AI and You

Transcript

Guest: Mark Lee, part 2

Episode 125

First Aired: Monday, November 7, 2022

Hello, and welcome to episode 125! Today we will conclude the interview with Mark Lee. He is Emeritus Professor of Intelligent Systems in the Department of Computer Science at Aberystwyth University in Wales, and we are talking about robots. Mark has a vision of conversational robots that learn from moving around in the real world and encountering people. Hence the theme of his book, *How to Grow a Robot: Developing Human-Friendly, Social AI,* which was published by MIT Press in 2020 and talks about how he trained the iCub robot to learn from its experiences as he took its development from newborn helplessness to ability levels equal to a nine-month-old.

Last week we talked about what meaning means to a robot, some connections with Alan Turing, the role of curiosity, and the relationship of computer vision to interpreting the 3-D world. Let's get back to the interview with Mark Lee.

You talk about engineering and with the other things you're talking about, do you feel that you're closer to engineering or closer to neuroscience and human development or biology?

Well, my view is as an engineer, so I say, "how can we build this?" and we've got all lots of data from these neurosciences, psychology, we've got all sorts of information. We've got stuff from evolutionary theory about why the brain is bigger in us than earliest species of *homo* and that's because the frontal lobe is bigger with *Homo sapiens*. Because, according to the anthropologists, we are more social than the early apes and animals. So, the social brain hypothesis says that our group size is roughly 150 people that we know very well. Whereas chimps have about 30 and their brain size is similarly reduced in the same ratio. But the frontal lobe is the bit that makes the difference, apparently, because the earliest species, like Neanderthals, had a big brain like us, but it was at the back where the vision areas that was developed, because they lived in the northern hemispheres in the Ice Age, and it was very dark and they needed better vision.

Where would advances in science help you the most in engineering and AI, or in neuroscience?

Neuroscience is useful, but I don't think it's going to -- most neuroscience works on the adult brain and even adult monkeys' brains you know. Very little work is actually done on infants' brains, because you can't intrude on such a brain. During surgery on an adult, you can actually do experiments, but you can't with children. And so there's lots of data in neuroscience about the structures, but not about how they grow and change. And there's an enormous amount of change in the first year and an enormous amount of neural growth, and decay of neurons. So what I would really like is more people working in this area, who are interested in making systems that behave like humans, and are happy to work with psychology data. And I think that's a problem that most AI people look to neuroscience; they don't look to psychology.

What would you like to see you if you look far down the road that you're on here as a goal, as a dream, as the product that is perhaps driving you to go in this direction, what would you like to see the outcome of that be?

Well, as I speculate in the book: conversational robots, social robots would be the goal of this sort of work. It would be very nice to see a robot you could sit down with and have a conversation. And the robot would know enough about the world that it could understand your interests and your motivations and talk to you about it. Once you can do that you can lead on to lots of other things, but I think conversation is the important thing and that's what this brain hypothesis, the brain size hypothesis about is the bit that matters is the conversation part and that makes us different from the other animals.

What do you think about the conversations that people are having with large language models at the moment?

Well, they're very impressive what they do, but they don't understand what they're doing, those language models. So Mary says something to a robot, Mary says, "John doesn't like Alice" or something. A language model could kind of figure out what that means, but you would never get the full context and nuances that you'd have in a conversation. And multiple levels as well, and John said that Mary thinks that Frank doesn't like her, multiple levels of mentalizing of what different people are thinking. And those sorts of understanding are based on what's going on in the world and if the systems don't have any connection really with the world, how are they going to reason about it? How are they going to talk about it? We're going to look at bits of text, of course, where similar things have happened.

Right.

But that won't last for long. Every conversation is unique, in fact, and you can't always look up the answers from recorded history.

