AI Theory and Practice: A Discussion on Hard Challenges and Opportunities Ahead

Moderator: Eric Horvitz

Panel Members: Lise Getoor, Carlos Guestrin, James Hendler, Henry Kautz, Joseph Konstan, Devika Subramanian, and Michael Wellman

■ *A special track on directions in artificial intelligence at a Microsoft Research Faculty Summit included a panel discussion on key challenges and opportunities ahead in AI theory and practice. This article captures the conversation among eight leading researchers.*

Eric Horvitz: Welcome to the panel discussion today.[1] We have quite an interesting and esteemed panel of experts — passionate researchers in the field of artificial intelligence, investigating different aspects of AI.

The idea is to ask the panelists to share their thoughts about the key challenges ahead in theory and practice across the broad constellation of research in artificial intelligence, which includes quite a few subdisciplines — and, in fact, if you asked a large number of researchers, "What is the main field of the people on this panel?" they might not even say AI, they might say human-computer interaction (HCI), e-commerce, or they might go down more deeply in the ontology of the topic space. So, we have a variety of people here with different interests and backgrounds that I asked to talk about not just the key challenges ahead but potential opportunities and promising pathways, trajectories to solving those problems, and their predictions about how R&D might proceed in terms of the timing of various kinds of development over time.

I asked the panelists briefly to frame their comments sharing a little bit about fundamental questions, such as, "What is the research goal?" Not everybody stays up late at night hunched over a computer or a simulation or a robotic system, pondering the foundations of intelligence and human-level AI. There's a variety of goals in the field, too.

 ISSN 0738-4602

We have here today Lise Getoor from the University of Maryland; Devika Subramanian, who comes to us from Rice University; we have Carlos Guestrin from Carnegie Mellon University (CMU); James Hendler from Rensselaer Polytechnic Institute (RPI); Mike Wellman at the University of Michigan; Henry Kautz at tjhe University of Rochester; and Joe Konstan, who comes to us from the Midwest, as our Minneapolis person here on the panel.

Joe Konstan: I was actually surprised when you invited me to this panel, because as you were saying, I don't think of myself as an AI person, though I've been to AI conferences and have worked in recommender systems. I think of myself at the core in human-computer interaction. So I went back and started looking at what I knew of artificial intelligence to try to see where the path forward was, and I was inspired by the past. I was inspired by going back to the vision of Turing, of Weizenbaum, of Minsky, and realizing that while people may have gone too far in trying to turn computers into thinking like humans, that actually the Turing test was remarkably inspiring if you treat it as the question of how do we engineer computer systems to interact appropriately with humans. I don't think Turing was exactly right. I don't think you have to fool a human into believing that you're human. But what you have to do is be good enough that you don't distract a human by acting rudely, automatically, inappropriately. I want to run through a handful of examples of why this isn't trivial, and why we're often doing the wrong thing, some of which will take us toward my belief that, yes, we have real potential.

In my own area of recommender systems one of the things that we learned is that it's much easier to optimize the quality of your prediction if you measure that by saying "how well do I predict some data that's sitting off in a database" than it is to come up with recommendations that people actually appreciate.

We've done a bunch of studies that show that people would prefer recommendations that are less good (in terms of accuracy) but more interesting. They would prefer more diversity at the expense of accuracy. They would prefer less obviousness. That's all about human qualities that we can engineer into a system, once we understand them.

We've also learned that some things that should be obvious, such as that explanations are remarkably difficult. If you actually tell somebody how you came up with a recommendation, we found it depresses willingness to believe that the data is any good. And that caused me to reflect back to my AI 1 course where I remember reading about MYCIN and how wonderful this was. And what was MYCIN's great failure? You could argue there were two of them: one, that the researchers didn't anticipate the liability and insurance industry; and the other one, that it was a human interface problem, that people don't necessarily want to go and type a bunch of yes/no questions into a computer to get an answer, even with a rule-based explanation, that if you'd taken that just a step further and solved the human problem, it might have worked. Related to that, I was remembering a bunch of these smart house projects. And I have to admit I hate all of them. I hate smart spaces. I think everyone hates smart spaces.

Here's a simple example of a question that AI needs to answer: If you're about to turn off the lights to save energy because the sensors think that there's nobody there, do you warn people and give them a chance to answer? There's no good answer to this question. I can tell you if that person is in bed asleep, the answer is no, don't wake them up to say, "hey, I'm about to turn off the lights." I can also tell you that if they're in the bathroom very still, the answer is yes, you don't turn the lights off on that person; they're dealing with problems enough on their own. How do you distinguish those two things in a system with anything other than ad hoc rules? How do we learn those behaviors, how do we model enough about humans to say what's respectful?

Why do we love to hate Clippy? Two obvious problems jumped out in the commercial implementation, which were less noticeable in the research world. One of them is understanding history and context: the first time you tell me something, it's new; the fourth time, it's annoying. If your stack is one deep, you never understand history; that's obviously not deep AI.

But the second problem, which I think is bigger, is understanding concepts like subtlety. Clippy will ask questions like "hey, it looks like you're writing a letter, I can help." As true as that may be, putting a little icon in the corner that says, "format as letter" or "letter wizard," lets the person take the initiative. If you look at the work going on in interruption and attention, here [at Microsoft Research], by Brian Bailey at Illinois, and by others in this area, to me it feels like a great AI problem. You have sensor fusion, you have so much different information, all of whose goal is to get computers to be respectful and take their appropriate place. Work in online community and how to get an online community that manages its members feels like a great problem in AI. Work in online health and persuasive computing is studying how to diagnose where someone is in their own mental decision making, in their own behavior change, and adapt the interaction to most effectively help that person get to where they want to go. Those are great problems in AI. So the challenges and opportunities go back to that original vision. The challenge and the opportunity is how do we build computing systems that may have their own goal, but that in their interaction with humans interact in such a way that humans can interact naturally and in a trusted way with those systems.

