RAW COPY

AI FOR GOOD GLOBAL SUMMIT
STORYTELLERS
MAY 15, 2018

Services Provided By:
Caption First, Inc.
P.O. Box 3066
Monument, CO 80132
1-877-825-5234
+001-719-482-9835
www.captionfirst.com

>> Our next session is ran by Stephen Ibaraki. Stephen Ibaraki, I'll hand it over trait over to you.

>> STEPHEN IBARAKI: Thank you.

It gives me real pleasure to be here. It thank the Excellencies, AI for good global Summit, distinguished participants, all of our Delegates, I'm Stephen Ibaraki, the moderator of this keynote session with, sir, Roger Penrose. The topic is why algorithmic systems possess no understanding, it is controversial and a stimulating discussion. The keynote will be followed by a dialogue session with questions and answers between myself, the audience, also with Sir, Roger Penrose. Let's provide a brief introduction. He's an English mathematical fist assist, mathematician for loss of science, he's a professor of mathematics at the University of Oxford and a fellow. The professor is known for his work in mathematical physics and particularly his contributions to general relativity, and he's received numerous prizes for his decades of work, including the 1988 wolf prize for physics, which he shared with Stephen Hawkings for the singularly theoroms.

I'll now ask you, professor, to take the stage, the podium, you can deliver the keynote.

>> ROGER PENROSE: I feel honored to present my point of view here to such a distinguished audience. It certainly is the case that Artificial Intelligence has advanced tremendously in the last years, and a question has been raised very often quite since the beginning of the subject which is whether and when -- if whether the subject of Artificial Intelligence will reach the level of human intel with against. Now, I think it is very likely that it will reach the level of one particular part of the brain, namely the cerebelum, I make a distinction -- there are two parts of the brain in particular, one, the part where people usually refer to up here, and the cerebelum in the back, it is a bit underneath, the two, they have a comparable number of neurons each and the ce rebelum is have more connections than the cerebrum does, you may think that the cerebelum has more could you terrible power than the cerebrum and that's probably true. It is possible with the advance of Artificial Intelligence that one might reach the level of cerebelum activity.

I like to think of the relationship between the two, that's to say the cerebelum and cerebrum is a bit like the programmer and the action of the programme because -- you're driving a car, you're learning to he are verse, something like that, initially you have to work out, you know, if I move my arm in this particular way, so on, then the car will do such and such and you think through in a conscience way. After a while you don't have to think about it consciously. The cerebelum takes over. We know that when people play the piano, they play extraordinarily fast and accurately, they're not thinking about how they move each finger, that's all done unconsciously by the cerebelum likewise, a tennis player, the skills of the tennis player, they would not work out which muscle it should move at each stage. All of those things are controlled unconsciously and the unconscious actions are the kind of thing that one can well believe may well be something that you could imitate by some Artificial Intelligence system. When I say Artificial Intelligence, I mean, computer controlled in the sense that we now understand that term.

Why do I say there may be something different about the cerebrum. It really dated back to the early 1950s when I started doing research in cambridge on a completely different subject. It is a mathematical subject, algebra, geometry, didn't have anything to do with AI and such. I was very intrigued -- I have heard about the theorem that said apparently, so I thought, that there are things in mathematics that you just can't prove. I didn't like that idea. I went to this course on mathematical logic given by a man called stein, a very good course and I learned about touring machines and the theorem and he taught me -- and others in the class -- about this theorem and it didn't say what I was afraid of, that there are things you can't prove. What it said, if you have a system of proof procedures and these proof procedures are such that you could put them on a computer to check whether they have been accurately carried through then these procedures end up by saying you feed in the theorem and at the end it says yes, proved, no,ish no, it is wrong, or don't know. Don't know, it doesn't say don't know, it just goes on forever. The thing is, what the theorem is, if you believe that the system never gives you the wrong answer, if you trust it, reason to believe it is right, you must believe a particular statement constructed from the rules must be true, yet unobtainable by means of the rules. I found this absolutely stunning when I heard that. It is not that you can't prove this thing, it is just that you have to use procedures which transcends the ones you're using. The key thing is, it is your belief that these procedures actually work.

