DOI:10.1145/3034429

Artificial Intelligence

Since its inauguration in 1966, the ACM A.M. Turing Award has recognized major contributions of lasting importance to computing. Through the years, it has become the most prestigious award in computing. To help celebrate 50 years of the ACM Turing Award and the visionaries who have received it, ACM has launched a campaign called "Panels in Print," which takes the form of a collection of responses from Turing laureates, ACM award recipients and other ACM experts on a given topic or trend.

ACM's celebration of 50 years of the ACM Turing Award will culminate with a conference June 23–24, 2017 at the Westin St. Francis in San Francisco to highlight the significant impact of the contributions of ACM Turing laureates on computing and society, to look ahead to the future of technology and innovation, and to help inspire the next generation of computer scientists to invent and dream.

For the first Panel in Print, we invited 1994 ACM Turing laureate **RAJ REDDY**, 2012 ACM Prize in Computing recipient **JEFF DEAN**, 2013 ACM Prize in Computing recipient **DAVID BLEI** and 2013 ACM Grace Murray Hopper recipient **PEDRO FELZENSZWALB** to respond to several questions about Artificial Intelligence.

What have been the biggest breakthroughs in AI in recent years and what impact is it having in the real-world?

RAJ REDDY: Ten years ago, I would have said it wouldn't be possible, in my lifetime, to recognize unrehearsed spontaneous speech from an open population but that's exactly what Siri, Cortana and Alexa do. The same is happening with vision and robotics. We are by no means at the end of the activity in these areas, but we have enough working examples that society can benefit from these breakthroughs.

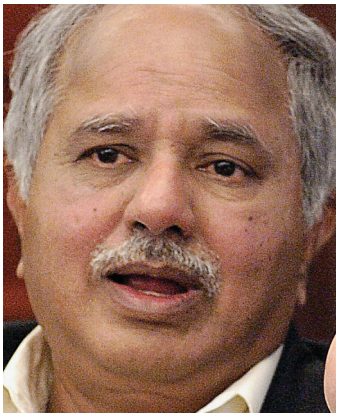
JEFF DEAN: The biggest breakthrough in the last five or so years has been the use of deep learning, a particular kind of machine learning that uses neural networks. Stacking the network into many layers that learn increasingly abstract patterns as you go up the layers seems to be a fundamentally powerful idea, and it's been very successful in a surprisingly wide variety of applications—from speech recognition, to image recognition, to language understanding. What's interesting is we don't seem to be near the limit of what deep learning can do; we'll likely see many more powerful uses of it in the coming years.

PEDRO FELZENSZWALB: Among the biggest technical advances I would include the development of scalable machine learning algorithms and the computational infrastructure to process and interact with huge datasets. The latest example of these advances is deep learning. In computer vision deep learning has led to breakthroughs in object recognition. The accuracy of object recognition in popular benchmarks has increased way beyond what most of us expected to see in the last few years. The impact of this progress still remains to be seen but I expect it will play an important role in building intelligent systems that can interact directly with our physical world.

What specific AI applications will most improve our quality of life in the next five years, 10 years?

JEFF DEAN: Three areas stand out for me: healthcare, self-driving cars, and general-purpose robotics. Machine learning systems will be able to offer suggestions and advice to doctors in ways that are very complementary to the strengths of human medical professionals, resulting in better care for patients, and more efficient healthcare systems. Self-driving cars will be incredibly transformative as well: our urban environments are built around the idea that people own cars and need to park them, etc., and we'll start to see dramatic changes in even things like how cities and neighborhoods are designed as self-driving cars become more widespread. General purpose robots that can operate in messy, uncontrolled environments like households or offices will also start to have a big impact in this time frame.

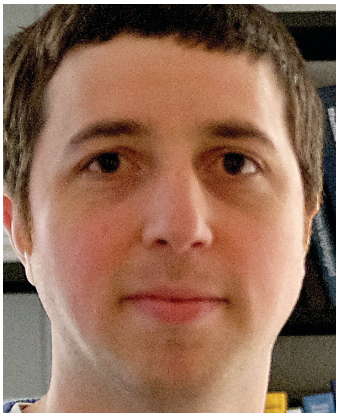
DAVID BLEI: I believe that we are now



1994 ACM Turing laureate Raj Reddy

2012 ACM Prize in Computing recipient Jeff Dean

2013 ACM Prize in Computing recipient David Blei



2013 ACM Grace Murray Hopper recipient Pedro Felzenszwalb

making major progress in two areas that will significantly improve our quality of life. The first is in natural language processing, both in language understanding and language generation. The second is in personalization, in developing software and methods that adapt to user behavior.

