What AI Machines Won't Be Able to Achieve

https://mindmatters.ai/podcast/ep218

Austin Egbert:

Selmer Bringsjord, a professor at Rensselaer, is a pioneer in artificial intelligence. On this episode of Mind Matters News, Dr. Bringsjord answers questions on topics ranging from quantum computing to his celebrated Lovelace test. For the ability of AI to be creative, he begins by introducing himself. Let's listen in.

Selmer Bringsjord:

I'm Selmer Bringsjord, director of the Rensselaer AI and Reasoning Laboratory. I have a few departmental affiliations, but basically I think that's my main job. Abbreviated it's the rare lab at RPI and it focuses on the intersection of AI and essentially logic to capture how reasoning works. And I've pretty much been doing this kind of thing, not running a lab, but this kind of logic and AI thing as far as I can tell since my senior year in high school when I fell in love with logic and the mind and computation. So that's what I do. And the lab is basically run on the brains of some brilliant graduate students and sometimes undergraduates. So I spend a lot of time figuring out how to work with them and with a number of other colleagues in the lab.

Announcer:

What is artificial intelligence and where do we find it?

Selmer Bringsjord:

AI can be defined by course turning predictably to the textbooks for the field. Unfortunately, that really does not give us an accurate definition of what AI is. I think a better strategy is to simply say, look, there's shallow AI and then there's real AI. Shallow AI tends to dominate and increasingly in the form of machine learning gets a lot of attention. I would characterize shallow AI as okay, any function, any input, output, transition that has a smidgen of intelligence, whether at the human level, the animal level, even some cases the insect level counts as shallow AI. Real AI is by definition human level AI and above, and there the definition operationally is pretty straightforward. You want to try to create an artifact that can emulate, match, maybe even exceed what humans are doing when they're exhibiting their rather remarkable intellectual skills.

Announcer:

If AI can beat us at chess, is it matching human intelligence?

Selmer Bringsjord:

When you define AI roughly in the way I've done with this rather arguably offensive division, our question arises as to what you do with something like chess, which of course is human level by definition, no canine mine can do well at chess, but that's still not matching and let alone exceeding human intelligence for a rather obvious reason. And that is that the game is completely solvable. There's a fixed algorithm for it, and most of what we praise humans for when they operate in intelligent fashion does not correspond to something that has a known completely efficacious algorithm. So anything in the area of creativity, for example, is out the window in comparison. So maybe if you're a true believer in

this strong AI sometimes called AGI, now artificial general intelligence, then you think humans will be matched by machines even in areas well beyond chess.

Announcer:

What are the heady predictions of future AI?

Selmer Bringsjord:

There are predictions for AI exceeding human intelligence based on the rather inevitable increase in raw computing power in a hardware sense, and this has been with us pretty much since we started thinking about these things. That is whether AI could match human intelligence Turing, Alan Turing, the great Alan Turing of imitation game fame in terms of films initially believed there was no way to actually have a physical process run fast enough to do the kind of things he wanted to see happen, including, for example, chess playing that was shown to be a groundless fear, he retracted. And then I think from that point on, the ball really starts to turn into a larger and larger snowball where we see indeed, unquestionably. Hardware gets faster and faster and faster roughly in keeping with Moore's law.

And we are clearly not exhausting the kind of hardware that we'll see right now. It's of course transistor based, maybe quantum computers will do something different, maybe computing at a physical level with light, which some people fantasize will mean some amazing speed up. But there's always been this tendency to make a prediction about AI eventually exceeding human intelligence simply because the hardware power is increasing exponentially.

Announcer:

Are there problems with hardware-based predictions?

Selmer Bringsjord:

The problem with what I described in terms of hardware focused or based predictions is that they reflect some kind of fetish, I think for physical computing speed. We've known since the dawn of AI in any kind of systematic fashion, I think that's with a liveness, so that's 17th century. We've known that the harder thing, the thing that's harder than having physical stuff that moves information around in faster and faster and faster ways is figuring out what those ways are. We've got to figure out the process that we're trying to run on the hardware. So these predictions based on speed of moving stuff around physically just reflect I think some kind of hardware fetish. The hard thing is finding out what those processes are and then once you find out roughly what they are representing them with sufficient rigor, that they can eventually be reduced to a process that runs at some low level faster and faster and faster. So they're two radically different things.

Announcer:

What is Ray Kurzweil's proposed singularity?