Henry Kautz: I think many people would agree that probably the greatest crisis facing the world, next to global warming, is coming up with the ability to provide health care to all individuals. What's interesting when you look at why we need health care is that most money is spent on conditions that either could be prevented through education and lifestyle changes or are completely nonpreventable because it's simply a matter of growing old. So the domain that I've been interested in for the last several years is applying AI to create systems that could interact with people to provide levels of caregiving and to help influence people's behaviors in positive manners.

It grows out of work in things like smart homes and things like interruption, to get to the point where you have systems that can, for example, monitor a person's activities of daily living, notice changes in their behavior, and ultimately interact with them to provide help, provide assistance, and so on.

It's a great domain because not only is it socially relevant, but it's a place where you can bring together basically all AI technology, sensing, state estimation, natural language, and so on.

I agree strongly with Joe that it's easy to get a negative reaction to an awful lot of the work in the area because it seems to be insensitive to human factors and human needs.

I always think that's not a reason not to do the work, but that is a reason that when you're doing the work you always have to spend time also talking to end users, doing focus groups with nurses, with families, with caregivers. And when you bring them into the loop, it goes both ways. It both opens the eyes of the public to what could be done with technology, and can also open your eyes as a researcher to what are really the core problems to address.

Michael Wellman: I'm glad that Joe pointed out that accuracy and prediction are not everything, because I was going to choose not to try to predict what the next best opportunity is, in part because I've never been especially accurate about that in the past, and also because I'm not sure that that is really the way that successful long-term research enterprises actually proceed. I think in the AI area opportunities are just so dense, that is, there are so many rich problems such that any really good idea in AI is going to have wide and important benefits, that we don't really have to be extremely tactical about this, except with respect to our own positions and our own opportunities and what really engages our own interests.

In thinking back to how I wound up in the problems I work on, I was originally motivated by dealing with markets, markets as a way to decentralize resource-allocation problems. And for many years if I was on a panel like this, I would proselytize for why our software agents need to be market aware, and that's the most important domain.

What I did not especially anticipate was the electronic commerce explosion from 10 years ago. Thinking back to 1995, if you were doing X research in 1995, you said, hey, I should do X research on the web. And I was doing markets and auctions, so that's how I wound up with it. It wasn't really foresight there.

I think it's interesting that just in this conference we've heard some related talks. I think Eric mentioned a little market-based task allocation for allocating computational resources. We heard a lot of stuff about the computing cloud and how maybe the decentralized resource allocation kinds of issues are going to come up big again. So, maybe that original motivation will come back.

I think I really latched onto the electronic commerce approach, not just because of the value of the opportunity, but because it gave me an excuse to stop trying to argue with people about why markets might be good. The point was that whether you like it or not, markets are out there, and it's a domain that is important to deal with. So, I think you all can pick your problems by again just finding something that engages you, and anything that you latch onto there is going to have wide benefits.

Eric Horvitz: And I could vouch for Mike being deeply focused on markets and the promise of electronic commerce way before there was the upswing of this technology with web applications. In fact, I think you actually visited Microsoft Research in 1995 or so, 1996, whenever it was, and together we were thinking about this world we live in, and we said, can you believe that eBay came to exist? We were just marveling at the prospect of the concept. It seemed so alien at the time that it actually was happening right before our eyes.

Michael Wellman: Well, I remember in 1996 when we found out about eBay, and we saw that they had done — they had claimed that they had done $60,000 worth of total volume, and —

Eric Horvitz: It was pretty impressive then.

Michael Wellman: — and that was very impressive. I said, if we could ever get that with our system, that would be great.

James Hendler: So, I've been looking around the panel. I think age-wise there's a couple of us who are close, but since I started doing AI as a freshman in college, I probably have been doing it longest. (Laughter)

For more years than I'm going to admit, I've been doing more or less the same thing, which at various times in my career has been mainstream

AI, not AI, sort of AI, and now I'm not even sure how to categorize, because some people in AI have heard of the semantic web, and think maybe there's some relation, other people think it's all about AI, which is wrong.

So thinking about what to say today, I thought back a couple of years ago we had a Fellows forum for the 50th anniversary of the Dartmouth conference, which I was not at. We were supposed to write a sort of one-pager on more or less the same topic as this panel, and the idea was not coming to me, future of AI, just couldn't come up with it, and I had a dream literally, a nightmare. It was one of those dreams where I was in front of the room and it was time to give my presentation, and I was fully dressed, but I didn't have my slides, and no one would tell me what my talk was. And I finally found the program, and the title of my talk was, computers play chess, humans play Go. And I woke up realizing that that was the answer that I'd been looking for. At one point many years ago, the reason chess was picked as a really hard problem to challenge computer science and motivated a lot of early AI was not because we wanted a chess player, it was because we picked a really hard thing humans could do and computers couldn't.

Now computers actually play chess better than most people, and some programs even better than the best people, and there are other games and other things, the learning stuff. So, predicting traffic in Seattle, we beat the pants off the average Seattle driver, according to Eric.

But what about all the stuff that humans do better? That used to be what AI was about, looking at that stuff and saying, you know, what is that, what's it all about, how do we do that? You can't solve Go using common things at least for another 40, 50 years if you just believe Moore's Law, and even then it will take till the heat death of the universe to do it computationally. Go has all these things we used to talk about in the planning community, like nonlocal effects and patterns and things like that. You go to a current planning conference, you won't find anybody talking about those things. You go to the learning conference, you hear almost all about mathematical models of learning and data-mining things. You hear almost nothing about how is it that children can differentiate the stories they're reading that are fables and the stories they're reading that are real-life things. After reading the one with the talking crow, very few kids go out and think the crow actually talks, right, and go talk to crows.

So, there are a lot of really hard problems that have to do with what intelligence really is that we have forgotten, that we have stopped looking at because we're looking where we know how to surpass people. What I see as the real challenge is once many years ago the cognitive side of AI and the computational side of AI were in something of a balance, and somehow they've gotten very, very badly out of balance. The people who think about humans, human relationships, trust, respect, reliability, we have computer definitions for all those things that have almost nothing to do with what humans mean by those things, right?