If you're prepared to trust procedures when it says yes it is true, that it really is true, if you have enough belief in that, then you also believe the statement which goes beyond them. You somehow can transcend the rules. I thought it was amazing.

It sort of tells you that your understanding of why the rules work is more than using the rules. That's really what it does say.

What does that mean? Understanding, well, you look at the rules, you say yeah, I think that's okay, that one, I'm -- that's okay too. You go through that kind of reasoning, if you're convinced it is okay. Then not only do you trust using those rules but you can trust the procedure for jumping beyond the rules. This struck me as amazing. It is understanding transcend something the rules. This made me think our whatever it is that makes us understand things, and that seems to require a conscience perception of things, that's something that is not governed by rules. You're not quite so clear on that. It is rules that you know and can appreciate. Maybe there are rules in our head that are so complicated we can't know what they are, we don't know what they are. Something like that. I don't think that's the case. That troubled me and it troubles others who complain about what I write about. The thing is, you have to ask the question, how did we come about with our understanding of mathematics? We have a cartoon which I can't really show you here which shows our ancient ancestors doing wonderful things, building mammoth traps, building houses, domestic ating animals and there is a mathematician trying to prove something and a Tiger is ready to pounce on him. It is really to show that there is no particular selective advantage in being a mathematician. In fact, I think it is a disadvantage actually. Okay. General understanding, yes, that could easily be selected, it was, by natural selection. I'm okay with this. This can be applied in different areas. There is a nice example that I should show you but I won't show you here, you will be grateful for that. It is a theorem that's nice to give if you have the time, it can be appreciated by people that don't know much mathematics. The thing about this theorem, you start with any number, when I say a number, I mean an actual number which means 0, 1, 2, 3, 4, whole numbers that are not negative.

If you take, we know about the procedure to prove a thing for -- one of the amazing things about understanding, it is that we can understand the infinite number of things. People say you can't understand infinity, can you read? Yes, you can. That's a thing you can do, a human being can do. A computer can't. That's really the point.

An example of how a human being may do this, we learn it at school. Suppose you have a proposition that depends on the number N. All you have to prove is two things. First of all, it is true for 0. Secondly, if it is true for N it is true for N plus 1, that's standard mathematical induction. It is true for all. The thing about this theorem, it can't prove it using that, it is a -- yet you can understand it. I'm not going to describe it. I'll give you a rough idea.

You are given a number, any natural number, and you apply to this two procedures, procedure A, and that is a way of getting this number, making it bigger. Procedure B, subtract more. A makes it bigger, B, subtract more, making it bigger, it makes it hugely bigger, absolutely hugely bigger. What's not obvious, it is that it is a tortoise and the hare, B always wins but it does it in so many steps you never would go through it. If you start with number 4, it goes through so many steps that no computer that's ever been conceived of could ever go through all of those steps. With a pencil and paper, you could see that by following the steps, oh, yeah, yeah, that's going to -- oh, I see it will come down.

What kind of reasoning tells you that? It is our understanding.

I think this shows that whatever understanding is, it is not something that you can incapsulate. I came to that view after hearing the talk. It was a bit worrying, what do we do? What is consciousness, what's understanding? Is it a magical thing that comes in our heads at some point and something like that? I don't like that idea at all. It seems to me that it must be something in the laws of physics because our brains are governed by the same physical laws as everything else. I don't think there is something else going on there. What are the laws of physics? This I took advantage of going to other courses one by one, you know, on quantum mechanics and I thought, well, you could put a thing about general activity on a computer, now we have the good Examples of this because people probably here have heard of the Ligo detection of black holes barreling into each other some galaxy millions and millions of light years away and waves come along and a specific signal that you see from the waves, it can be seen in the gravitational detector. We have to pick this out of noise and stuff. How do we know that indicator is two black holes spiraling to one another, an extraordinary amount of computation. It is high powered -- I don't know about AI but direct computation of this particular problem. Really powerful stuff.