Selmer Bringsjord:

The singularity, this moment in time when, in the future of course, when instantaneously the first level is achieved, that is the level at which AI, now we use the term in a way that refers to machines. The machines match human intelligence first at that first instinct and then they say, wait, now that we've reached that level, we'll just build smarter versions of ourselves. We go the next increment and we iterate and iterate and iterate, and the singularity is the explosion of machine intelligence leaving us in

the dust. Maybe we end up being to them like mice are to us currently. And certainly Kurzweil believes that'll happen. The latest date I've heard is 2045, but this idea is as far as we can tell, one that is in its precise form originated by J. J. Goode, a statistician/mathematician, and he did give an argument almost exactly like the one I did, and if you do it rigorously, it turns out to make use of what's called mathematical induction in mathematics to produce the reasoning.

And there's some in the limit intelligence and the machine case that would not just be finitely but large beyond our capacity, but some kind of gap that's infinite in size between machine intelligence and us. And so Kurzweil certainly on this bandwagon, I've been hearing about this event since I was very young. Viewers may be surprised here, I was young. I remember hearing these ideas in my first AI class, loud and clear. That was what, '80, '82 or '83 at Brown University and the date always keeps getting moved out. But yes, there are such views and Kurzweil is prime example.

Announcer:

You sound skeptical. What is the source of your skepticism?

Selmer Bringsjord:

You're right that I'm skeptical, but I'm not just skeptical because I think this is a mathematical impossibility and that category includes a fellow denizen, now deceased, Kurt Gödel. Gödel said this is not possible. So I'm not just skeptical, it's not possible. One of the reasons it's not possible is in the scenario imagined by the mathematician Goode, where the machines get smarter and smarter and smarter. Remember it's the machine or class of machines at each level that creates the smarter machines. If we know anything about the mathematics of computation, we know that a machine at one level cannot create a machine at another level. If we didn't have that theorem, that would mean that any restriction on a category of machines at this level would be meaningless. We would not be able to make a blanket statement in the form of a theorem about these machines and say, hey, there's a ceiling on their capacity.

If they could just flip a switch and say, oh, by golly, we want to get beyond this ceiling so we'll just create smarter machines that work with us and take credit, it doesn't work that way and it really just amazes me. This is an elementary fact about computability theory that Turing knew full well. So the only thing you're left with since you can't get qualitatively more powerful as you start climbing up in the well known hierarchies of computing machines and an AI at the end of the day is a computing machine. The only thing you could fall back on to make sense of this view that you'll have of the singularity is speed. Things just in the sense of raw processing get faster and faster. But that basically trams everything we've learned for a few millennia now about the nature of intelligence. I mean, do we count Aristotle as a smart dude because Aristotle worked quickly?

We have no idea how quickly Aristotle came up with, he was the first logician he wrote the Organon. Who cares if it took him a month before he had that eureka moment? We just don't do that. Speed is great track and field, sports, basketball, but look, I have no idea how long it took Rembrandt to paint his... Picasso was known to be super fast. Is Picasso better than Rembrandt if Rembrandt produced some of these masterpieces over a period of months? I don't think so. Proust arguably the greatest novel ever written, took him a long time sometimes to write a sentence and craft it. He doesn't get dinged for that. So if you take speed out of the equation and say for reasons I've just sketched, that doesn't really entail an increase in intelligence in light of the theorem that I mentioned or class of theorems, I frankly have no idea how it is that there are apparently respectable folks in the intellectual landscape today who think the singularity is going to happen and not just think it's going to happen, they're really loud about this.

Perhaps they should be encouraged to wager with some people about whether the dates they've picked will see this singularity event happen.

Announcer:

What things can humans do that AI can't?

Selmer Bringsjord:

I think humans are capable, I would sum it up of the big three, the big three C's, consciousness, cognition, and it's related to the third C, but creativity, this is the problem. The challenge if you are someone in AI who thinks you can eventually get to the human case, what do I mean by the three Cs? Well, consciousness, let's take that. Much of what we do that's impressive, leverages the fact that we have an internal sense of what it feels like to do something, okay? For example, it's just not possible to write a great opera, to write a great novel, to do anything certainly in the literary artistic sphere without having an understanding of what it feels like to be someone you are not like one of your characters. And so then we go from something as exotic to that, to just getting through the day.