Right; I think that at the moment they're pushing the boundary of how long those things can fool you in a conversation but the underlying model is what I believe Timnit Gebru described as "stochastic parrots": probabilistically-based engines that are operating on manipulating text that they've already seen. This goal of having a conversational robot is interesting in that it doesn't involve any actual movement, and so I wonder if you envisage the robot that could have these kinds of conversations like the one we're having now, where we're not actually constrained or required to move as a result of that, learning the things that they need to by sitting in a classroom listening to a philosophy professor at some point, but they have clearly already been embodied to the extent that they have that embodied understanding of the real world. Because you're not going to get an embodied understanding of philosophy, at least, not much, right?

Well, I mean, to get to that conversational robot stage, you've had to go through all the learning about the physical world to get there. So, it knows lot of that. And of course, it would have the

Web and the Net available. So, it could look up things; if somebody says a funny word, you can look it up in a dictionary, and find out what that means in terms of other words. So, when you get to that stage, it can be quite useful and quite powerful.

And could that then result in a reusable module, a software model that could be transported and implanted in a system that *wasn't* embodied, to have the conversation?

Yeah, you could download the robots' brain into a computer. But of course, it would then be frozen in time, in the sense of development. It would be able to look up things and interrogate questions, answers, and it would be able to provide but it would not develop further in the same way. So, things like jargon and the funny things we say that don't make sense, when you analyze them, the robot would not understand those, and then have to ask questions: "When you say these different bits of jargon, you say, what does that mean?" Then it would remember, of course, but all of that would have to be learned in the real world as we do.

What about the use of robots for physical tasks? Many of the applications that people are wanting to put robots to for instance, in Japan, they're very focused on eldercare and things like giving people baths, moving them around; very hard tasks, it seems to me, to get a robot to do, to even build the necessary hardware for doing that. I'm just reminded of how much of what we do, what I do, with my hands is dependent on these incredibly dense sensors on nerve endings on the fingertips that tell me whether I'm holding something firmly enough. And that's where a lot of money is being put in robotic development at the moment. What do you think of how that future is going to play out in terms of robots that have that, for say, that kind of application level of ability?

Well, it's quite a difficult area, I mean, the carer in the care home doesn't actually pick up people, they use specialized equipment to do that. So, the robots would have to do the same thing, they'd have to be very specialized for different tasks, and I think that can be engineered. But of course, that's rather different from AI, that is designing something to do a mechanical task, like wheelchairs, automated wheelchairs and things like that. They can be quite sophisticated. But I don't think the android type of robot that looks like humans would be the answer there. Partly because our own bodies are very efficient in what we do and our fingers are very flexible, and the ratio of sensing modules and muscles and relative force and so on, we can't get anywhere near that with robot hands at the moment. And I'm not sure we can, because it may be that biology is necessary to get it packed to that density.

Right. I occasionally see reports of development of some sort of artificial skin and various sensors, but they're nowhere close to the density of what we have in a piece of human skin. Seems, to coin a word, we'd have to *grow* that to be able to make it properly. But I'm just wondering to what extent we need that to be able to do the average real-world manipulation. I think I forget who said it, but one researcher said, "nothing stumps a robot like a bag of oranges." It's not obvious how you should pick it up, it's going to shift around in unpredictable ways, and if you're trying to grab it, you run the risk of creating orange juice. What do you think in your approach, what's the answer to that problem?

Well, I think you could make quite a lot of progress with a learning system which would crushed the orange, of course, but learning through experience is the way to do it really. As, we do you know, I would say that, when it's skilled enough, you'll find it's able to do it. But of course, that is a big problem, the bin-picking problem of getting sorted objects in a bin and trying to pick them up and I think Amazon is working on that quite hard. Because that will save a lot of people in their warehouse.

Right, and just to be clear here that that's the bin packing problem is when putting things into a bag. This researcher was talking about the problem where a bag like a netting bag with oranges in it that you're picking up and it's got half a dozen oranges in it, and they're going to move on you, and how do you pick it up? It doesn't have a handle, and how do you even figure out where to grab it?

Yeah, well, that depends on your experience of those physical objects. So you have to experience liquids as things that that change shape, things like wire and netting, and all those different things.

So, in your development then, of robots that are going to learn that, do you have or are you building a lab where robots are indulging in and experiencing all kinds of interactions with the real world?