This shows the accuracy of this kind of comp takings. General relativity is something that you can really put on a computer and do it very, very precisely. How about quantum mechanics. That's a big evolution of the physics, and quantum mechanics, yes, you can put that in the equation, there are problems with it, you can put that in the computer, you can imagine how to stimulate the brain, whatever, put it on the computer and it chugs away and it solves the equation. Then there is a catch. Because this equation, it was clearly pointed out, it does not tell you what happened in the world. You probably heard about this, this idea was introduced to really show the absurdity of his own equation as applied to rather extreme situations. This is something that you put in a state that according to this equation, it is dead and alive at the same time. Look, this is ridiculous, he didn't quite say it like that, it was ridiculous, there must be something wrong with the equation.

He really did think there was something missing. Einstein did, and others did, even -- he wasn't quite clean on making the views clear. We thought the same kind of thing. The first lecture I went through, he illustrated this thing called the super position principle, which is used in this dead and alive at the same time, you see, you have in the quantum mechanics, the particle is here and here at the same time. That's a great thing to get your mind around. The theory works. The thing is, that's not the whole story because this equation says it doesn't allow it at the same time. That's not what you see. You see either dead or alive.

Something happens -- you have changed something down there -- excuse me.

Something happens that is not part of the equation. That's what you call making a multilinear and making the measurement in the device which is probably the device just like everything else, what's the difference. Any way, there was something happening which doesn't follow this equation. It makes a choice. The universe makes the choice between that and that.

You can make a criteria so combining the theories that I was talking about to say if you move so much mass into a super position of two locations how long does that live. You can work out a lifetime if you believe something to do with quantum mechanics and generality together which is not part of the quantum mechanics, it does not say that one thing is part of the other and you wheel in another procedure and it is inconsistent. You get used to the idea when you use the quantum mechanics.

What is the other procedure? Nobody knows. It is the collapse of the way function or the reduction of the state.

For me, that was the gap. That's the one thing that you could not put on the computer.

In fact, I liked to call it this, I collaborated later and said I couldn't see how neurons could make use of this thing, but I was told about these things called micro tubials that inhabit all cells in the body, particularly neurons and these things which would shield information away from the outside world and use quantum mechanics in an essential way. I should make the point, people often say quantum mechanics is irrelevant to the brain action or something, of course it can't be true because chemistry is crucially quantum mechanics, you don't mean chemistry but something else, something beyond chemistry.

There is lots of that. We see that in quantum mechanics, they go beyond the ordinary procedures of chemistry. In the brain, it could be easily something like that, particularly in the micro tubials and I think that's likely correct, these are tiny tubes, nano scale and according to a colleague whose day job was to put people asleep reversibly as an anesthesiology and he's interested in what he's doing, his view is that the general anesthetics affect the micro tubials and it sends you to sleep in a reversible way. The thing is, the micro tubialses would be making this choice, and what's the choice? The choice is what we don't know about in nature. Nature just says it is random.

I don't think it is probably just random but something extremely subtle, something beyond comp takings, which is my view, the view would be that you can associate with every action of this state reduction process, collapse the process with an element of what we call pro toe consciousness. When pro toe consciousness, it is the building block out of which again went consciousness is built. We have to understand that part. Nature has made choices and they don't mean anything. In the brain, these choices do mean something. That is when we maybe make a conscience choice, it is a very interesting idea. A lot of people are skeptical of this kind of idea. It seems to me do we need somewhere that you take advantage of this part of the physical world as we understand physics, which goes beyond computational simulation and this is the place it has to lie on and that's the point of view we have taken and there are experiments that can shed good light on that.

Thank you very much.

>> ROGER PENROSE:

>> STEPHEN IBARAKI: I want to get a time check, we're ahead of time by the older schedule. Is there an updated schedule? I believe the tea time is 10:45? We're fine? Good.

This is the question and answer portion of our dialogue with the notable iconic legendary professor Roger Penrose, we're fortunate to have a man of his stature scientist of his statue across so many domains. I'll ask you for a bit of a favor, how many of you agree with my favor, show of hands? What! Come on! I just want to see who will agree with my favor. It is a simple favor.