These two threads of innovation will result in a more seamless interface between people and AI software, enabling AI to help our lives and society in more ways. For example, we will be able to carry on intelligent and useful conversations with an algorithm, especially around question answering of existing facts. The seamless interface—powered by natural language understanding and personalization—will change how we interact with knowledge bases such as libraries and the internet and thus change how we are able to access, find, and use information.

PEDRO FELZENSZWALB: I believe medicine and public health are areas where the potential for AI is very big and we may see significant impact in the next 10 years. Consider the problem of medical diagnosis. Conceptually this is a simple problem, involving figuring out what condition someone has based on their symptoms. But in practice the problem is very hard.

We rely on specialists, as no one doctor can master all the complexities of the human body. An AI doctor will have access to a database with all of our medical knowledge and the necessary computational capabilities to reason about this data. This AI doctor could be much more easily accessible than the best doctors in the world. The bottom line is that medical diagnosing requires doing statistical inference with lots of data, something that computers can probably do better than humans.

What are some of the major hurdles that AI still needs to overcome in the next 10 years?

DAVID BLEI: Right now, AI is revolutionizing technology through prediction, e.g., "What will I buy next?" or "What face is this in the picture?" I believe that AI will next revolutionize science and scholarship, i.e., how we understand our world through observation. In the context of many fields—astronomy, genetics, sociology, history, and many others—AI can help us analyze massive collections of data to form an understanding of what happened and how things work.

"In my opinion, we are still quite far from realizing the potential of AI. One meta-hurdle is to define what we mean by intelligence."

But there is a significant hurdle to this vision. Finding causal connections, e.g., for science and history, is a deep statistical problem. We must develop the field of causal inference in the context of modern AI to realize its potential in this way.

I will add that using AI to find causal connections will also have an impact technologically. Problems around medical personalization—such as how will a particular patient respond to a medicine—might seem "predictive" at first, but are ultimately causal questions. Indeed, using AI for causal inference will only bolster our predictive capabilities.

PEDRO FELZENSZWALB: In my opinion, we are still quite far from realizing the potential of AI. One meta-hurdle is to define what we mean by intelligence. In the history of AI we have had some specific goals, such as building a computer that can play chess as well as any human, or getting a computer to recognize objects in pictures. However, the AI community has often looked down upon practical solutions to such problems, citing among other things that large engineering efforts and special purpose solutions have little resemblance to intelligence and will not generalize to other problems. It appears that as soon as we figure out how to solve a classical problem in AI we no longer consider the problem to be part of AI. Perhaps the solution simply demystifies the problem too much. It is not clear if we will ever attribute intelligence to a system that we fully understand.

Much has been made of the potential for AI in pop culture. What are some of the biggest myths you've seen? Can you think of examples where science fiction is getting close to reality?

JEFF DEAN: Probably the biggest myth is that AI is one singular thing that you can just "flip on" like a switch, and suddenly you've got human-style intelligence. In fact, AI is a huge field involving many techniques, only very loosely inspired by human intelligence. The good news is these techniques are already quite practical for some kinds of real-world applications today—this is why you can talk to Google on your phone, and it understands what you mean and can give you good answers. It's not magic, but it already works well enough that it's really impressive compared to what we could do just a few years ago.

RAJ REDDY: The best example is Ray Kurzweil and Vernor Vinge's description of the singularity which I believe will happen. Where we disagree is on "when" it will happen. I think it won't happen for at least another 100 years, if not longer.

