

Stuart Russell on the Future of Artificial Intelligence

AI may not take over the world but it will provide new and powerful tools. Smart microwave ovens?

*No big deal. Full-size humanoid robots that walk, climb stairs, open and close doors,
and pick things up? Now that gets our attention.*

Stuart Russell is a leading researcher in the field of artificial intelligence. He is a Professor of Computer Science at the University of California at Berkeley, Associate Editor of the *Journal of the ACM*, and author of "Artificial Intelligence: A Modern Approach" (Prentice Hall, 1995, 2003), the leading textbook in the field. His research interests include machine learning, limited rationality, real-time decision-making, intelligent agent architectures, autonomous vehicles, search, game-playing, reasoning under uncertainty, and commonsense knowledge representation.

UBIQUITY: The original grand vision of artificial intelligence (AI) in the 1950s and '60s seemed to dissipate into many small, disparate projects. Should this fragmentation be written off as an inevitable Humpty-Dumpty problem or is it possible to bring the fragments back together into a single field?

RUSSELL: I think we can put it back together in the sense of being able to join the pieces. Of course, the pieces won't be subsumed under one Über theory of intelligence. The subfields that focus on vision or speech or language, for instance, will still exist but I think we'll have compatible theories that will enable us to connect.

UBIQUITY: Will they be based on a probability theory?

RUSSELL: Yes. Speech has already gone this route. Speech recognition is a giant calculation of posterior probabilities from evidence. Once you do that it's easy to connect to the raw data and relate it to high-level representations. In the case of speech, the representations are word sequences. In the same manner, I think ultimately we'll come to understand vision. In the case of vision, it will be descriptions of scenes and objects in relation to an object. At the same time, logical AI tradition has broadened to include probability theory. A lot of high-level representation, reasoning and planning can go on in a probabilistic formalism.

UBIQUITY: Do you suppose the AI grand vision could be an umbrella for more than AI itself? Could it be used as a vehicle for interpreting other things?

RUSSELL: What might happen is that as we understand better how to build these artificial systems we'll start to have theories that are powerful enough to explain some aspects of human cognition in ways that would influence people's understanding of culture and sociology. Currently, there is not much constructive flow from cognitive science to the humanities. One area that I think we'll see substantial progress in the near future is flow into scientific methodology itself.

UBIQUITY: Can you provide an example?

RUSSELL: One of the developers of Bayesian networks, Judea Pearl, recently developed a way of interpreting data so as to determine the cause of things in the real world -- does A cause B, or does B cause A, or is there a third hidden cause for both of them? The statistical view is that you can never determine causality. You can only determine a statistical correlation. The tobacco companies have used this excuse for many years in court cases. They said, "So what if lots of smokers get cancer? Whatever it is that causes people to smoke also causes them to get cancer. It has nothing to do with the smoking." Pearl developed a theory that mathematically determines causation. That

theory, or some refined version of it, will play a huge role in areas such as sociology where determining causation is the big argument. I believe that we'll start to see people constructing very complex scientific hypotheses using probabilistic modeling and machine-learning techniques. But this is a long way from saying that AI will constitute the intellectual enterprises of the human race.

UBIQUITY: It is a long way in time, but is it also a long way logically or rhetorically? In other words, if it can do such things, is it plausible to suggest that it could be the starting point intellectually for political science, sociology or anything else?

RUSSELL: I don't see it that way. I think it will produce useful tools. Think about physics for example. Particle physicists make massive use of computers for interpreting their data, detecting events in giant accelerators, and making predictions from theory because it's too hard to make a prediction from a theory without a computer. But I don't think that's had a huge impact on the subject matter of particle physics. It's a useful tool that's made particle physicists much more productive.

UBIQUITY: In that case you are talking about computing itself. Isn't artificial intelligence on a different order intellectually than computing?