Like this interview here, there's something that it feels like to be in this interview for me and I know what that feels like, I'm feeling it and I'm thinking about it and I'm leveraging what it feels like hopefully to utter at least fairly coherent sentences. So consciousness, okay, are machines conscious? No, they're not. They're just hunks of stuff, moving symbols around and they feel nothing. There's nothing it's like to be my Apple laptop. Okay? Nothing it feels like to even be a robot in my laboratory. Now, cognition, people think this one is not a problem if you're super optimistic about AI, because after all AI today learns a whole lot. We have transformers, these categories, category of AI based on machine learning can speak to people and have semblance of conversation, yeah, but we're just selecting parts of cognition for these interactions with machines that are in the sweet spot of machines. Okay?

If we just ask a machine a question, the data for which or associated with which does not exist on the internet and has not been conveniently provided to it, it can't handle the situation. If I say machine, I want you now to take a sentence that I'm going to give you as a seed for a short story. I'm just going to give you one sentence. Please take the sentence and instantly make a coherent, at least somewhat gripping tiny little 500-word story from the sentence. Can you do that? Absolutely not. If plagiarizing something that's already on the internet perhaps, but all you have to do in the cognition area to paralyze a machine is ask for a demonstration of intellectual power that doesn't use data that already exists remarkably in the case of Watson, which won the jeopardy competition, the very problems in jeopardy that require thinking on your feet, novel mathematical problems.

Hey, Ryan has three apples. Okay, Jonathan has five apples, how much do we have to add to the number of apples they have to reach the seventh prime number or eighth prime number? Tell me and also tell me why, no chance. And then third, so we've got consciousness, we've got cognition, and then we have what is devastating for AI of today, so devastating that you generally don't even find coverage of this in the textbooks because we don't know how to do it. I mean, if you get the wonderful textbooks we have for AI and look for that chapter on hey, somebody must have figured out at least in part how to get a machine to be creative the student thinks, wait, I can't find that chapter. Wait, I looked in the index and I only see one or maybe no occurrences of creative. So what challenges within creativity?

Well, I've mentioned literary creativity or let's say linguistic creativity, but I think painting, if it's classical and must have some kind of narrative punch, not modern art where there doesn't have to be an underlying coherent semantic orientation for the art, this has proved to be impossible for machines. We don't talk about it much. We don't talk about how much effort has gone into trying to get a machine to paint these kind of paintings or to produce music from scratch. What we've done in the music area is

we've basically now gotten to the point where machines can, plagiarize is somewhat a harsh word, but virtually copies say Mozart. Yes, but where's the machine that creates something brand new in music, new and coherent and semantically meaningful? Nowhere to be found.

So the three C's create a real problem for someone super optimistic about the future of AI. On the other hand, it inspires me to strive to do these things because it's fun to try and I don't want somebody to say, well, you don't really know what you're talking about when you're as negative as you are, you don't really know. I want to be in a position to know as well as anybody on planet Earth how it could be that some machine is creative.

Announcer:

Why can't AI possess cognition or creativity?

Selmer Bringsjord:

One of the entrancing things for many people regarding contemporary AI, let's say 21st century AI, is that the way machine learning currently works, you end up with a black box that only uses numbers to compute some function. The reason why it's entrancing is that it seemed and maybe still seems to some, remove the burden of having to express a problem and a solution to a problem in precise terms. Indeed, so precise that you could write a computer program that captures the problem and its solution. So generally, if that's a dead end, if the magic approach to AI, it's going to dead end and it will because all it does ultimately is produce a machine which on the base of processing numbers roughly through time arrives at a function that approximates the function the human wants. Is there a cat in this photo? The inputs of the function is the photo.

It may or may not have a cat give me a one or a zero, a yes or a no. When that strikes out and becomes a total commodity, which is gradually starting to happen, then we will come back to looking at the barrier, the fundamental barrier to bringing AI to a human level, which is that we actually don't know how to capture much of what I've been talking about that's remarkable in the human mind in mathematical terms. For example, who can tell me that a thought that I might have can be reduced to something mathematical? If I decide right now I want to have a thought of some massive unicorn flying over the city of Albany, over the state house where Teddy Roosevelt, where some of his exploits were amazing, I just don't even have to close my eyes, I can look ahead, there I see it, I've about got it sized.

Okay, the thought there, is that a physical thing? I mean, what Descartes said was, well, if it is a physical thing, that's great. If it's a physical thing, then it must be a physical thing in here. If it's a physical thing in here, then we could chop your thought in half. Because we can chop every physical thing in half and that's just the start of the incoherence you arrive at. But even if one is some kind of thoroughgoing physicalist and just insists that, oh well, science must be based on pure physicalism and leaving out of the equation the problem that math presents in that case, because many infinitary structures in math clearly can't be physical. We still have the problem of figuring out how to render something precise even if it is physical.