They want the favor first. The favor is that we have a golden opportunity here with Sir Roger Penrose to make history, to make history to add to the history of the AI for good Summit. I want all of you to do a selfie later on during the session, not the end, I want you to stand up, face away from the stage, professor Roger Penrose will be here, I'll put a thumbs up, I'll be behind him, beside him, a thumbs up and I want you to capture a selfie of yourself and professor Roger Penrose and I want you to put it on social media. Let's make sure we have an impact on the world today courtesy of Dr. Roger Penrose.

I want you to check the time. Now let's see if I can get a show of hands on that favor. Will you do that favor for me and for AI for good and Dr. Roger Penrose? Better. I still don't see all of the hands up! Come on! That's good.

Now we'll get to the question and answer portion. As you know, we have this app where you can submit questions and I have some myself. I'll start the Q&A portion. I'll sit down.

Professor, you know, a very interesting kind of research you have done especially the latest work where you are indicating you have understanding and understanding -- the professor says it cannot be algorithmically determineddisticly done on a computer. It is something that transcends the computer. It is an interesting idea. As you heard the professor indicate, something more is going on, there is a quantum affect with the micro tubialses which is a structure in the brain that transcends what computers do. Using that as a context, the first question is, professor, what's right and wrong with the term Artificial Intelligence?

>> ROGER PENROSE: When I heard that term, I was nervous about the word intelligence. To me, there are three words one can consider here.

One, intelligence, one is understanding, one is -- let's see. Yeah. Awareness. Awareness.

Is this switched on or -- can you hear me.

Okay.

The thing is, intelligence, these are words that I wouldn't like to define any of them. It does seem to me that there are implications between them. The word intelligence seems to me is something which requires understanding. That's to say if you have a device which you would call intelligent and it has no understanding, that's a misnomer. Whatever understanding is, it is an ingredient of intelligence. What about intelligence and awareness? Again, it seems to me inappropriate to say of a device, that it -- that it possesses understanding if there is no awareness. The usual usage of the words, it seems to me that we're taking the two together, intelligence requires consciousness. It causes awareness as to say.

It doesn't say AI isn't -- the devices are not intelligent. It does say that if they're not aware, they're not really intelligent in the normal use of the world. I didn't like the term intelligence. It seems to imply also the thing understands something. I think it is more likely the cerebellum, a device can achieve amazing things and the things it can achieve and I certainly agree with that, a great advantage, yes, tremendous. But they don't seem to know what they're doing. We, the person whose done the programming may know what they're doing. You can also point to systems like -- what do they call them -- the neural networks, where you set the thing so it improves the performance and you don't have the remotest on what it really is doing. The human being may not understand either. You understand the again principles under which it improves, so on. The understanding, it is something that we seem to present -- possess, present to the system as a whole. I'm just saying, never quite liked the word Artificial Intelligence, artificial cleverness would be okay. Artificial Intelligence implies it has qualities that intelligence has and I don't think artificial AI systems, I call it that, they actual do possess that. Of course, other people will probably disagree with me here. You just asked me my opinion. Yes.

>> STEPHEN IBARAKI: Maybe we shouldn't call it AI, right? We call it --

>> ROGER PENROSE: AC.

>> STEPHEN IBARAKI: Artificially clever. Maybe that's a bit of a conflict with AI researchers.

There's a question that's being posted by the audience, how do we incentivize young researchers to get involved towards programs like AI for good and focusing on these kinds of projects that are really motivated, motivating in many respects.

How do we incentivize --

>> ROGER PENROSE: The question is to me?

>> STEPHEN IBARAKI: Getting young people to focus on AI.

>> ROGER PENROSE: There is excitement in the subject anyway. I don't suppose that the excitement comes from people who want to think that they're simulating the cerebellum. I think it comes from being able to construct devices, it comes from being able to construct devices which do amazing things. These amazing things, sure, there are lots of them, robots, which walk around, so on. Self-driving cars, what have you.