It's time for us to actually go back to the thing we were originally looking at, which is intelligence, to look at the different kinds of intelligence, to look at the different models of intelligence, and start saying, what are things we don't know how to do. And I think that's a real challenge. It's not grand challenge problems, let's make a faster robot, it's let's make a bigger robot; it's let's make something that can attack some of the stuff that we don't right now know how to do.

Carlos Guestrin: The last few years I've been having some AI completeness envy. I've been thinking about what is a bigger interesting AI complete challenge, an AI complete problem that would be interesting to tackle, that does not involve, let's say, a robot that saves the universe. So, it doesn't involve some big hardware, but maybe involves a system that would be accessible for most AI researchers, maybe information that's available on the web as an example.

I would like to have a problem with an aspect of data collection, with an aspect of high-level chaining of information and an aspect of decision making. I would like something that ends up being pretty cool or very cool, and the really cool case will be AI.

So, here is my first proposal. Have you seen FactCheck.org? Pretty cool place. You have statements there that politicians have made, for example, and they try to analyze, collect information, and try to justify whether it's true or not. You may not believe on their analysis, but that's part of the system.

So, you can imagine an automated fact-checking system where you provide some fact that you're trying to figure out, but it's not just what's the capital of Finland, it's something that requires you to chain multiple bits of information. It includes a user interaction part where you can learn about what the user trusts or does not trust about this information. I think if we had that system, it would be extremely useful for all of us.

This is one example I think where it would be cool to discuss some high-level goals.

Devika Subramanian: Well, I say don't knock these robots that will save the universe. I don't know how many of you got the mail from Life Foundation asking you to become a board member. I did. I went to their website, and they're actually making a robot to save the universe. So, there you go; there are people who are doing that.

I like what Carlos just said, and I'm fully respectful of all of the panelists before me. I just want to reinforce that AI has made strides over the last 10 or 15 years on the computational side, perhaps losing touch with the cognition side, but I think I know a way we can get back in touch with the cognition side. Because though we accumulated a whole host of methods and techniques and algorithms, we haven't accumulated that list of great showcase applications in which to demonstrate them. So, most of the work I do starts with the mantra, what can AI do for you, riffing off of a UPS tagline, "what can brown do for you?"

I think there are plenty of opportunities in that arena. We'll gravitate to these based on sort of our own interests and our background and so forth, but I want to quickly run through four examples and four levels of granularity, just to give you an idea of the breadth of things that we can adopt that will really make our peers in computer science aware of what we do, our peers in the rest of science aware of what we do, and then our peers on the planet aware of what we do.

So, first off, there is the individual level, which really ties together the computational and the cognitive end of AI. Today, we have all these modalities. We can observe the human mind at work through FMRI, through EEG, and so forth, eye trackers that can look at what we're looking at, devices that can tap into our motor actions.

So, one of the things I've been working on is can we understand how humans learn specific tasks by tracking such information and fusing them, and understand why some people have difficulty with learning certain families of tasks.

A particular task I've been looking at is something the navy uses to differentiate between people who are going to be future submariners or not. But my dream is to see that being used in the K through 12 classroom. Imagine, I mean, by this Christmas every kid in America — well most kids — will have this Wii cap that they're going to wear along with their Nintendo Wii system, which will allow the machine to infer their emotional state, making the game harder or easier depending on their level of frustration. If we can process that kind of information and give it to the classroom teacher in a third-grade math class, can you imagine what we can do? That is an individual level.

We can do this at the city level, too. One of the things I'm doing right now is working with the city of Houston to help it plan for evacuations under disasters such as hurricanes, and the simplest thing that we've been able to do is bringing together decades of research on structural engineering, assessing the viability of homes with respect to wind, flood, and so forth, and making that information available to all of the citizens at appropriate times so they can make rational decisions. Often people flee because they can't answer, "is my house going to blow down if this hurricane comes over?"

At the country level, can we build models of the evolution of conflict by tracking news media over time, longitudinally from whenever online information is available? The answer seems to be a qualified yes.

For example, our system could have seen the Kuwait takeover by Saddam Hussein about four weeks before it actually happened. Why could it do that? It turns out we had a nice model of Saddam Hussein, who turned out to be a fairly predictable fellow. Before he would engage in serious conflict with one of his neighbors, he would engage in a very strategic dance with his neighbors, which I'll characterize in sort of third-grade playground terms. If you're going to fight, take on somebody on the playground, you want to find out who's with you — to use our fearless leader's words, who is with me, who is with us, and who is not. So, you can see that pattern. AI technology, vision networks, and the kinds of beautiful things that Eric talked about,[2] for example, can be used to do that.

Finally at the societal level, what can we do that will impact society? I loved Henry's suggestion of taking on health care or taking on energy or something like that.

The creativity is going to come from us and our students, who are going to see these opportunities. I have a 10 year old, and the greatest difficulty for me is to make her do word problems. She can add, subtract, divide, and do all those operations. We can do Bayesian inference, do all of the computational stuff, no problems at all, but when a problem comes knocking, can we see that here is where we can apply technique X for machine learning and so forth? How do we train the next generation of students to recognize and leverage these opportunities? To me as an educator that is the biggest challenge, and I'll leave you with that.

Lise Getoor: I'm going to say things that definitely echo things that have been said so far, but one of the things that I wanted to mention, following up on what Jim said, is recently there has been a number of AI anniversaries. So, first off, there was the anniversary of AAAI, the organization. There was also the 50th anniversary of the coining of the term *AI*. So, there's a lot to celebrate, and you can see a lot of the things that have happened. And in case any of you haven't been to the main AAAI conference in a long time, I would really encourage you to go, because in the past few years I think there's been a lot more energy, there's been a lot of new developments that have been added to the programs and so on.

But if you look at these retrospectives, you do see some common concerns. One of the common concerns is the basic kind of fragmentation of the field, that there is a lot of subconferences and so on, that we've lost sight of the bigger AI goals, human-level AI, and in general the crisis in computer-science (CS) education.