Two of my favorite examples of science fiction in the movies are *Minority Report* and *Her*, not because they are completely realistic, but because they provide a plausible scenario of things that could happen. In my Turing talk, I speak about teleportation, time travel, and immortality, but then I go on to redefine what I mean by those terms. For example, if we can observe things happening in 3D Virtual Reality without physically being there, that, in my mind, is teleportation, but of course that's not the same definition you get from things like *Star Trek*. The same thing happens in mathematics. If mathematicians don't like a particular outcome, they will define a new complex number world where such facts tend to be true. The issue is, if you don't like the world that you are in, then make a world where what you are imagining is true. There are lots of possibilities, some are reasonable and others may not be, but that depends on the date and time when you ask the question. C

DOI:10.1145/3051455

Cybersecurity

Since its inauguration in 1966, the ACM A.M. Turing Award has recognized major contributions of lasting importance to computing. Through the years, it has become the most prestigious award in computing. To help celebrate 50 years of the ACM Turing Award and the visionaries who have received it, ACM has launched a campaign called "Panels in Print," a collection of responses from Turing laureates, ACM award recipients and other ACM experts on a given topic or trend.

ACM's celebration of 50 years of the ACM Turing Award will culminate with a conference June 23–24, 2017, at the Westin St. Francis in San Francisco. This unique event will highlight the significant impact of ACM Turing laureates' achievements on computing and society, to look ahead to the future of technology and innovation, and to help inspire the next generation of computer scientists to invent and dream.

For our second Panel in Print, we invited 2002 ACM Turing laureate **LEN ADLEMAN**, 2014 ACM Prize in Computing recipient **DAN BONEH**, 2015 ACM Grace Murray Hopper Award recipient **BRENT WATERS**, and ACM Fellows **PATRICK MCDANIEL** and **PAUL VAN OORSCHOT** to discuss current issues in cybersecurity.

The cybersecurity discipline has developed rapidly. Do you think we are staying ahead of, or falling behind, the threats?

LEN ADLEMAN: I think that we are behind. Cybersecurity is a cat-and-mouse game. There can never be a final victory. The Internet is developing so quickly, along so many paths, that while we address current problems, we cannot even anticipate those that are emerging.

BRENT WATERS: In the research realm of cryptography, we have made significant leaps in the past 15 years in terms of which new functionalities we can realize. These include solutions to problems such as identity-based encryption, attribute-based encryption, and fully homomorphic encryption, and potentially can realize an exciting primitive called indistinguishability obfuscation.

Where we seem to be facing problems is filling in the gap between sound cryptography and sound deployments. These deployments can fail for any number of reasons from bad software implementation, to poor design of new cryptography, to use of legacy cryptographic protocols.

What do you see as the top cybersecurity threats in 2017 and why?

DAN BONEH: Social engineering attacks remain one of the top cybersecurity issues in 2017. Phishing and related attacks are still effective at stealing user credentials. Targeted emails continue to be effective at fooling end users into installing unwanted software such as adware, malware, or ransomware. These are common occurrences, and are often the easiest way to gain a foothold on a targeted system. Two-factor authentication and application whitelisting can make the attacker's job harder. They are tools to be used as part of a broad defense strategy.

PATRICK MCDANIEL: There's been an interesting transition of threats and attacks over the last 10 years, and what we're seeing more frequently is professional attacks that more effectively monetize the vulnerabilities in computers. In particular we have seen the rise in things like ransomware, which has become a very serious problem for businesses, government agencies, and organizations that don't have full-time professional cybersecurity staff.

Just looking at what is happened over the last six months in the U.S., it's clear that misinformation has become a major weapon in the cybercriminal's arsenal. I think we will see even more attacks where misinformation is used to try and shape public policy, sway public opinion or even to alter people's behaviors. Obviously the use of misinformation is nothing new—we've seen it before with stock market manipulation, etc.—but I think we're going to see much newer and inventive uses of misinformation as a means of cyberattack.

PAUL VAN OORSCHOT: It's a long-standing problem: software vulnerabilities allowing compromise of user devices and remote control of these compromised machines. This is easily addressed in theory, but harder to fix in the real world. Problems stem from long-ago and deeply entrenched architectural choices in operating systems; use of software applications which favor rich functionality over conservative


Brent Waters


Dan Boneh


Len Aldeman


Patrick McDaniel


Paul Van Oorschot

design; and the freedom allowing users to download and install software with a single click makes us all vulnerable to being socially engineered to install software without any sound evidence or knowledge of what the software will do. It's not likely that we will be able to retrain 10 million software developers, nor to motivate them to voluntarily spend extra time on security when the economic benefit of doing so falls to others. That's an economics of information security challenge—as is that fact that historically, almost all software is sold on an as-is basis without liability for consequences.