So where is the precise description, not just of consciousness, but where is the precise description of what's entailed in writing a novel? We don't have it, so we have to issue if we're a true believer, we have to issue a promissory note. We have to say, yes, Selmer, we don't have it, but we'll get it. Well, I've been hearing the, we don't have it, but we'll get it since '29, I'm 62, I always hear it. We're going to get it. Surely there has to come a time eventually, or we're not talking about science, we're talking about magic, there must come a time after failure, failure, failure, massive gap remaining between machine and human mind where we say, okay, inductively, we have to say there's a fundamental problem here.

And the longer we go postponing the brute fact that we can't capture formally, mathematically what we need to capture. The longer we go without facing up to that fact. I think the more embarrassing it is, and I think some young people key in on this. I see this, they come into the class, we're all excited about AI, I love AI, but they're sometimes expecting to see a way to capture rigorously the fancier, grander things about the human mind and they're just not there.

Announcer:

You wrote a paper concluding that cognition is non-algorithmic. What does this mean?

Selmer Bringsjord:

I think I'm probably guilty of writing a number of papers the chief claim in which in each case is that there's something the human mind is doing that's non-algorithmic. Now, when we say non algorithmic, perfectly acceptable phrase, what it means to someone well versed in computer science is, well, if something can't be reduced to an algorithm, how do we execute it on a machine that's limited by doing exactly that? That is what computing machines do. Computing machines execute algorithms, and if we then provide suitable inputs, we get back presumably the outputs we like. So what are some of the things that at least I've claimed are non algorithmic in the case of the human mind? Well, how can reasoning that is based on infinite structures be expressed as an algorithm given that algorithms by definition right out of the whatever textbook one likes to consult must be finite in size and each element thereof must be finite.

So what do we do with the work of Kurt Gödel in figuring out that, look, there are sets that are just extremely large and bi prophesy, maybe not expect, but bi prophesy in the future we are going to have to come to grips with these sets that are so large, we just can't wrap our heads around them currently with our axioms. So we're going to have to invent new axioms. So there's a branch of mathematics within set theory, the systematic study of very large infinite things that by definition cannot be put into, nor can there be an algorithm for processing or handling these things, they're just too big. And so this is the first immediate problem that rears its head. Even if you're in first or second grade or third or fourth grade, they eventually start showing you there's a number line doing this arithmetic, these crazy numbers don't end.

And then you get a little older and you're like, you're telling me that between zero and one, if I include irrational numbers that never ends, and actually there are as many numbers between zero and one that are reals as there are all of the rationals, and you're telling me between zero and one, the level of infinity is more than the natural numbers themselves. Yes, we figured these things out as human beings, and when we did this, we went into this wilderness where infinite objects are par for the course, but an infinite object cannot be put into an algorithm. And by the way, even if someone thought, oh yes, well all that work was algorithmic, okay, it's the put up or shut up situation, great. Show us that algorithm so we can finally have our AIs do mathematics alongside us and make contributions or at least join us symbiotically in making meaningful contributions to the formal sciences. And it's never happened.

Announcer:

What is the Turing test?

Selmer Bringsjord:

The Turing test is the test obviously, in which in two rooms you've got in one and AI and in the traditional form a woman and then a judge, the judge is not allowed to look into either room. So our players are sequestered and walled off, Turing said, well, just teletype communications a little bit

obsolete now. But you get to ask questions, try to figure out which is which, which is the computer and which is AI and which is the woman. And when the judge cannot tell the difference, the Turing test is passed by the AI. Descartes said this wasn't going to happen. Turing said it would happen by the year 2000. He explicitly made that prediction. I'm still trying to figure out how, Turing is a brilliant person, don't get me wrong, how Turing's a household name and people are lucky if they hear "Cogito, ergo sum" and that's about it for Descartes, when Descartes actually wrote quite a bit about mechanical processing and intelligence versus the real thing in the human case and he looks like the prophet not Turing.

But anyway, that's the Turing test, no question about it. One could say, oh yes, Selmer, but you're not giving me the parameters in the test. You're not telling me the qualifications of the judge. You're not saying it just has to be, say the person on the street versus an AI expert. You're not telling me how long the game can be played for. Well, look, obviously the game has to be played for a long time with an arbitrary judge or it's meaningless because people can just stack the deck. I mean, over a period of 10 seconds, a chat bot could fool Einstein, over a period of 30 seconds, maybe the chat bot could fool Einstein or for that matter, the ultimate judge in the game, if you wanted an expert, I'm firmly convinced would be Leibniz, right? So or Newton, take your pick, the two last universal geniuses. That's the test, but passing, it's nowhere to be found unless you cheat by dreaming up parameters that make it feasible to play the game.