I want to argue that actually these are now turning into opportunities, that there are a number of ways in which these have changed, and the first one is in terms of fragmentation of the field. I think we've gone from fragmentation actually to collaboration, and many of the things that people have talked about on the panel illustrate that.

My research area is representation, reasoning, and learning methods for combining uncertainty and logic, and I think — of course, I'm completely biased — but I think that they're great for allowing us to deal with noisy, heterogeneous data, and provide the kind of context-sensitive, adaptive, resource-constrained reasoning that we want. And I do want to hit on also supporting the kind of social intelligence that I think Jim alluded to, and which we've seen in this summit discussed a lot more, not so much in the AI track but in other parts of the conference, and I think that that's important.

At least for me, and I think this is true of the other panel members that I've talked to, this has allowed collaboration for me across subdisciplines within AI, but also across CS and outside CS. So, I think that there's a lot of exciting opportunities, and these opportunities actually segue into supporting going to newer and bigger goals, which are these AI complete kinds of problems where we really are kind of making a difference. We've seen and heard a number of these. So, Carlos highlighted these very well in his talk, Eric as well, and Devika and also Henry.

The one that I haven't heard talked about quite as much is these new kind of social information processing kinds of things. There was a AAAI spring symposium[3] — Kristina Lerman was one of the organizers of this — on how you collect together information from a bunch of sources, how you integrate it and align it.

And the thing that really impressed me was the really innovative applications that people had for this, things as diverse as saving the rain forest in South America to intelligent map building and so on. I think that this is a really exciting area.

And then in terms of CS education there's been some recent events that I think are really exciting. There was an AAAI spring symposium on using AI to motivate greater participation in computer science, and there was also a teaching forum at AAAI that had a lot of neat things going on, including the AI and education colloquium. I encourage you to look at these things. Mehran Sahami was one of the big organizers, also Marie des Jardins, and Adele Howe, and a lot of others.

One of the exciting things here is this notion of AI has developed enough. I remember the first AI class that I took I did not like at all. This is going to date me. But we studied arches and semantic networks and so on, and it seemed like a random collection of algorithms to me. Now when I teach AI, we go through representation, different types of representation, reasoning, and learning, and that gives the theory that I think supports computational thinking, and also the applications are actually compelling and relevant, and I think that that's really exciting.

So, the message, first off, is collaborate; it's fun. I joke with my database collaborators, the other good thing about collaborating is AI conferences are usually not in the greatest locations, but database conferences, for example, tend to be in better locations. Work on problems that matter. So, this is echoing a number of the things. And educate. And there are lots of challenges in terms of computational complexity, privacy, and I think the visualization to support the inference, the HCI kinds of issues that started off, Joe started off the panel with. And I really think that you can have theory and apply it too.

James Hendler: The day I really became an AI researcher was the day I stepped into my first AI course; Roger Schank was the professor, and he went on to become very controversial, but Schank said something that changed my life. He said, you know, what you're going to learn about in this course is a lot of stuff we don't know the answers to, and any one of you can go on to become a big player in this field and can really solve hard problems, because there's so much we don't know. The problem is, 50 years later I feel like we still don't know a lot of that stuff; we just know a lot of other things, and we have forgotten that we have to keep reminding our students that there's a lot of excitement about the stuff we don't know how to do.

So, Lise, I have a slightly different philosophy from you when I teach my AI class.

Eric Horvitz: A surprise about your comments — not a big surprise — is that I didn't hear very much from any of you on technical issues and opportunities. Let me ask one question maybe to break the ice in that department. Let's say we assume that a decade from now you're told, looking back, that there were two big surprises — there might be more — but two big surprises happened technically in AI, that for us old-timers, appeared amazing in retrospect. What are these surprises that we might encounter technically in terms of things becoming more doable, or a new discovery, for example? I'm giving you the partial result now, and I'm asking you to compute what the surprise is. Anybody have a response to that?

Michael Wellman: I don't want to ruin the surprise.

Eric Horvitz: Consistent with the earlier remarks, Mike.

Henry Kautz: I would say a couple things that have surprised me. One would be the surprising effectiveness of approximation algorithms for uncertain inference, and belief propagation, different kinds of local search methods. I think another more recent surprise is the advances in the statistical relation models. With each paper it just seems like, oh yeah, that's obviously the way to do it, and I think that's kind of surprising going back like 10 years.

Eric Horvitz: And going forward? Even the category of the surprise coming? You're looking back 10 years from now, 20 years from now, what were the two — at least two — big surprises that occurred technically?

James Hendler: Let me do two, one positive, one negative. I'll start with the negative. I think 10 years from now we're going to be stunned by the failure of the integrated AI system. I think it's a wonderful thing we're trying it, and, they're doing wonderful stuff, but when we actually look at what they do versus millions of dollars, a team of the top people, 18 schools involved in making this thing, 5 years from now we're going to look at the demo that some high school student does, and say, gee, what were we thinking?

So, I think one of the things is not that it's bad to do integrated AI, but that most of the big integrated AI projects are trying to do things that are already well understandable within the context of the single AI problem.

I think the second thing that's going to surprise us — one of the big unknowns in AI right now — is memory. We as humans deal with memory drastically differently than databases or computers do. We're learning a lot now at the neuro level about what some of that does, but we're also seeing a lot of people working on very different models of what kind of information space you create, and as the computers start to catch up to that, I think we're going to see the ability to actually start doing some of the things we've been ignoring like what does this remind you of and things like that, and I think that's going to make a qualitative difference in AI in a way we can't even imagine right now. So, I'm hoping for that surprise.

Carlos Guestrin: I'll be a little bit more controversial, since this is supposed to be a panel. One thing that I've been thinking about quite a bit is the complete death of models. Graphical models, the thing that I know and love, that's going to end. I think the reason is that we focus too much on having one model for the way the world works, and then committing to that model, which is an approximation, and then trying to do inference on top of this. So, I think this entire pipeline is, in my opinion, not the right way of solving the problem. So, I've been with one of my students who is interning here, Dafna, rethinking this pipeline, and I think we're going to kind of change the way we're thinking about problems.