RUSSELL: Let's take the case that we actually develop somewhat useful, predictive theories of human cognition. Then, let's say you want to start doing sociology. Those theories will be part of the armory of a sociologist but sociology will still be about trying to make sense of what's going on and deciding what are the interesting questions. They'll just have an extra tool. I don't think the tool should ever become the primary focus. If you're a sociologist, your primary focus should be on society, not on the tools that you use to help explain what's going on in society.

UBIQUITY: Is it on the same level as mathematics perhaps?

RUSSELL: In a sense, yes. Computer scientists use a lot of mathematics, but we're interested in computation. Mechanical engineers use lots of mathematics, but they're interested in mechanisms and design. And maybe sociologists and economists will use lots of AI models, but they'll still be interested in societies and economies. The only sense in which AI would take over would be if in fact we built artificial scientists that were a damn sight better than human scientists were and then it might take over.

UBIQUITY: Do you think that acceptance is important to the future of AI?

RUSSELL: I do think it's important to the future of AI. If you build something useful it will tend to get adopted but people won't necessarily care about how it works. They'll just assume that, "Of course my microwave should be able to tell when a thing has been defrosted." What examples do people have of an intelligent system? They have humans. Humans are very robust and work in a wide range of circumstances and can deal with all kinds of exceptions, whereas the intelligent systems that are being deployed in their everyday lives right now, for example, the speech recognition system that you get when you call information, are extremely fragile. If you have a vacuum cleaner going in the background, it will probably fail to understand what you're saying. People will find it irritating when it doesn't work.

UBIQUITY: Do you think that AI is having greater acceptance in the society now that embedded devices are becoming more commonplace in cars and toasters and everything else?

RUSSELL: I think we're still very much in the early days of the integration of any interesting intelligent systems into human life. When we start having humanoid household robots that are

interestingly competent, that will change things. Now there are full-size humanoid robots that can walk around, climb up stairs, open and close doors, pick things up, and put them down. There's something about a human-shaped thing that hits you at a physical level. A box or a trashcan-shaped robot just doesn't have the same effect. I think that the deployment of humanoid robots on a wide scale will probably start people thinking more seriously about what the future will be like.

UBIQUITY: Are any such things in use in real applications now?

RUSSELL: Not yet. Right now they're incredibly expensive -- probably one or two million dollars for the full-size one. Plus, you need to hire half a dozen full-time engineers to keep it working. But people have done demonstrations showing that it can operate a backhoe. Other examples are NASA, which is developing the top half of a humanoid robot for operations in space and Sony, which will be selling a somewhat smaller humanoid robot. It's called an entertainment robot and it's only about two feet high.

UBIQUITY: It's not easy for laypeople to know what is AI and what's not AI. Is that a good sign or a bad sign? For example, is TiVo considered AI? TiVo, of course, is a system that allows recording of TV programs, searches for shows it predicts the viewer will like, edits out commercials, and does various other tricks.

RUSSELL: Most people would call the predictive part AI, and it's partially a sociological accident of who developed those ideas and what field they identified with. There can't be a hard-and-fast line because every AI system is based on a computer program, but in practice AI tends not to be simple algorithms that are straightforward preprogrammed sequences of commands.

UBIQUITY: Is it the case that an AI program necessarily does learning?

RUSSELL: No, it doesn't have to learn. Deep Blue, the chess program, didn't do much learning, and certainly not while it was playing. But people would say that was AI because chess programs were one of the things that people in the AI community worked on. But the algorithm that's used for playing chess is something that could equally have been invented by a theoretical computer scientist. There are other gray areas too. Some people would say that Google is AI. Some people would say its databases. Some people would say its algorithm or theoretical computer science.

UBIQUITY: And YOU say?

RUSSELL: I'd say it contains some elements of all of the above. It's like asking, where is the dividing line between trees and bushes or bushes and shrubs? It's not clear that there has to be a dividing line. There's a cliché that as soon as something starts to work people no longer call it AI. There's some truth to that because once it starts to work then people can explain how it works. Once the mystery is no longer there, people say that's just an algorithm. There is a misconception that AI is only AI if it has a black box that produces intelligence in a mysterious way.

