, 4 (2001), 1063; I. Kerr, "Ensuring the success of contract formation in agent-mediated electronic commerce," Electronic Commerce Research 1, (2001), 183–202; E. Weitzenboeck, "Electronic agents and the formation of contracts", *International Journal of Law and Information Technology 9*, 3 (2001), 204–234. Various international trade agreements such as those formulated by the UNCITRAL or national legislations such as the UCITA have not as yet resulted in clarity in these areas.
7. S. Chopra and L. White, "Artificial agents—Personhood in law and philosophy," in *Proceedings of the European Conference on Artificial Intelligence*, 2004 and S. Chopra and L. White, *A Legal Theory for Autonomous Artificial Agents*, University of Michigan Press, to be published.

Samir Chopra (schopra@sci.brooklyn.cuny.edu) is an associate professor in the Department of Philosophy at Brooklyn College of the City University of New York.

Future Tense, one of the revolving features on this page, presents stories and essays from the intersection of computational science and technological speculation, their boundaries limited only by our ability to imagine what will and could be.

DOI:10.1145/1562164.1562192 Jaron Lanier

Future Tense
Confusions of the Hive Mind

Cherish the individual.

BE CAUTIOUS ABOUT the artificial intelligence approach to computer science. It is impossible to differentiate the actual achievement of AI from the degree to which people change when confronted with what is purported to be intelligent technology. We humans are vulnerable to bending over backward, sometimes making ourselves significantly stupider, in order to make an algorithm seem smart. A great many people in the U.S., as well as elsewhere, demonstrated this danger when they interacted foolishly with deeply flawed algorithms related to the credit and mortgage industries.

There is an even greater economic danger ahead as it relates to the idea of AI. If we are gullible enough to expect emergent large-scale intelligence to arise from the vast connections of the worldwide Internet, as has been proposed with increasing frequency in *Communications* and elsewhere, then we risk undermining the value we place on human labor and creativity. We might thus ruin the most successful design yet invented for the purpose of generating and preserving individual human dignity and liberty—capitalism.

Those who believe in the imminent arrival of global AI (possibly emerging from the computing clouds) pretend that all the information we humans upload actually comes from some mysterious supernatural dimension. There's an economic component to the way we lie to ourselves to support this confusion. Millions of us anonymously upload our online offerings—thoughts, pictures, videos, links, votes, and more. Or, if not anonymously, we often express ourselves in such a fragmentary way, as with tweets, that there is no room left for personality. Under these circumstances we accept that we will not be paid for our acts of expression, as if we are engaged in a massive economic ritual to reify the falsehood that a global supernatural brain is speaking, instead of us.

PHOTOGRAPH BY RAMI K

The idea of creativity emerging autonomously from the computing clouds has the potential to ruin what might be called the endgame of basic technological development. Will technology good enough to provide comfort and security usher in a golden age for all? Or will we diverge into two species, one relatively lucky, the other relatively left out, as predicted by H.G. Wells in his novel *The Time Machine* in 1898?

The rarified beneficiaries might turn out to be the owners of the computing clouds, while the rest might be inundated with [CONTINUED ON P. 111]

LAST BYTE

[CONTINUED FROM P. 112] advertising. The bifurcation of humanity could be sustained only so long as those on the receiving end have money to spend. But as more things become free in order to support advertising, fewer of us will be making money. The dénouement would probably be some sort of violent swing toward socialism.

This might sound like an extreme scenario, but consider how much more difficult it is for certain creative people to earn a living today than they did before the public Internet became a global social phenomenon. The most tormented examples are probably recording musicians and investigative journalists.

Alas, it is now common to hear suggestions that people in this predicament should revert to retro (inevitably more physical) strategies of sustenance, like selling branded T-shirts and other merchandise. This is a sad reversal of what had been one of the brightest aspects of technological progress. Prior to the centrality of "open culture" and the rise of online collectivization, technological progress generally supported ever more cerebral, creative, and comfortable means of making a living.

Now extrapolate: How long will it be before cheap fabricating robots are able to download T-shirt designs from the cloud and automatically manufacture customized clothing as easily as one downloads music today? And how long after that will it be before personal robots are able to build copies of the latest medical implant or other gadgets from an online design? The answers are likely to be measured in decades, not centuries. If robotics is eventually good enough to harvest the garbage dumps of the world for materials and transform them into manufactured products, then a plateau will have been reached. At that time, all consumer technology will become media technology. Even those who hoped to make a living from T-shirts will join the investigative journalist and recording musician in poverty.

How far back in history toward the stone age will people have to devolve in order to find a way to make a living when fabricating robots are that good? Will people be forced by the marketplace to work the fields, as academics did under various Maoist-type regimes? Not with good robots around. Surely, robots will eventually also do a better job tending the crops.

How long will it be before cheap fabricating robots are able to download T-shirt designs from the cloud and automatically manufacture customized clothing as easily as one downloads music today?

If you go back to some of the earliest thinking about how information technology might interact with the patterns of human life, you'll find examples of people who thought ahead to this potential dilemma. For instance, Ted Nelson, probably the first person to really think through how something like the Web might be built and how it would influence human society, proposed in the 1960s a design in which each copy of a file existed, from a logical point of view, in only one instance. Any user could make micropayments to gain access. The conflict between file sharing and DRM would be defused because there would be little motivation to make copies. Accessing files would be enticingly cheap, but everyone would make some incremental amount of money from sharing files with everyone else. A new social contract would emerge based on self-interest. This was not just a proposal to extend capitalism, but to broaden its benefits to a greater variety of people, since all would be able to upload interesting bits as needed.

A popular objection when Nelson proposed this design was that few people had anything of interest or value to say, and if they tried to say what they could, no one else would be interested. Fortunately, the rise of social networking has proved these objections unfounded.

I directly experienced a later period, in the 1970s and 1980s, when Nelson was no longer a solitary pioneer. Much of the underlying architecture and ideology that guides the public Internet today appeared in rough cut during those years. The ideas had shifted. Nelson was attacked by the campus left of the time over his willingness to imagine a future in which money continued to be important. Meanwhile, the culture of AI fascinated engineers, drawing their attention away from the problem of how to reward human creativity that had so fascinated Nelson.

We ended up with an Internet and Web that is, for the moment, a sort of cross between mass collective implementation of a Turing Test, through designs like Twitter, and the clumsy fantasy of armchair pseudo-Maoists. I realize these words could strike many as alarmist. If this is the case for you, please look into the history of collectivist design in human affairs. Such designs often appear enlightened at first, with a special way of enchanting idealistic young people. But they have also engendered the worst social disasters of the past century.

That's why I reject the idea that a collective or emergent intelligence is appearing through the computing clouds. We'll never know if it's really there, or if we have collectively become idiots.

Jaron Lanier is a computer scientist interested in interpersonal perception, biomimetic computing, and new displays and sensors. He received a Career Award from the IEEE in 2009 for his lifetime contributions to virtual reality research and is presently working at Microsoft on intriguing unannounced projects.