Well that's the idea yes, I mean, we haven't got that far, to such a variety of things. But we do have different objects, objects that move, objects that are stationary objects, objects like buttons you can press, and lights come on and things like that. Yes, but we have to go further on that road to get there. I mean, other labs have done it through repeated trials where they set20 robots going overnight for several days, trying different things out and recording the successful ones. But it's not terribly intelligent, it's a data recording process really.

Well, and it would be good if there was a way for the labs that are doing this, to not have to duplicate each other's work so that once one learned how to walk, that they could share that model with everyone else, and then the others could work on making coffee or something.

Yes. But they'd have to be the same sort of robot.

Where in this development pattern does the idea of values come in? Because as humans are developing this understanding of the world, we're also developing values: what's important to us. We get a lot of this our parents and then from other people in the environment, teaching us don't do that, don't hurt people, be respectful to your elders, don't hit your little sister. All of which is essential to safe conduct in the real world. How do you give a robot that?

Well, I think that there's two things here really, one is the sort of ethics of the robot. I think at some point, you'd have to enter into the system some rules that stop it from doing certain dangerous things by accident. The other thing is, I think the robots who had gone through the developmental process wouldn't have goals of their own; they would imitate humans, and you could train them to do things that you want them to do. But they would be platonic and their demeanor, they would they wouldn't have strong passions or wants or intense. Despite what the

media often portrays; I think that's a biological thing. Because we live in a framework where we have some very essential things, we must have shelter, we must have food, and we get born and we die, and those facts of life don't apply to robots at all. So, in terms of emotions, I don't see that robots should have strong emotions. I don't see any reason for that. Because they don't have a concern.

Right; I'm thinking about -- well, in a grand scale, it's the value alignment problem here -- but robots can play around with all kinds of things in the physical world to learn physics and they will get an understanding of the physics of everything from fish to boards of wood. But if I've got a practical robot and to just come up with a trite example, I say, "go down to the store and get the newspaper for me," I don't want to have to specify every step that it's going to have to take and every condition that it might have to deal with along the way, I want it to figure that out. Somewhere between my house and the store, it could encounter an unpredictable situation that requires the exercise of ethics. It sees someone that's hurt, or it has to avoid hitting someone; all kinds of situations that that we can imagine that we deal with automatically that require the exercise of values, which are things that we learned from our parents at the same time we were learning how to run up and down stairs. So, where does that training part of the process come in?

I think most cases where accidents occur, the robot would be observing them and not necessarily being involved in them.

But that might still be something that it should know to do something about; you see someone bleeding to death on the side of the road, you don't just keep going because your goal is to go to the store and get a newspaper.

If it recognized that, it would know that bleeding is bad for humans, it we know a lot about human emotions because it would work in human states. So, this is a distressed human here, it might not be bleeding even, and it would say the human needs help and it would phone up for the necessary ambulance or whatever is required. But I don't see it intervening in itself.

I mean calling for help is intervening; it's taking an action.

But what I mean is it wouldn't have physical, the strength or the abilities designed for all the circumstances that might occur. If it's just a conversation robot wthat anders around people, you probably won't be strong, you know?

Well if we looked further ahead, not that we're necessarily building robots this physically capable just to go to the store and get a newspaper, but the robots might be equipped by default with first aid kits, just because they're out in the world and they might encounter situations where they're needed.

Yes, well, they whether they could use the first aid kit themselves is the question. They could certainly carry one and give it to people who needed it.

Wouldn't it be as easy to train a robot to apply a bandage or a splint as it would be for it to learn how to play cricket?

Yes, it would. I mean, this is up to what you want to do with these robots, if you want to go around as first aiders, yes they could do that.

Or both, I guess it just question of how much training you want to give it.

It is, and you'd probably need some sort of training school or something, you'd have to decide how much training these robots need before you release them for any useful thing to the public and even then, the user might want certain functions, out of the robot and they should have to be trained for that.