There is a huge amount of excitement in developing these devices. Maybe there is the hope that some people would have to try to build the device which could actually have conciousness. This raises a you will sorts of problems, ethical problems, goodness knows what. It is just as well that we don't have such devices yet. If they were conscience, you would have to worry about turning your computer off, you have to ask it, is is it all right for me to turn you off, you would have to say that each time. There is something that really has researched that. Do you really mean to. Yes.

I think -- it would raise questions that are nice to talk about. If you really believed it, it would be a really blockage to all sorts of developments in AI and we wouldn't want to send out something that destroys it if you really thought it was conscience. Sorry, I'm straying away from your question which is how do we excite young people. I would have thought there is an awful lot in AI, in what AI does that doesn't involve really being conscience or anything like that which is terribly exciting for young people. Whether it is games, whether it is trying to build robots which do wonderful things, so on. All I can say, I don't see any obstruction to that with the point of view that AI as we now know it, which is not involved, some new thing about physics, it is computers as we understand them, even whether they're quantum computers or classical ones, they still involve something beyond what we understand in physics as of yet.

Another question, professor, your colleague who you spent so many years working in collaborating with Stephen hawkings, what would he think about the current ideas on quantum affects causing consciousness in human begins?

>> ROGER PENROSE: We collaborated together on a project that you mentioned about singularlies and black holes and all of that, but after a while we started to diverge in our views. I

think it was really that he -- although he initially when he put forward the idea of black hole evaporation it seemed had there was information lost and he put forth the idea that there should be information lost in black hole evaporation which I agree with. We agreed originally and he had his ideas on modifying the quantum mechanics and he changed his mind and said no, no, there has to be no change -- there can't be loss of information. You can't do this with quantum mechanics, to me, he was being too reactionary at that point. we departed. Remained friends, of course. We departed company as regards to what we believed in. I'm sure he wouldn't like my ideas on AI because they suggest that you do need to go beyond standard quantum mechanics and say that standard quantum mechanics, they have no explanation for the reduction of the state.

He would say, well, maybe all of the different alternatives that exist forever, there are many worlds, all coexisting, so on. I don't go along with that view. We would have disagreed on that. You probably would have disagreed with me on my view about AI. I'm pretty sure he would. Yes.

>> STEPHEN IBARAKI: Next question from the audience, if we can get the questions up, again, this is for you, professor, right now machine learning is great at spotting patterns but it doesn't truly understand data. How do we transition to learning reasoning? That's the question. Machine learning, great at patterns but not truly understanding data. How do we transition to learning reasoning? That's for you.

>> ROGER PENROSE: I presumed it was the audience. Yes.

How do we -- that's good. The trouble is, I'm saying that it is not going to understand. You have to put your own understanding in it. The thing is, even -- this is striking, you -- you get the point, you see this is beyond the system we have had before. Then you say, okay, why don't we put that into the system? That's perfectly good. You go another step. You use your understanding to go to the next step and put that understanding into the system p you could keep doing that and although you say it never actually had the understanding, you can use the human understanding to improve the algorithmic system you had before and you can keep on doing that.

Maybe you could do that enough. You just -- the human provides the understanding and then if you're clever enough, you can see how that particular element of understanding can be translated into an algorithmic calculation. That's fine. I have nothing against that. I can see -- it could be hard in a particular case. Very often -- I made up this chess position, which you can use a bit of understanding, and you feed this to the chess programme Fritz which is the grand master level and it makes a stupid mistake and loses the game. You see why it loses the game, you can -- I expected it to lose the game, but the thing is, you then see what is the understanding that I was using that it didn't have and then you see how the programme is with the thing and it doesn't make that mistake again. I think you can see this in a positive light. Although it won't itself have the quality of understanding, if you can see clearly enough what this ingredient is in any particular case, and that particular instance of your understanding, you could make into algorithmic system and it goes beyond what you had before. It seems to me, as an optimistic statement. You are useful at some point. human insights, so on, they have their relevance. When you can see how to incorporate that in a way which could enhance the could you terrible system, sure, that's fine.

-- enhance the computational system, that's fine.