Why are cyberthreats/attacks becoming more sophisticated with each passing year?

BRENT WATERS: One can attribute this to two basic reasons. First, technology in general becomes better and more sophisticated over time. One would expect the sophistication of cyberattacks to also flow in that same direction. Another important factor is that with more and more data stored on computing devices, the value in launching attacks increases. For example, there have been multiple attacks that exposed private communications and photos of celebrities. Ten years ago, without smartphones, these photos either wouldn't be taken or wouldn't be accessible. As another example, in this election cycle we have seen that attacks (for example, the DNC emails) have the potential to shake up organizations and possibly shift outcomes in elections. This type of power will not only interest the usual attackers, but will also attract extremely well-funded state sponsored adversaries.

With the public more concerned about cybersecurity than ever before, what should be the top cybersecurity priorities for the new U.S. administration?

DAN BONEH: The highest priority for the new U.S. administration is to shore up the cyber defenses of government systems. Events like the 2015 attack on the office of personnel management that exposed the personnel records of over 21 million people, or the compromise of the IRS systems that may have exposed personal data from over 700,000 taxpayers, should not happen again.

PATRICK MCDANIEL: Our federal IT systems, as we have learned repeatedly, are very much antiquated, due to things like underfunding. But if our society is to become more secure, then we need to focus on updating and fixing those federal systems. One way in which we could do this would be for the current administration to immediately prioritize creating a national two-factor authentication system, either for federal employees or even more broadly. Although that sounds somewhat boring, I think that is the single simplest thing we can do to reduce the threats against the information systems we have in this country. A good friend and colleague of mine, Farnam Jahanian, the Provost of Carnegie Mellon University, has said, "We are not good at doing the easy things, and we need to get better at them."

"The Internet is developing so quickly, along so many paths, that while we address current problems, we cannot even anticipate those that are emerging."

What are the biggest challenges faced by industry in defending against cyberattacks, and what technologies/approaches can help them overcome these challenges?

LEN ADLEMAN: I think the issues raised in the question transcend industry. From my forthcoming book, *Memes: How Genes, Brenes and Cenes Shape Your Life and Will Shape the Future of Humanity*:

We will soon see religions, nations, and economies rise and fall in cyberspace. These entities will be no less powerful and have no less impact on our lives than their current "brick and mortar" counterparts. Political, economic, and even military power will be diffuse; the physical locations of like-minded people will be less important than their numbers and connectivity.

If the U.S., Russian, and Chinese governments are not working on black hat programs that, in the event of war, will knock out the computational infrastructure of the other two, they aren't doing their jobs. Such programs are weapons of mass destruction, and, if used, the death toll could be colossal. A first world country with no computational infrastructure is a country with no economy, no food, no power and ultimately not a country at all.

What are your biggest security concerns as they relate to the influx of connected devices in the Internet of Things (IoT)?

PATRICK MCDANIEL: When it comes to IoT and the future of security, I have a vision of two possible futures. The first scenario comes at a significant cost. But I believe this to be the more optimistic future, because we will understand the trade-off between cost and value, and we're going to pay for it so we can live in a world in which we have much better security than we do today.

The second, and in my view, the more pessimistic scenario is a world in which we have just become used to insecurity. There is a kind of really toxic resignation among some members of the cybersecurity research community as well as industry and government, that today's systems are unfixable and that we don't have the technology, time or resources to make ourselves more secure. I think this is a particularly dim and uncomfortable scenario, not only because the kinds of benefits we see from technology would be greatly diminished, but our potential for changing life on this planet—from healthcare, to society, to communications, to quality of life to energy efficiency, to protecting the environment—will be vastly diminished, if we just accept insecurity. C

DOI:10.1145/3061359

The Internet of Things

Since its inauguration in 1966, the ACM A.M. Turing Award has recognized major contributions of lasting importance to computing. Through the years, it has become the most prestigious award in computing. To help celebrate 50 years of the ACM Turing Award and the visionaries who have received it, ACM has launched a campaign called "Panels in Print," which takes the form of a collection of responses from Turing laureates, ACM award recipients and other ACM experts on a given topic or trend.