I see. So, in the developmental path you're pursuing towards conversational robots. Do you see them being released? And in what way?

Yeah, I could see them wandering around, say, in a park, or in a university campus, and sitting on seats, and people come up to them and talk to them say, "What's your name?" and they'd have names. They wouldn't have gender, because they don't have gender, but they'd have, they probably have male or female names because their speech synthesizers would match that. And you could talk to them and tell them about yourself. And then another day, you could meet the same robot in the park and carry on the conversation because it would record the history. And it would say, how are you feeling today, you weren't that well, yesterday? And it could advise you that paracetamol is useful for headaches or something. You could talk to it like a person and we're talking about care homes: a lot of the human interest is having conversations with people. They like to have conversations and the problem with current computer systems is they're very poor at conversation.

Right, and when you're describing your work to the public, there are certain tropes that the media loves to attach to any mention of the word *robotics*. Most of those articles have a picture of the *Terminator* next to them and the ones that that are more thoughtful, will start talking about Asimov's Laws of Robotics. To what extent do you find yourself pushing back against unhelpful or obstructive narratives to help people understand what you're doing? Or is, for that matter, Asimov's work at all useful?

Well, I don't find the predictions very useful, no. A lot of these seem to be produced by physicists and people like that, not people who are actually programming the systems. And the software people I find are generally much more cautious. You know, even the AI people, the top AI people, are much more cautious about it. It seems to be a strange thing. A few years back, there was big headlines about they would take over the world you know, there would be machines that would connect together, decide that they were all more important than humans, and take all the energy resources and so on and all the scenarios. I don't think those are going to happen. When you look at it in detail, we would go, we would intervene, we were not going to let it happen. So, it's a bit unfortunate that and you know, self-driving cars are talked about all the time. But getting to level five in self-driving cars with no steering wheel, means that the car really has to have something like human intelligence, because it can be stopped by the police or

something; say the police wants to talk to the car and tell it that there's been a problem, and it's got to go on a different route. You know, I mean, all sorts of cases like that, and the problem with self-driving cars is the same as all the other problems really, that it works very well for 90% of the time. But the long tails on the distribution, the odd cases, the kind of problems that come up, each one of those tends to be unique. So, you cannot group those together, you can't classify these togetherand solve them each time. Each one is a unique one and there are enough of them to be a problem, even though individually, they're very rare and that's the same language things, it works well, most of the time. But there are these odd cases and the errors that come up are horrendous with some of the systems that are being used.

90% isn't good enough, we need something that's 99.9%, which actually makes me think of the Uncanny Valley, and the anthropomorphic types of robots, and it's very easy to build something that people relate to empathetically. Pretty much just putting eyes on it, that don't even have to move, puts it in the realm of "teddy bear" and now we start having a reaction to it. And so when we start talking about robots that have any kind of -- not even necessarily humanoid form, but mammalian form -- you can create this empathetic reaction, this emotional reaction, and that's been done with lots of toy robots like the Jibo and the Pleo. And Kate Darling explores this in her TED talk about the Pleo robot, which is very simple compared to the things we've been talking about. And so when you're talking about having a conversation with a robot, to what extent do you think about the way that you might want to anthropomorphize that robot or give it these kinds of qualities, so that we *want* to have a conversation with it, so that we feel like that's useful, or that there's something on the other end that's worth talking to?

I think that's right, I think we do need to make the thing attractive to people and there are lots of robots that aren't attractive, and people don't like.

Boston Dynamics.

Quite scary aren't they.

Yes.

We use the icub, which is based on a three-year-old child and that sort of non-threatening size is important. But also, there's other biological data; like we tend to blink our eyes at the end of sentences, or double blink sometimes when we're emphasizing things, and so on, and you can make the robot copy these patterns of behavior. And gesture, of course, is very useful. So, some of those techniques can be used to make the whole thing more friendly, or human friendly, but quite a narrow path, I think to get that right, and it's easy to make a mistake and people will reject it quite easily then.