Devika Subramanian: My prediction will be that the biggest surprise is not going to come from inside our community, but actually from someone outside, working on a hard enough problem that pushes the limits that will inform us about our own models. The reason for the death of the models I think, which I agree with, by the way, even though I know and love and use models in all of what I do, is that we're going to make a new family of what I will call lightweight models. The models we have right now are heavyweight models. We're going to make models quickly, because the world is changing. We're going to attack nonstationarity at its core, and build very lightweight, throwaway models, and keep redoing that process, and integrate that in sort of the inner loop of a pretty fast computation, as opposed to the get me 20 years' worth of data on X, and I'll tell you what will happen today.

Carlos Guestrin: This is exactly where we're going.

Devika Subramanian: Oh, good. Then you and I should talk.

Joe Konstan: This is obviously a no-win question, because if we're right, then we predicted it, and it won't surprise us.

Eric Horvitz: That's okay. We'll call it prediction.

Joe Konstan: But I'm going to take a guess here, I hope. I don't think so, because everyone is listening to us. But I think we're going to see the end of the era of attempting to solve things solely through computational intelligence, and greater embracing of systems that bring in at the very least human intelligence, whether that's explicit human in the loop systems, whether that's involving purposeful games in the style of Luis von Ahn, but possibly also animal intelligence, that we may end up that you shine lights on a colony of ants as part of the computation that solves a hard problem because we realize there are things that we just don't know how to compute but that we can infer from others.

I also think the idea that this is the same problem from 50 years ago shouldn't be taken as a criticism of the field. That's actually one of the strengths of the field. It's the same thing as the challenge to go out and explore the universe that you will never meet in anything you achieve; you just realize there's more universe or understanding the origin of the universe. There's something really powerful about pursuing a challenge that you know you will never achieve, because it allows you to celebrate everything you did along the way as an accomplishment.

Lise Getoor: Connecting to the fragments on models and also to Joe's comment, I think that there is the opportunity for keeping track of our context, and having multiple roles that we are able to store and to actually take with us different places, and we're starting to see that. And I think the thing that will be really interesting to see is how that develops in broader society. There's lots of discussion about privacy and things like that, but then if you look at what's happening with kids and how they view privacy, it's not from the model that I deal in terms of "I don't want anybody seeing my e-mail" or something like that.

So, this notion of managing identity and keeping account of context, and then being able to share that with others so that you can do more things than you could on your own, and connecting that into kind of augmenting intelligence and so on is I think fascinating.

Michael Wellman: Just to follow up on Joe's point about working on these problems that are well beyond things that we're likely to be able to do, I think that's a typical form of a kind of surprise. If you look back to the early days of AI and you see these people who are working theorem proving and chess and natural language understanding, and nobody was working on word processing, other things it turns out computers were productive at long before they solved the problems that people were addressing.

Frankly, that's one of the things I love about AI, and attracted me to it originally and keeps me in it. Even though I don't view myself as working on the AI complete problems primarily, I think it's very stimulating to be around the community that is, and so that I think that it's just a mark of the ambition.

Just an observation to Carlos: If you're looking for an AI complete problem, I can reassure you that it doesn't matter which one you solve, by definition. (Laughter.) And if you don't quite solve it, then it didn't matter that it was AI complete.

Carlos Guestrin: Thank you.

Eric Horvitz: On that note, let's open it up to the audience.

Audience Member: I'm not sure your problem is AI complete, and that's what I wanted to press on. I don't know everyone on the panel, so it could be that's just the way Eric picked everybody, but there was nothing about the fact that we're making great strides in understanding how the brain actually works, and that kind of technological advance, tools to do that, and how it might influence AI, and I'm a little surprised.

Eric Horvitz: So, to amplify that, I was going to — during a lull in the conversation — throw this to the panel, too, resonating with the intent of your question. But we actually do have existence proofs, unless you are a deep believer in something else going on. We have existence proofs of computation creating all of this, this cognition, and our abilities as humans and even the magic of other vertebrates and invertebrate creatures that have nervous systems. So, one question also is: Might there be a surprise in the link between the two? Might we actually understand per representation or modelless reasoning or just-in-time modeling or small models — what's really going on with these naturally evolved tangles of cells that seem to be so marvelous in their abilities? I hope that you view that as an amplification of your comment.

Audience Member: Yeah, I mean, it's related. It's only been a few years that we've had tools, and I'm just surprised that there's no comment on that.

Lise Getoor: I think Eric mentioned it in his talk [earlier keynote presentation at the meeting by Eric Horvitz]. The amazing things going on with the kinds of imaging that you can do now, and then trying to kind of connect that with other kinds of information that you have about the various functions and so on, and being able to do some sort of statistical analysis to propose models, and then have those models be things that you potentially go in and verify in some experimental way, I think there is a potential to use the advances in computational intelligence to help do the science and vice versa.

Carlos Guestrin: There's amazing things going on. You probably saw Tom Mitchell's stuff on where they can predict what you're thinking for words that they've never seen data for, which is pretty cool.

But I will say to this, that I'll be highly surprised if the things that we do in AI in the next 50 years will be highly influenced the other way, meaning that the systems we'll build will directly mimic in some way the way that the brain works. I think some of the models we're building might give us insight into making predictions from the brain, but not necessarily the other way.

Eric Horvitz: I'd like to interject, that I started out as a Ph.D. M.D. student in neurobiology, and in my first year decided that sticking little electrodes into cells, even though I was getting somewhere near where thought was happening in these creatures, was probably about as relevant to cognition as putting a little wire into the Apple II computer next to me at the time, and trying to infer the operating system or application level of semantics.