For our third Panel in Print, we invited 2009 ACM Prize recipient **ERIC BREWER**, 2004 ACM A.M. Turing Award co-recipient **VINT CERF**, 2016–2017 Athena Lecturer **JENNIFER REXFORD**, ACM Grace Murray Hopper Award recipient **MARTIN CASADO**, ACM Fellows **NICK FEAMSTER** and **JIM KUROSE**, and ACM member **GEORGE ROUSSOS** to discuss the Internet of Things (IoT).

What do you see as some of the biggest transformations that have been brought through the Internet and where do we go next?

ERIC BREWER: The most important transformation brought about by the Internet is the kind of self-empowerment it has caused. If you don't know something, you can find it out. If you want to educate yourself on something, you can learn it fairly directly. People feel like they can learn anything, in any country now.

NICK FEAMSTER: The early Internet was a network of trusted research universities with very few stakeholders. There was no business aspect to it, there were no profits to be taken, and there was little to no concern over security. The chief goal was connectivity, and the primary challenges were technical in nature.

Today, the situation is much different, with each of the previous points having been turned completely on their head. We see increasing tensions between stakeholders, especially between Internet service providers and content providers on to issues like pricing of Internet access, network neutrality, performance guarantees and quality of experience. We also see tremendous tension in cybersecurity between attackers, businesses and end users.

JENNIFER REXFORD: Recently, the Internet has become an amazing way to collect and analyze data about people and their behavior and the kinds of things they do online. This, in turn, has allowed the information we see on the Internet to be much more customized, like Google search and so on. Which brings us to the current evolution, the connecting of the Internet to the physical world, or Internet of Things. This is where we are actually effecting change in the physical world based on the information that gets collected over networks. .

VINT CERF: Projecting into the future, we can see much higher-speed access to the Net, more wireless access and increasing amounts of artificial intelligence and machine learning adding to our ability to accomplish our objectives. It's a rich environment we are heading into.

There are reasons to be concerned—for example, about safety, security, privacy, resilience, and robustness. I am particularly concerned about what I'll call "autonomy," which stems from my concern that you don't want to have a highly automated house that doesn't work when it's not connected to the Internet. So, you need to have local capability independent of or in addition to interactions through the public Internet.

> "The most important transformation brought about by the Internet is the kind of self-empowerment it has caused."

There are still more people in the world offline than on. How will connecting these individuals help neglected and underserved communities around the world?

MARTIN CASADO: I agree with the United Nations in the view that connectivity to the Internet is a basic human right. Beyond the intrinsic benefits of better communication within the community, it provides access to the grand marketplace that's erupted within the Internet. In many ways, that can become a great equalizer. If it costs me less to produce a good or a service, and the distribution cost (in this case the Internet) is the same, then I have an advantage in an open market. Of course, it isn't as simple as that, but it certainly does inject underserved communities directly into the economic nervous system in which they can participate.

GEORGE ROUSSOS: The two main factors limiting the ability of people to access the Internet are affordability and lack of literacy and language skills. While getting online can provide benefits, connectivity is not a panacea for all ills. Lifting these communities out of poverty and getting the basics right such as access to clean water, vaccinations, or in some cases a less corrupt government, would be a priority. Moreover, joining the connected world as a latecomer involves significant hazards as well as opportunities, so developing the appropriate skills and safeguards is a precondition.

There are already interesting cases highlighting how innovations can be created from the bottom up: for example, through microlending and using the mobile Internet to broaden access to financial services.

JENNIFER REXFORD: I think there is a lot of opportunity to collect data that can help people make better deci-

sions. For example, farmers could determine the going rate for their crops, rather than relying on a third-party intermediary to determine prices. Knowing what the weather is going to be like in a few days to make decisions about farming practices, and so on. That being said, having access to information for education and training and awareness doesn't replace having access to clean water and very basic needs.