Announcer:

If a machine passed Turing's test, would he admit that it was a thinking machine?

Selmer Bringsjord:

Exactly what Turing would say if he were the judge or if you will, off to the side and witnessed unbelievable performance, arbitrarily long period of time, say it goes for a whole day and the judge is knocking his head against the wall, what would Turing say, would Turing say that's an outright thinking machine? Would he say, well, it's a good bet that it's an outright thinking machine? Would he say it's effectively some kind of empirical proof that we have a thinking machine on our hands? Scholars are still debating this. I can only express my idiosyncratic view that if we take him at his word in what he says in his paper computing machinery and intelligence, he would be forced to say, as a matter of non-revisable, very, very, very, very confident pronouncement, this is a thinking machine. His escape, if that seems too aggressive, might be to say, well, thinking is still a little squirrely. I think that's not going to work for him because he seeks to operationalize the term by way of the very test he's given us.

So Turing, brilliant as he is, I think deserves to be put in the hot seat. Not only because his prediction has proved to be way off, but because I think it's not really that plausible to say if we did have a chat bot that even worked its magic for two hours, that's a thinking thing. I mean, that's a subject or should be the subject of much subsequent debate. Perhaps if we get closer to the time when the test is passed. Right now, it really is an issue for the Turing scholars to focus on since it's utterly hypothetical and we're nowhere near having the hypothetical arrive as authentic reality.

Announcer:

How can we test if a computer is creative or not?

Selmer Bringsjord:

Arguably ultimate test of creativity for that matter, not only a computing machine or an AI, but human beings as well is if when you set them to some task, if you express some desire for what you want from

them in terms of an artifact, we tend to measure creativity, I think with great justification in both cases by expecting to get something we can say Jones is the most creative human being on the planet, but if we don't have anything that Jones has produced, I think this is the height of charity. So what can we demand of a computing machine that in terms of an artifact and it's surrounding context, that would be sufficient to warrant saying this can computing machine is absolutely genuinely creative?

Well, I think we need two things. We need a remarkable artifact and we need the artifact that is in the category of not being anat from what has come before it. And so there are people with deflationary views of creativity. Well, don't talk to me by creativity, people are just rearranging the symbols that were already there before or something like that, or the patterns. It's just a new pattern. No, because we have absolutely decisive reputations in the case of human beings. In this session, I've mentioned Proust. Would someone like to show me the antecedent to Proust and Proust's prose? I would love to see it, please. Not just his philosophy of time and consciousness, I just want to see the prose that came before him. No, it's new. Okay, can someone scoundrel though he was? Can someone please give me the antecedents to Wagner and his creative work? Awful person. The artifacts though are not anticipated, so we need that from the machine.

And then secondly, we need to make sure no one cheated. We need to make sure there isn't a Wizard of Oz situation going on, that there isn't some human in the background that enabled this to happen. But we just don't see that. And the way to do that is to ensure that everyone who knows anything about the system, including the system's designers and engineers, look at all the details from overarching system, architecture, background mathematics to details inside, if you will, the code, and they say on our oath, we have no clue how the machine is doing this.

And that's what I've called with my co-authors, Paul Bellow and David Ferrucci, the Lovelace test. If it takes those two things, a machine passes that, it passes the Lovelace test, and then I'll say with everyone, wow, something is really going on here in terms of creativity. So we don't have one, the first condition satisfied even on the horizon, and we clearly don't have two. Someone might say, what are you talking about? You admit we have black box systems that have learned. Whoa, whoa, whoa, whoa, wait, we know what's inside those systems, okay? We engineered the start of the process of learning the function, that's what data scientists do. We stacked it all together. We knew what the data was that was going to be used for learning. It was old data. Without massive data, you're hamstrung in this approach. So no, that doesn't count, that's not what I'm talking about. So that would be the test I would urge people to use in order to ascertain whether a machine is genuinely creative.

Announcer:

Can Watson or Deep Blue past the Lovelace test?