>> STEPHEN IBARAKI: You have this work, quantum processes occurring in the human brains which gives us understanding which computers are incapable of, and that's what separates human understanding from algorithmic determine fistic processes that standard computers have. There is a gap. But at a base law of physics, what you are a talking about, the new ideas, interesting ideas about quantum processes. Could a computer ever do these kinds of quantum processes that you talk about that are unique in the human brain.

>> This is what we talk about computer, careful to say as we understand them today, a computer -- by the nature of the word, we say computer, that means it computes. It does what we understand in the ordinary sense. We have a good notion of that, it is a machine, what it can do, that's a had mod were computer, I should qualify that a bit. It means a had modern computer which never actually makes mistakes which is pretty close to what the truth is with them which never makes mistakes and which has an unlimited storage. That's not true. You can -- the storage may run on a certain point. the idea is that at that point you bring in more. So on.

It has a an unlimited storage which is not totally realistic.

It is, some of these problems mentioned, you would exceed the storage of any conceivable -- conceived I should actually say computer. So even though that is a deterministic thing that a human being can look at and see -- if you just plug that in the computer, it never would get there. Then, of course, you say, look, I have this understanding, I put that in, then it gets there. The point I'm raising here. We're talking about computers as we know them now. Now of course, I'm not saying that whatever is going on in the human brain is outside what you can build in the lab. There could be some day, some latter day Frankenstein constructing some device -- I'm not sure I like the idea -- some device that incorporates what I'm claiming is going on in this going beyond quantum mechanics as we understand it today. The elements would be brought into the action of this device, and then, of course, you have to worry about it, you have to -- you have to know what pain means, you will say if I do something to it, is it going to be painful to it. I'm glad we're not there. It raises ethical problems which we're not ready to face. Fortunately, computers as we now know them don't have consciousness and you can do what you like with them, you can kick them, throw them in the bin, turn them off, swear at them, all of these things are fine. It doesn't feel a thing.

>> STEPHEN IBARAKI: Interesting idea, in fact, there is a question from the audience which is about the late and great Stephen hawkings was worried about our existence because of AI, and that ties into what you were saying, if something happens, maybe you want to comment on that question, the existence of AI, and worried about that aspect.

>> ROGER PENROSE: In my view, it would become a worry if we made these beyond computer devices which took advantage of this element of the physical world which I'm claiming is what consciousness is built from, if you like. The choices that the universe makes in the quantum system, it does this, does this, does that.

Nobody says you should make it, decide, what you should make it, as you like.

It is this decision that the universe makes, it is that it is somehow organized in a way in our brains which we make use of or which is us in a certain sense, but you could in principle according to this view have something in the lab which is not biological, with I is not human, which is not animal and which could perhaps take advantage of it.

Fortunately, I would say this is way in the future whether the human race survives enough to get there is a question I would say.

>> STEPHEN IBARAKI: In your 9th decade of contribution, amazing career in mathematics, relatively, philosophy, is there any -- this is the last question and then we'll do the selfie. Is there one final challenge or book or some quest that you still want to do over already an amazing career?

>> ROGER PENROSE: Oh, yes! There is something in -- that's exciting too, but that's not got so much to do with this discussion. All right. What else. We have both of you writing, I was supposed to write this book for at least a decade. A book which takes the works -- amazing works, and develops various themes of mathmatic, I think of themes that could be explained in simple terms using these explanations and I really want to get down to this Bock but I have been

distracted by too many other things.

>> ROGER PENROSE: This is not a distraction, to share with leaders of the world, really on this idea of a common good, we're now going to execute on my favor. As many of you can, I want you to stand up, turn around, hold the camera up, we'll get professor hawkings -- professor hawkings?

>> I'll put. Mic down. You won't get to hear me.

>> I stand up too I guess. I'll grab the --

>> Okay.

>> Are you able to do this.

We have the thumbs up. Remember, you're going to socialize this through social media.

Quite honored. Thank you. Thumbs up. That's great. Not yet. Wait until we're finished. .

You can all say that now you have done a selfie with Sir Roger Penrose. How many people can say that! A thunderous applause for the professor. Thank you. I guess there's a break coming up.

That was fun by the way. Thank you for agreeing to my favor.

>> We have a coffee break now and we have the first demo outside. We have a break.