One problem in a lot of the developing world is that much of the Internet traffic is routed back through more developed areas; traffic in South America being routed through Miami, or traffic in Africa going through Amsterdam or London, etc. So there is a missed opportunity to host local content locally. For example, if you're in Kenya, a local Kenyan website will be hosted outside of Kenya, making it very expensive and slow to get information. What we are starting to see more are efforts to have Internet exchange points in the region so that the multiple network providers within Africa and within South America can directly connect with one another and provide a stable platform for hosting of local content.

For organizations and individuals to be confident when conducting transactions and exchanging information, the Internet has to be secure. How does the IoT impact the security of the Internet?

JIM KUROSE: With an ever-increasing array of devices being connected to the Internet (between 26 billion and 50 billion devices in manufacturing, business, and home applications by 2020, by some predictions), the question of resilience—knowing that a device will continue to perform its tasks safely and securely in the presence of unintended as well as malicious faults—is increasingly important.

VINT CERF: There are technologies that allow people to protect themselves better. Two-factor authentications are a good example of that—the best practice of which is to encrypt everything from the laptop or mobile all the way to the server on the net. All of these are practices we adopt at Google.

NICK FEAMTSER: There are a couple of reasons why IoT raises the stakes as far as the security of the Internet is concerned. An Internet attack may now involve physical inconveniences or threats such as security cameras, door locks, thermostats, etc.

The issue here is that most businesses are fundamentally focused on the market they serve. In other words, a hardware company is just a hardware company, a consumer electronics company is just a consumer electronics company. They are not thinking about the security of the software they put on the devices they sell. So it won't be long until we have an abundance of fundamentally unpatchable, insecure, and difficult if not impossible-to-patch devices affecting nearly every aspect of our daily lives. It's a perfect storm.

ERIC BREWER: Even though "less-connected devices" sounds paradoxical in today's scenario, I believe it's an option. As an example, if a device has to connect through the user's phone or home laptop or computer, maybe that is a bit safer because then, at least, the gateway could be secured. Another option is to stop making these devices so flexible. They are really just doing one kind of reporting, and all the rest of the data is in the cloud. It's more plausible that you could make that secure.

What makes security hard is if you are trying to have a lot of flexibility in the device, or complexity, or if you're trying to change what the device is doing over time, and that's why you're having upgrades. All this makes it much more like a phone and then it really needs to have a more automated form of security patching.

What are the possibilities, and repercussions, of IoT capabilities such as smart cities and connected cars?

MARTIN CASADO: There are obvious answers here around energy efficiency, traffic, safety, etc. But I feel that those are already easy to see from where we are today. So perhaps I will take a bit of a longer view and say that in the limit IoT could very well make the notion of a city anachronistic. Cities are largely products of organic growth and physical constraints; close enough for protection and commerce, and far enough away for privacy and access to resources. However, IoT changes these constraints. Drones can deliver goods without requiring traditional roads or supply routes. Advances in connected and urban farming can allow sustainability just about anywhere. And the Internet provides a social overlay that is independent of geography. We are heading toward a future where cities are more defined by common interests than by geography.

What do you think are some of the potentially most exciting/important applications of IoT beyond the ones already being actively developed?

JIM KUROSE: It's difficult to predict future Internet applications. But I'll make *one* prediction. I believe education and skill acquisition have yet to be truly disrupted by the Internet, or by interactive and/or virtual reality/augmented reality technologies. As a long-time teacher (and learner), I don't think there is anything as good as learning with inspired and engaged teachers and students, using interactive learning and team-based activities in the classroom. But that approach is neither uniformly affordable nor scalable. So I do believe a next generation of interactive software/textbooks/classes is increasingly important to meet the pace and need for training, skills updating, and acquiring new fundamentals.

GEORGE ROUSSOS: One specific way that I hope the IoT can bring about change is by shifting the emphasis away from our current predominantly visual mode of interaction with information, which I consider to be the key ingredient enabling a sedentary and passive contemporary lifestyle. IoT technologies afford interactions engaging the whole body through touch, proprioception, equilibrioception, interoception, and perhaps a few new artificial senses that can hopefully rebalance the focus on the brain as the only locus of intelligence.