Devika Subramanian: I think there is work going on within the field. Since I am familiar with my own work, let me just throw it one direction there. I think interpreting the output from these amazing devices, and actually figuring out what it means — one example, right now by correlating visual, motor, and EEG activity we're building models of how humans learn complex tasks with strategic as well as visual motor components. We're finding that some people have great difficulty in actually translating their strategic components — you know, for their frontal lobe activity back, and they don't have the visual hand-eye coordination. If you observe people at this task and they're failing, you can't tell by just looking at the fact that they're failing whether they're having strategic difficulties with the task, unable to basically come up with decision-making rules, or just an inability to execute those rules. So, there is work going on now, which uses all these new modalities to interpret and diagnose particular types of learning difficulties. So, that's one way in which I think AI and AI techniques, and particularly in, say, learning, can come and help. I would hope that we'd go beyond and shed some light on brain architecture, but we're not there yet.

James Hendler: I think there's a lot we can learn as we do learn more about brain architecture and things like that, but I think there's still a limitation, and I think it's an inherent problem in some assumptions we made roughly 50 years ago about how to study AI, which is the study of the individual entity. I mean, there's a lot to be said, but sort of let's use the opposite existence proof, right? If you have a kid, you lock them in a closet and you take them out of the closet 30 years later, you don't have a very intelligent entity. Or put them in a nice closet, put them on a desert island where all of their dreams come true, I mean, you still don't have an intelligent entity. So, again it's not the deprivation thing I'm talking about. We still aren't at the point where we can start looking at two people communicating within FMRIs. We're still not at the point of saying how does hearing something from someone you trust somehow affect your memory later differently from if you had heard it from someone you distrust, sometimes in surprising ways, et cetera, et cetera. So, I think there's a lot to be done there, but I think when we start looking at what we don't know — again, most of what we're trying to get out of the current brain modeling is how did the stuff that we're actually starting to get pretty good at in AI work. I think there's a lot of distance, a long way to go before we really can say the brain-inspired stuff is taking us to the really hard problems.

Audience Member: I was very intrigued by the statement made that the greatest discovery might come from a really hard problem that pushes AI beyond the study of the brain and cognition. Do you have any other examples of what that might be, any field?

Devika Subramanian: Well, limited to the examples that I have worked on, I'll just say if we can figure out a way to kind of solve hard combinatorial optimization problems, the ones that are naturally occurring, that humans today solve also by approximation methods, if we're able to kind of try to tackle that, as Carlos has shown it's possible, if you can do some analyses of it, I think that would be the recipe for the breakthrough. I think it's not by studying how other organisms solve the problem, though that's one way of doing it. I want to use the problem itself, independent of who else solves it, as the motivator for it, and an absolute benchmark on how well we can do on it as the driver for the problem, as for innovation.

Joe Konstan: So, to give a completely different problem, take management of volunteers. If you think of the number of voluntary organizations that are out there in the world, and the small number of people who are really good at running these, and you ask, what could I build into a computer system, if I fed it the data or if it could gather the data on what people had been assigned to do, who they're doing it with, what they thought when they were done, all of the data I have, and you come up with this system, not necessarily to do this autonomously, but to support somebody in keeping your volunteers engaged, healthy, developed, all of that, and you think of all of the cognitive, the social, the sensor data, the fusion of different information involved, I think there's a huge amount of AI, as well as HCI, as well as perhaps non-AI computation that will have to be solved in order to make a dent into a problem like that, that has huge social importance.

Audience Member: It seems to me that AI has mostly approached the question of deciding about action as maximizing or approximately maximizing a well-defined objective function. And it also seems to me that there is very little evidence that this is the way that people decide on how to act. In fact, there's even well-known experiments that show that people's behavior isn't even consistent with any objective function, let alone are people aware of what objective function they are responding to. So, I wonder what you all think of the persistence of this kind of objective function model of action in AI, and what alternatives to it you see.

Michael Wellman: I think first we need to separate the different scientific goals in AI that people are pursuing, and a lot of AI really is still about engineering competent behavior. For that purpose, having well-defined objective functions follows good engineering principles, and it's from that perspective somewhat irrelevant what humans do. AI people throughout have debated these different goals. I think the really fortunate thing is that now we can really pursue both without any conflict in that there's so much relevant demand for humanlike AI, that really is humanlike because it's going to be used for training or for entertainment or for other things, where actually being like humans is itself important. So, we'll have the opportunity to develop those theories that are going to make decisions the way people do for good reasons, and we won't have this conflict anymore. So, those of us who are in the maybe near term concerned about competence can still keep the principles that we know, and then later maybe if when we find out how to make things human like we can now compare and contrast their strengths and weaknesses.

Carlos Guestrin: If I could take this into a slightly tangential but controversial level, I think one of the big advances of AI in the last couple of decades or so is the definition of objective functions. There's a lot of work in AI before that was about I did this, then I did that, and then look at my answer. So, I think it's been a good thing for us, although we might have overfit to this idea. So, I agree with Michael with everything you said.

James Hendler: I can't disagree with any of the specific words Michael spoke, but I had such a visceral negative reaction to it, that I know I disagree. (Laughter.) And really where I'm coming from on that is again the fooling ourselves into thinking we're making progress, because techniques we already understand well can be applied to yet another problem and yet another problem and yet another problem and yet another problem. That's not even — not only is that not science, that's not engineering, right? Engineering and science are about solving problems we don't know how to solve yet. They're about attacking new and different things. It's not about application building. In fact, my visceral reaction to a lot of things I've heard on this panel is AI has become about building applications, specific applications, not engineering principles, which a lot of the people on this panel got famous for doing, is not solving a particular problem using *X,* but inventing the technique *X,* which is now being used to solve a lot of problems. Okay, your work in inventing this, well, what have you invented for us lately, right? Now you're just applying and applying it. And I say this — I say this in a funny way, but if you think about it, we as a field have forgotten about innovation. I don't care whether you're doing it for cognitive reasons, I don't think it matters if you're doing it for anything, but we have forgotten that the world out there is this amazingly complex and interesting thing to view from an intelligence perspective, and that building a better cell phone isn't the job of the scientist. It's understanding the principles that let someone else build better.