Selmer Bringsjord:

We've seen a litany of supposedly creative AIs working their wonders in games. First it was checkers. The AI was Chinook, then it's chess. The AI is Deep Blue, and then it's jeopardy and the AI is Watson, and then it's Go with a higher branching factor than chess, but by the way, no more difficult than chess, there are theorems that tell us this, but then it's AlphaGo. Again according to some doing creative things in the game. If we apply the Lovelace test to any of these machines, it fails on the first criterion. It has to produce an artifact that is genuinely in the realm of the creative. All these machines are just following algorithms slavishly. We already know there's a perfect winning strategy for Go. If there's a perfect winning, well have a devil of a time implementing it, but if we know there's a perfect winning strategy as a hard and fast algorithm for doing it, then marching along with the algorithm, it could be a bunch of turtles that are following the algorithms or birds that are strung together following the algorithm.

The minute there's an algorithm to just crank it out, you fail on the first score of something that's genuinely creative. With all due respect to Go players, I know they make creative moves, but I'm sorry, that's failure. And then secondly, we do know what they're doing. We do know what these machines are doing in terms of the second criterion. Nobody is at all stupified, not even amazed that AlphaGo was able to do what it did. How do we know? We can just look at the transcripts of the engineers along the way to know that they knew exactly what they were doing and that it was going to work out.

Same thing for Deep Blue. When Deep Blue ran into trouble against Gary Kasparov in the match where Kasparov lost, it had a problem figuring out King's safety. Okay, how did the engineers know that it would be doing a little better in the near future when it came to King's safety against Kasparov, they put it in, they figured it out, they consulted with the grand master or master Joel Benjamin, I forget, but it was Joel Benjamin. Figure it out, put it in. Does that make the machine creative when it does something Now that's innovative in the case of King's safety and chess? No, it just followed the basic heuristics that were handed over to it.

Announcer:

Will the Lovelace test ever be passed by AI?

Selmer Bringsjord:

I don't think the Lovelace test will ever be passed by an AI if there's no restriction on the depth, richness, of the artifact produced. So a full length novel, it's going to be very, very, very, very, very difficult, I don't think so. If it's not the novel, then I'll pick anything in the realm of the frontiers of the formal sciences, mathematics, logic, game theory, the formal sciences overall. So my answer is no. The human mind will always be with sufficient, dedication, enthusiasm, training, perseverance, thereby able to reach its full potential. Unbeatable. All bets are off, I don't want to come across as an elitist. I think maybe inevitably that's what it sounds like. When I'm talking about the human mind, I'm talking about neuro biologically normal human minds that have been educated and trained on the best previous generations have produced. And so, no, I don't think a machine will ever relative to humans in that category and what they judge to be in the first criterion, a genuinely creative artifact passed, I don't think so.

Announcer:

What would you say to people who claim AI is creative?

Selmer Bringsjord:

Some claims regarding AI as creative, that is some positive claims to the effect, that listen, this AI is creative, are true. I think world's leading authority on AI and musical creativity Cope explicitly says this and says this repeatedly in writing said, here's my exemplar. This is a musically creative machine. The problem is he then to his credit, offers a definition of what he means by creativity. And for him, problem-solving counts as creativity, pretty much generic problem-solving. I don't count as creativity the solving of SAT quant problems, though, that's problem-solving one after another. I'm sorry, I don't count that as creative. But there are other people in this general camp, Aaron, the AI painting system, people have said that's creative. Yes, it is in your sense of creative, that might be emotional impact for a human looking at it.

They have an experience. Art produces experiences on the part of viewers. You look at an Aaron painting, it produces an experience. No, for me, I have higher standards in both cases. I've got to have a semantic meaning that I can attach to something that is supposedly a creative artifact. Otherwise, this becomes entirely subjective. And anyone's super optimistic about whether AI can be created creative is

just entitled to their own view and can declare it. It's really hard to have an opera that's great without some kind of storyline. That's why I know some of my friends who non-accidental atonal music, they say, well, you're just brought up in the west. But there's got to be some semantic structure, some meaning for it. And it can't be derivable from the antecedent, which clearly happens in the case of an SAT problem. In fact, the SAT problems are essentially clones.

And that's why when you get, I suppose, proficient at solving them, you can develop a degree of confidence if you're doing this for a living that throw at you what they might in terms of these problems, you can solve them. So I will admit that there are true claims to the effect that AIs are creative, but you always got to check the definitions. I'm not aware of any credible claim to the effect that this AI is genuinely creative where the artifact passes the test that I'm talking about. Interestingly enough, the economists would probably say, well, so what you're saying is prices aren't affected by the artifact? Yeah. If you've got an AI that solves SAT problems, unless that serves some kind of purpose of generating the test to save money, what good is that? What good is Aaron the painter? What good is Emmy?