In particular, my hope is that the IoT will play a key role toward improving the health and the sustainability of the planet: overconsumption of raw materials, pollution from fossil fuels, and industrialized farming, the destruction of forests and numerous other effects of modernity are setting massive challenges ahead. I believe the IoT has to play a central role in addressing these challenges and ensuring the welfare of future generations. C

CREDIT TK

DOI:10.1145/3079064

Big Data

Since its inauguration in 1966, the ACM A.M. Turing Award has recognized major contributions of lasting importance to computing. Through the years, it has become the most prestigious award in computing. To help celebrate 50 years of the ACM Turing Award and the visionaries who have received it, ACM has launched a campaign called "Panels in Print," which takes the form of a collection of responses from Turing laureates, ACM award recipients and other ACM experts on a given topic or trend.

For our fourth and final Panel in Print, we invited 2014 ACM A.M. Turing Award recipient **MICHAEL STONEBRAKER**, 2013 ACM Prize recipient **DAVID BLEI**, 2007 ACM Prize recipient **DAPHNE KOLLER**, and ACM Fellow **VIPIN KUMAR** to discuss trends in big data.

Gartner estimates that there are currently about 4.9 billion connected devices (cars, homes, appliances, industrial equipment, among others) generating data. This is expected to reach 25 billion by 2020. What do you see as some of the primary challenges and opportunities this wave of data will create?

VIPIN KUMAR: One of the major challenges we are going to see is that the data being gathered from these connected devices and sensors is very different from other datasets that our big data community has had to deal with.

The biggest successes we have seen for big data are in applications such as Internet search, e-commerce, placement of online ads, language translation, image processing, autonomous driving. These successes have been enabled, to a great extent, by the availability of large, relatively structured datasets that can be used to train a broad range of machine learning algorithms. But the data from multitudes of interconnected devices in its raw state, can be highly fragmented, disparate in space and time, and very heterogeneous. Analyzing such data will be a big and new technological challenge for the machine learning and data mining communities.

DAVID BLEI: The key idea here is that just the data from something as simple as Netflix watching habits doesn't provide the recommendation of a new movie; it's that data alongside all the data from everybody else that helps make recommendations.

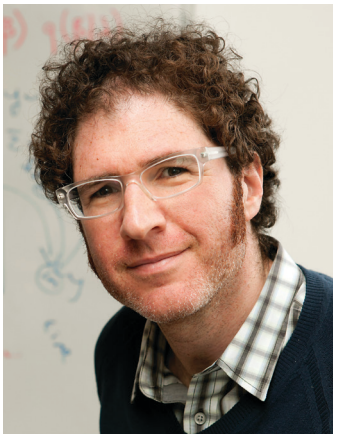
It's an exciting world because we are personalizing our interaction with devices through the aggregate data of everybody using their devices. Of course, this all comes with a challenge around privacy and what we give up when we make our data available or the spectrum of how much we can give up against how much personalization power we get in return.

The other opportunity is in an unprecedented way to learn about the world through these huge collections of many individuals. This is a massive dataset, and patterns of communication, interaction, and movement—including all types of other macro-level descriptions of society and people and the world—are now available to us.

As more data is collected from a growing pool of devices, has the individual lost the right to information privacy?

MICHAEL STONEBRAKER: Imagine this simple example: you show up at your doctor's office and have an x-ray done and you want the doctor to run a query that shows who else has x-rays that look like yours, what was their diagnosis and what was the morbidity of the patients. That requires integrating essentially the country's entire online medical databases and presumably would extend to multiple countries as well. While that is a daunting data integration challenge, because every hospital chain stores its data with different formats, different encodings for common terms, etc., the social value gained from solving it is just huge. But that also creates an incredibly difficult privacy problem, one that is not a technical issue. Because if you're looking for an interesting medical query, you're not looking for common events; you're looking for rare events, and at least to my knowledge, there aren't any technical solutions that will allow access to rare events without indirectly disclosing who the events belong to.

I view the privacy problem to be basically a legal problem. We have to have legal remedies in this area.


David Blei


Daphne Koller


Vipin Kumar


Michael Stonebraker

There are tons of examples of data that can be assembled right now that will compromise privacy. Unfortunately, the social value to compromising privacy is pretty substantial. So, you can argue that technology has rendered privacy a moot question. Or you can argue that preserving privacy is a legislative issue.