I really, really think that we've lost this.

Eric Horvitz: Let me just defend the people on the panel, that Henry, for example, has done some wonderful recent work in model counting, and his work on foundations still continues, even though he's been working on some interesting applications.

James Hendler: I think the definition of what's a foundation has changed drastically.

Carlos Guestrin: I totally disagree. Yesterday I talked about work with my student Andreas Krause[4] that I think is extremely general, and was applied to a wide number of applications, and I tell you how that work started. I was working at Intel Research in Berkeley on a group that does sensor networks, and they were deploying sensors in a forest to understand the microclimate around Redwood trees. I had a chitchat with one of the people who were deploying sensors, and I asked, how do you decide where to put sensors in this forest? And he said, well, wherever it looks good, here or there, I just put sensors up, so it's all good. And I thought, okay, I'll do a project for a month, and help them out, and move on. And somehow we ended up in this huge, very interesting area, which Andreas had a big impact on, that I think is a fundamental principle in new understanding in AI, which totally was motivated by an application domain.

James Hendler: Well, look, don't get me wrong, I absolutely am not saying it's bad to work on applications.

Devika Subramanian: I too had a visceral reaction to what you said, and I used to be a theoretician in my former life, and that was only 10 years ago, so it's not that old. I think it's a little bit naïve to think that we developed theories — so, in my former life I developed theories, and all I've been doing for the last decade is just punching out, you know, working on a factory going stamp, stamp, stamp as the applications roll by. (Laughter.)

Eric Horvitz: That's a great metaphor. I love it.

Devika Subramanian: In fact, I have had to forget all the theory I did, and rethink it. What has emerged instead is much stronger theory. I gave you three examples, one actually interpreting building models of how humans learn tasks, doing evacuation planning for a major urban city of 4 million people, predicting conflict by reading news reports across all online sources over hundreds or however many years are available. They all share a core set of computational principles and models and methods. And I understand them much better, even though I was responsible for creating some of them, and I think a much leaner, meaner theory base has emerged by immersing myself in actual problems. So, I don't think the advances in AI over the last 10 years are an accident. The spurt in theory has come because we have been forced, in many cases by our funders, to actually find actual relevance for this. This has made us more creative I think. It's not an either/or with you do theory or you do applications, but really they go hand in hand and we've got each to drive the other. I can't go and write a paper for AAAI saying, and here's what I did, look at what a great system. Even Eric in his [earlier plenary talk at the meeting] kept punching it….

Eric Horvitz: Even?

Devika Subramanian: Right, even Eric. He gave us all these lovely examples, right? I loved your talk, because you had all these examples, but if all I walked away from your talk was, oh, and you can do Smartflow, and you can do this, and you can do that …. What I saw was, oh my God, Bayesian networks and decision-theoretic reasoning can really influence this whole plethora of things, and I'm sure you had to innovate in so many different ways to make each of them.

Eric Horvitz: I should say that the reason I'm passionate about these applications is that they help me explore the problems with taking closed world models into the open world, and to better understand theories in each case.

Devika Subramanian: Absolutely, and I think it's crucial for AI.

James Hendler: So, I'm the scruffy on the panel, right? So, I'm the last person to say applications are bad, but an awful lot of what we are now teaching our students to do is not really to think as creatively as many of us were taught to think when we were students. Let me see if I can explain it this way. This was the thought experiment I played when I was at DARPA, to try to convince some people to put some money into AI. Supposing you take the things we know how to do in AI and you kind of take this big table, and say that's the space of applications that we know how to do AI, and you take this technique and you say which of those could it do, and it covers a big piece, and you've got the next one that covers a big piece, and the next one that — okay, so now you've got your table covered with these circles, right?

Well, two problems. One is there's a lot of stuff outside the table that those circles aren't covering, but the second one is we still don't have a technique where we can cover that whole table, because each of those circles is focused on a different way of looking at things. They make contradictory assumptions to each other. So, the metareasoning that was originally in MYCIN, and was talked about 50 years ago and 40 years ago as a key thing in AI, this notion of really trying to plan through a space of techniques and a space through problems to solve, right, you don't hear about that so much anymore. And Lise was right about talking about bringing some of that back, that these are opportunities, but again very often the application space follows what we know how to do with the technique. What's nice is when you get a big problem for us, when you like the sensor problem, where you say, hey, the stuff we do doesn't work. That's where new invention comes from. But, in fact, in AI we've become much more averse than we once were as a field to work on the stuff we don't know how to do.

Audience Member: A couple of the panelists expressed some dismay at grand challenge problems, and I wanted to draw you out a little bit more on that. If you take something like the challenge of driving a car through a busy city, that requires a lot of forms of perception, audiovisual perception, fusing these forms of information, planning, understanding what the people driving around you are doing, trying to decide if the person to your right is trying to cut you off, and if so, if you're going to be a nice guy that day or you're going to be aggressive and try to do a counter maneuver. What's not to like about that sort of thing?

Lise Getoor: I think it's great, and as a matter of fact I have a much lower AI complete problem and I think it should satisfy Jim as well, because it's something I don't know how to do, and it's help dealing with information overload, with my e-mail inbox. If I did have an intelligent assistant that can help me sort through and be more productive and figure out which things are important and get it on my calendar list correctly and so on, I think that that would be great. It's something that helping to understand how to organize that information, reason about attention, reason about resources, and reason about the social context of the messages — the stuff that's not actually in the message that I know, learn from feedback in what I do, I think this would be great.

Eric Horvitz: So, is the dream you come in, in the morning, and look at your sent mail folder to see what's going on? (Laughter.)

Lise Getoor: That would be awesome.

Michael Wellman: The question of challenge problems is actually related to something that I've thought about a lot, which is the role of research competitions, which are more and more common.