What good is the musical system that emulates say Mozart? Well, what good is it? From an economic standpoint? We have a tip off the economist would say, is that very valuable? It has no effect. It might change the careers, maybe it indicates human capital's been used to reach that, but that should make us a little bit skeptical. If the novels start coming into the slush unsolicited slush pile to the literary houses in New York City and the editor says, oh wow, we have a blockbuster on our hands and it's literary fiction and it's written by a machine, that'll change everything economically.

Announcer:

What do you think about Roger Penrose's argument that computers will never be creative?

Selmer Bringsjord:

Human beings who take positions on big questions about really anything but the nature of the human mind versus the nature of the machine mind going into the future and offer some cogent rationale for why they believe what they believe can be right, even if their reasoning is wrong. And unfortunately, in the case of Roger Penrose, he won the Nobel Prize for excellence in mathematical physics, not for excellence in mathematical or formal philosophy or logic. So I have a couple of papers on Penrose's arguments in Emperor's New Mind and Shadows of the Mind. I think the reasoning is poor and fails, but he is right, he is fundamentally right. So I'm in a tricky position and a number of people have explained, I think in print or tried to explain why in the details he is incorrect. However, someone with that intellectual heft that of Penrose who has taken this position, tried to teach essentially himself all the relevant material, being trained as he was trained, some training that would be relevant to what we're talking about in this domain and do what he did is remarkable.

But that's what you would expect from a Nobel Prize winner in physics. But unfortunately, it's really a perpetuation of an older line of argument. I think that originated with JR Lucas claiming that because Gödel has a negative result in his, or two really, but the first incompleteness theorem can be used to claim that the human mind exceeds the machine mind. I think everybody tempted to go that route ought to look at the history books, the technical history books, and see that there was a guy named Zermelo, mathematician/logician who looked at this and said, really negative result. You're telling me, wow. Yeah, you're telling me that under certain assumptions there would be propositions that can't be proved, nor can their negations be proved. And you're telling me these are number theoretic propositions, they're about arithmetic. What are your assumptions? Assumption number one, it's all finitary.

The axiom systems are decidable as well. So the axiom, you have to be able to decide when an axiom is part of the axiom system. And Zermelo said, and I'm with Zermelo, this has nothing to do with a limitation on the human mind. Trying to use these results in naive fashion is why Gödel himself became very angry with a famous book called, I Believe Gödel's proof, it's a book by Nagle. He was invited, Gödel was invited, you can find this online if you dig a little bit. You can even find the correspondence in the terms that Gödel laid down. He would not countenance naive reasoning from his results to the superiority of the human mind over machines because he knew full well there were intense, meaningful determinate assumptions that if you threw them out, the argument would have, if you were willing to throw them out, I should say, the argument would have little power.

That said, I do have a paper in which I try to renovate and circumvent the concerns that Gödel had by saying, well, we do have to examine these restrictions and consider what would happen if we did. And unfortunately, the paper's a bit technical, it's called The Modal Argument for Hyper computing Minds. But the long and short of it is, you can be right, but not for the right reasons.

Announcer:

How is Penrose wrong on this topic?

Selmer Bringsjord:

The great mathematical physicist, Roger Penrose in a series of publications, not just one book, Emperor's New Mind or Shadows of the Mind, but most recently, and after those books in a small journal called Psyche, makes the argument that in light of what Gödel's negative results tell us, they tell us there are limitations to a certain kind of mathematical reasoning. Okay? The idea is well, that those limitations can be exceeded by the human mind, they cannot be exceeded by a computing machine and therefore by an AI which is a computing machine ultimately. Penrose says basically the human mind is greater, it'll never be matched. And again, this is in a series of publications, not just one. And with contributors to the Psyche paper where a number of people quite knowledgeable about formal logic say, well, Roger, cool, maybe we'll convince some, but there's some serious technical issues. So I find myself very sympathetic. I think he's right, but not for the reasons that he gives.

Announcer:

What are your thoughts on John Searle's Chinese Room example?

Selmer Bringsjord:

Unfortunately, computers cannot understand anything and that they can't is rather famously established by John Searle. It's galling to the AI folks. It's sort of out of vogue now to express being upset by Searle's argument. But there was a point in time when there wasn't an AI person on the planet who hadn't read the argument and had strong reactions to it. So I'm with Searle. Not everything, Searle, like all human beings says is correct, but a lot of it is, and Searle just nailed it. Computers can't understand anything because as he explains, they just mindlessly manipulate symbols. And whether it's a quantum computer, whether it's the greatest chips available on the planet today, whether it's some future light-based form of a computing machine, only symbols are being manipulated and that's it. There's no meaning attached to those symbols for the thing that manipulates them. And that's what you need for understanding.