As predictive models are increasingly used, how do we avoid biases when interpreting and using data?

DAPHNE KOLLER: Bias will always be a challenge, and there isn't a single, magic solution. The bigger question is: "How do we disentangle correlation from causation?" The gold standard in medicine is that of randomized case control. In the case of Web data, it's called AB testing. Although not perfect, randomized case control, or AB testing, is about as good a tool as we have been able to develop for addressing some of the confounders. Unfortunately, this type of control is not feasible in all cases. Then processes must be carefully scrutinized to check for different confounders and to look for any and all correlations that give rise to the phenomenon being viewed. It's a process that requires a lot of thought and a lot of care and cannot be overstated in its importance.

For example, sometimes there are biases that are reflected in the conclusions that are drawn from the data. In searches on certain sites for example, "Steph" auto-completes to "Stephen" rather than "Stephanie" because Stephen is a more common search term. Some would say this is a gender bias and should be eliminated. As a woman in tech, I can certainly relate to and understand that perspective. Some would also say that the data is what it is, and if Stephen is a more common search term than Stephanie—then do we really want to make the algorithm do something other than what is best for user efficiency? It's a real quandary, and one can make legitimate arguments either way.

MICHAEL STONEBRAKER: The trouble with predictive models is that they are built by humans, and humans by nature are prone to bias. If we look at the most recent presidential election, we see a spectacular failure of existing polling models. Twenty-twenty hindsight shows that nobody thought Trump could actually win, when in reality, it is far more likely the polling models were subtly biased against him.

"The trouble with predictive models is that they are built by humans, and humans are by nature prone to bias."

So, the problem with predictive models is the models themselves. If they include fraud, bias, etc., they can yield very bad answers. One has to take predictive models with a grain of salt. We put way too much faith in predictive modeling.

What role can big data and machine learning play in helping scientists understand data (for example, in the Human Genome project) and bring forth some potential real-world opportunities in health and medicine?

DAPHNE KOLLER: One of the main reasons I came back to the healthcare field is because I think the opportunity here is so tremendous. As costs go down, our ability to sequence new genomes increases dramatically. And it's not just genomes; it's transcriptomes and proteomes and many other data modalities. When we combine that with wearable devices that allow you to see the effect of phenotypes, there is an amazing explosion of data that we could access. One reason this is beneficial is that it will improve our ability to determine the genetic factors that cause certain diseases. Yes, we could do that before, but when faced with tens of millions of variations in the genome and only a couple hundred examples to use, it's really difficult to extract much out of that except the very strongest signals.

Are there potential technological breakthroughs on the horizon that could transform this area again in the near future?

DAVID BLEI: I think we are in the middle of a transformative time for machine learning and statistics, and it's fueled by a few ideas. Reinforcement learning is a big one. This is the idea that we can learn how to act in the face of an uncertain environment with uncertain consequences of our actions; it's fueling a lot of the amazing results that we're seeing in machine learning and AI. Deep learning is another idea—a very flexible class of learners that, when given massive datasets, can identify complex and compositional structure in high-dimensional data. Another idea is 60 years old, but it's optimization: I have some kind of function and I want the maximal value of that function, how do I do that? Well, it's called an optimization procedure. Optimization tells us how to do that very efficiently with massive datasets.

VIPIN KUMAR: New types of sensors and communication technologies can be quite transformational. The kinds of sensors that we see today, we could not even have been imagined just a few decades ago. Mobile health sensors such as Fitbit and Apple Watches that can record our physiological parameters at unprecedented detail have been around only for the past decade or so. New types of sensors based on advances in electronics, nanotechnology, and biomedical sciences are already enabling deployment of small and inexpensive satellites that can monitor the earth and its environment at spatial and temporal resolutions never possible before. Without technologies such as RFID, it would be very hard for someone to imagine that you could walk into a store and purchase something just by looking at it or by being close to it—something that is now possible at Amazon Go, a grocery store in Seattle that has no checkout counter. New sensors based on quantum technology may open up entirely new applications that we are not even considering today.

Final thoughts?

MICHAEL STONEBRAKER: All of the fancy social benefits we expect from big data depends on seamless data integration. Solving the problem of how to improve data integration is going to be key in getting the most benefit from all the data being created. C