Well, this is fascinating, obviously could talk for hours about this; we don't *have* hours to keep going. So, I'm going to ask you to help wrap things up here and tell us how people can find out more about what you're doing, follow your work, and get your book.

Okay, well, the book, I should say, includes nearly all of what we've talked about, including some speculation about the future, and what social robots might be like, and how they would talk, and so on and that's based on our work over the last 10-15 years in which we've been doing experiments on infant development. And that work is continuing in a community of people called developmental robotics, or epigenetic robotics. And there are conferences and series on that, and a few books around, if you really want to get into the technical stuff. It is growing and you'll find that people are getting more interested in the developmental approach. Even in bigger systems, they don't do the whole thing like I'm talking about, but they will introduce the developmental idea. So I think it is getting more traction at the moment. But there's an enormous amount of work to do and so we need more people to start working on projects in it.

And if people who would be useful to you are listening to this podcast, what would you say to help them decide whether they should get in touch and offer their abilities?

Well, read some of the stuff, see if you like it, have a look at the videos and see if you get any resonances with it.

Great. So, thank you. If you want to build a conversational robot, now you know where to go. Iif you want to get involved in that project then Mark's the person to contact. So Mark Lee, thank you very much for coming on the podcast again, The book is *How to Grow a Robot*. Mark Lee, thank you for coming on AI and you.

Thank you. It's been a pleasure.

That's the end of the interview. That really expanded my view of how robots might evolve. You know, I'd heard about the grow-it-from-a-baby approach from documentaries on, like, Soul Machines and Baby X, but it's another thing to have a dialogue with someone who's doing that themselves. And what a great field to be working in, I mean how exciting, and how wonderful to be at a time when there's money available to work on that.

In today's news ripped from the headlines about AI, researchers at the University of California at San Francisco developed AI that could turn a paralyzed person's brain waves into text. Now, they were only selecting from a vocabulary of 50 words that they had trained the AI on, but that was enough to form useful sentences. The patient, codenamed BRAVO1, lost his ability to speak in 2003 after a severe car accident that left him paralyzed. They trained the network by asking the patient to think of, or attempt to say, a number of short sentences constructed from the 50 words, which gave them patterns that the AI formed a network on and then was able to decode at the rate of 18 words a minute when the patient thought of new sentences. The research, sponsored by Facebook among others, demonstrates, according to researcher David Moses, "the potential for this approach to give a voice to people with severe paralysis and speech loss."

Tesla had their annual AI Day a couple of weeks ago, and we got to see their progress with their android robot, called Optimus, and what they call full self driving beta. I've had the FSD Beta for a couple of months now, I had to earn it by being a very good driver, no sudden stops or turns, I was on my best behavior for about three months and then it showed up one day. It's simultaneously impressive and underwhelming. Impressive for what it can do, like drive on roads that don't have any lines at all, and underwhelming for how much it still can't, like make a turn quickly enough to not tick off the car behind.

So the AI Day, which you can find on YouTube, had a lot of data – really, really, technical data – about their progress with FSD, and also had a live demo of their current Optimus design, where the robot took a few steps on stage and waved. They didn't want to do more in a press conference because if it fell over that would be the only thing people would remember, but they showed videos of it doing other things like moving packages around and watering the office plants. What was notable to me was that they didn't say what you might use it for. Presumably watering the office ficus isn't a killer app. And that's really the big question. They might be able to sell it for $25k like they say, but to do what? Now it does look sleeker and infinitely less threatening than the Boston Dynamics Atlas, which is certainly far ahead, and if Boston Dynamics is looking for some press, they'll issue a wrestling challenge, because Atlas would certainly rip Optimus in half. What we've heard from Mark in this interview helps me put these robots in a better perspective.

Next week I'll be talking with Mark van Rijmenam, a digital futurist whose new book is Step into the Metaverse: How the Immersive Internet Will Unlock a Trillion-Dollar Social Economy, and he'll be helping us understand what the metaverse is, why it's important, and how it uses AI.

That's next week on *AI and You*.

Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

http://aiandyou.net