UBIQUITY: I'm sure you have a crisp definition of what AI is, right?

RUSSELL: A lot of people ask me this, and I made an attempt at a formal definition a few years ago. . . . An intelligent system is one whose expected utility is the highest that can be achieved by any system with the same computational limitations.

UBIQUITY: Is it important to have the highest utility?

RUSSELL: Obviously, you could be intelligent to a degree. But the key thing is taking into account the computational limitations and insisting on trying to achieve the maximum utility given those limitations.

UBIQUITY: Is the AI community in the US pretty much the same as it is everywhere, or is there a European school and a US school and so forth?

RUSSELL: Within the US there are several schools. There is still a substantial community working primarily within logical representations. There's what some people call the modern AI school, which has more emphasis on probabilistic representation, reasoning and machine-learning methods. There's also a connectionist school where people look at neuron net models of cognition. To lesser or greater extent they're inspired by the desire to model how the brain works and to connect up to cognitive phenomena. There's also the fuzzy logic community, which is very strong in Japan and in a few European countries, but is not very prevalent in the US.

UBIQUITY: Lotfi Zadeh, creator of fuzzy logic, is at your institution. How are his ideas viewed there?

RUSSELL: We're very fond of Lotfi but he's much more celebrated outside of the US than within the US. By and large, the US hasn't bought into the arguments in favor of fuzzy logic.

UBIQUITY: Is it possible to say that the different themes or schools that you mentioned are prevalent at certain specific institutions?

RUSSELL: It's actually mixing up fairly quickly. For example, Berkley is primarily a modern AI probabilistic school but Stanford has hired three dyed-in-the-wool probability and machine-learning

people so they now have a very strong modern AI group. CMU and MIT have also hired people in that area. It's pretty clear that the modern AI view has a positive gradient within the overall community. I think the logical AI community maybe feels that it might be becoming less relevant. I'm not sure about the connectionist AI community. I think it's very active in cognitive science, which is a version of AI that is actively trying to model and interpret human intelligence. On the computational side, the part that isn't doing psychology, the upper echelons in the neural net community have pretty much switched over to a probability modeling and statistical learning viewpoint.

UBIQUITY: So do all these different approaches merge into a happy family? Is it a happy world, or a contentious world?

RUSSELL: I think of the mainstream AI community as the non-fuzzy group. Where they come together -- which is not very often, as they manage to separate themselves very well! -- there's definitely friction. In the past there's been a lot of contention in the cognitive science community between the connectionists, who sometimes call themselves the sub-symbolic modeling community, and the symbolic modeling community, which was primarily people like Allen Newell from CMU and others. It was quite contentious. Whereas the relations between the logical AI community and the modern AI -- the people doing probability and statistical learning -- have generally been very friendly. My assumption is that gradually they will become a unified community.

UBIQUITY: One topic we haven't touched on in this interview is the perennial question people have: Should we be worried about AI taking over the world or worried about intelligent weapons?

RUSSELL: One could argue that artificial intelligence could be used, for example, by the military or by unscrupulous corporations in ways that wouldn't be great for the human race as a whole. There needs to be serious thought about how such technology should be used and controlled. The example that people use is atomic power and how with treaties and organizations atomic power can be used without creating a huge threat of proliferation. Eventually we will need to do something similar in terms of controlling how AI is used, what types of systems are built, and for what purposes. For example, if we build smart weapons, the next generation beyond the cruise missile could be something that can independently retarget itself and defend itself against attackers. You can easily imagine serious problems. If you had a few thousand of those and there were errors in the programming, for example, it might be very nasty indeed. I would like to see more discussion of that.

More on Dr. Russell can be found at <http://www.cs.berkeley.edu/~russell>

Source: Ubiquity, Volume 4, Issue 43, Dec. 24, 2003 - Jan. 6, 2004