Basically since there are so many worthwhile problems to solve, it's not that we need to invent new ones. It could be that there is some overall misallocation and no one is focusing on this key combination of capabilities that would put it all together, and so you want some kind of coordinating force to do that, or you want a coordinating force to get people to focus on some kind of common domain just for the purpose of being able to build on each other's results and compare them and combine them. But you've got to have a balance, because you don't want central creation of problems, because that may miss the opportunities that you get when you have a whole community of people also inventing the problems as well as the techniques. That's the trade-off.

Audience Member: So, in the remaining three minutes I'll shift the discussion to the small topic of privacy. It was striking to me that in the panel's opening statements there was a reflection on this kind of cognitive versus computational approach to AI, and a recognition that privacy posed some sort of challenge.

I'm curious to know whether you think it's just a stumbling block or something that AI in one version or another, depending on your flavor up there, could actually help with and what that would mean and how you would approach it.

Eric Horvitz: Well, I did comment in my talk yesterday in the opening keynote that this is a critical opportunity area for some of the methods that we work with, but we might have other comments here on the panel.

Michael Wellman: I think it's a big problem for AI because if AI succeeds, then that proposes great threats to privacy, because of the ability to use information. Potentially AI could be part of a solution. However, it doesn't seem like there's a great deal of work going on in that direction. The way I think my own view of the foremost solution is having better ways of accounting for use of information. I think that obstructing collection of information is not going to work, but if we could somehow have better systems for — either though audits or online — making sure that information is used for intent that it's purposed for, that's something that potentially AI could contribute to, but there's not a whole lot of current backing for that, as I see it.

Joe Konstan: On top of that, if we understand the human side, if we understand what people will regret, find disturbing, find objectionable, we can build hybrid technological human systems, because we've already seen people are not very good at anticipating what they're going to have trouble with in the future. And if you can bring in AI support to help people prevent situations that they're going to regret on the privacy dimension, I think that's another area where AI can help.

Eric Horvitz: I'm curious to hear if you have reflections about potential disruptions to our society, good and bad, that might come based on developments that come out of the fires of our technology, in the next 20 years, for example, 25 years.

James Hendler: I'm going to channel a colleague of mine, a colleague of all of ours, Noel Sharkey of Sheffield. Noel has been writing about the notion of military battlefield robots for a while now, and pointing out how as we've moved forward with the technology, we've been lowering our expectations of the criteria before we're going to let the machine pull the trigger. People, surrounded by computers and seeing so much that can be done by their machines and by the web and by things like that, who don't understand the technology, actually think there's far more capability in the system than they have right now. And I think it's that lowered expectation — that lowered opinion of what a human is compared to a machine opens the door for just this huge amount of abuse, and I think there's plenty of people out there who will be very happy to abuse it if we let that happen. In fact, one of the places where a lot of my thinking about needing a bigger definition of AI or needing to embrace what we can't do comes from is again because of people's expectation that we could do all this stuff.

Carlos Guestrin: I want to take a more positive view, if possible. I don't think it's a disaster. I think the web has really changed things, as we all know, and I believe that AI has had a big impact on this, even though we don't get much recognition for it. As this technology changes and improves, I think the way that the web has revolutionized the way that we think right now, I think AI will do the same way.

If you think about how machine translation systems, for example, could bring people together, how automating the number of tasks that we do could actually let us think more and get away from more of the issues of every day, I think this could be really, really impactful and really amazing for us as a society.

Joe Konstan: I think that's wonderful, but I'm not willing to give up on doom and gloom yet. I think there are some examples out there that show that it's easier and there's greater incentive to develop systems that support individuals than systems that support communities and societies. You see this in the stock market. Why do we have mechanisms where humans stop program trading?

Carlos Guestrin: How about this? You name one, I name one, and then we'll go back and forth to see

Joe Konstan: Well, we could, but the question was about disasters.

Eric Horvitz: We have a couple minutes left, Carlos, so I don't know about getting to a Nash equilibrium there.

Joe Konstan: There's hope here, and there's a challenge here, because we see a lot of systems that basically help people with greed or with greed without respect to the good of society. If we're going to have this not lead to decay, that means people have to adopt challenges of developing systems whose goal or whose client is the collective rather than the individual. And I think we're capable of doing that, but I think the incentives haven't been set up to induce people to put nearly as much effort in that direction.

Eric Horvitz: Okay, we'll stop there and thank our panelists and the audience as well. Thank you very much. (Applause.)

Notes

1. This article is a based on a transcript from the 2008 Microsoft Faculty Research Summit, July 29, 2008 (research.microsoft.com /en-us/events/facsum2008/). A video of the panel is available at research.microsoft.com/ en-us/UM/redmond/events/MSRNVideo-Content/FacSum08/16063/lecture.htm.

2. Earlier keynote presentation by Eric Horvitz at the meeting, "Reflections about Directions in Artificial Intelligence," talk available online at: research.microsoft.com/ en-us/um/redmond/events/msrnvideocontent/facsum08/16075/lecture.htm.

3. See Social Information Processing: Papers from the AAAI Spring Symposium. AAAI Technical Report SS-08-06. www.aaai.org/ Library/Symposia/springsymposia-library. php#SS08

4 Earlier plenary presentation at meeting by Carlos Guestrin, "AI, Sensing, and Optimized Information Gathering: Trends and Directions," available online at: research. microsoft.com/en-us/UM/redmond/events/ MSRNVideoContent/FacSum08/16087/lecture.htm.

Eric Horvitz is a distinguished scientist at Microsoft Research.

Lise Getoor is an associate professor at the Institute for Advanced Computer Studies, the University of Maryland College Park.

Carlos Guestrin is an associate professor in the Machine Learning and Computer Science departments at Carnegie Mellon University.

James Hendler is a senior professor in the Departments of Computer Science and Cognitive Science at Rensselaer Polytechnic Institute.

Henry Kautz is a professor and chair of the Department of Computer Science at the University of Rochester.

Joseph Konstan is a distinguished professor and an associate department head at the Department of Computer Science and Engineering, University of Minnesota.

Devika Subramanian is a professor of computer science at Rice University.

Michael Wellman is a professor of computer science and engineering at the University of Michigan.