You also need consciousness for understanding. Okay, if I say to someone, a finely prepared halibut with just searing on the outside in butter and then taken out of that frying pan as one of my favorite chefs does, and then let it gradually cook because it's very thick on the inside, when you taste that, oh man,

it's just astonishing. Well, to understand what I just said, you have to have consciousness. Now, Searle, establishes this in his, I think it's Minds, Brains, and Programs where he introduces what we now call the Chinese Room. And in the Chinese room argument, you have Searle in the room who doesn't understand any Chinese, you'll have Chinese speakers, native Chinese speakers outside the room and they send to the room pieces of Chinese and they get back pieces of Chinese Mandarin I suppose technically specifically in return. It seems to them that the room understands Chinese.

Unfortunately, the fact of the matter is that Searle is inside the room following as he puts it a rule book. But to make this in modernized fashion, we have to say he's got the ability to hand simulate some computer program and that's all he does. Does he have to work fast? Yes, it's a thought experiment. Just like Einstein had thought experiments that don't make really any sense at the end of the day in terms of naive physics or even established physics, we understand the point of the thought experiment in this case, and he's right. The thought experiment shows that there is Searle operating precisely as a computer manipulating symbols only. But since Searle doesn't understand any Chinese, he understands nothing of what he's doing and he is directly analogous to what the machine is doing when it follows the program. So I am one of these folks I've tried to meet in the argument, I've tried to make it more precise, but I can take zero credit, nor can anyone else, in my opinion.

He really did come up with a landmark objection to AI and it'll come back, it's going to come back when we go through another peak and valley turn where people are like, okay, this machine learning is getting us machines that chatter and speak, but wait, they don't seem to understand what they're talking about. Right they don't, they're following patterns established by prior data and learning, shallow learning applied to them. So they have no understanding and that's why if you ask them a few simple questions, they're like clueless.

Announcer:

Will computers ever have common sense?

Selmer Bringsjord:

Common sense has been the Achilles heel, at least in the minds of many critics of AI for a long time. I'm not so sure. I think this is a mantra. I think we're better off if we want to be honest in assessing AI's power relative to the human mind to look at the extraordinary things the human mind is able to manage. And that may be clear from some of what I've said in the past. However, I think ambiguity and specifically context in communication, I don't know if I would call it common sense. I would say it's more specific than that. Ultimately, I believe common sense, if it's background knowledge that's brought to bear, if it's getting that background knowledge, I think that will happen. And I think the vision for doing so that famously, at least for people in AI, Doug Leonard is still pursuing, I think that's going to work.

I'm not worried about that, it's an unbelievable challenge. What I am worried about is in communication when humans talk to each other or even they're reading a document, anything at any moment can make perfectly good sense only because of its context. And so there's ambiguity. You don't know if I say to you, well, take a sentence that seems to be absurd. The camera ate the mouse. Well, how can a camera eat the mouse? How could that be coherent or anything like that? Well, in any number of contexts, that could be perfectly understandable. If someone had just asked the cameraman, did you actually track the mouse that ran across in the background there and get that? Maybe it's some clever way for him to say, yeah, the camera swallowed the mouse, something like that. Then the context makes it clear.

The problem is this is ubiquitous in human communication and context continuously clarifies and makes sense of what we're talking about in any particular instance. How many people today working in the part of AI that pertains to this, it would be natural language understanding are working systematically,

energetically, and with some degree of funding on this problem? Almost zero. Because what's happened in natural language understanding pursued by machine learning is that we assume all the past data regarding conversations are going to help us. Yes, but what if we get into situations where the context in that particular case decides the issue? How is the AI going to figure it out?

So that's a specific thing that I think haunts these transformer systems as they're called. There's GPT3, people can watch plenty of, no doubt there'll be subsequent ones, but watch plenty of videos on YouTube now to see it having conversations with humans. As we've discussed it doesn't understand what it's saying, but leaving that out of the equation, go to the transformer and ask it something, which if it is to understand it, is based on the context right there. So bring the transformer into a room or have a video and an avatar and say, gee, it's amazing that as smart as you are, you fit into a box that's less than a foot square.

I don't think the transformer will have any idea what is being talked about because the context is the surrounding environment with a computer and a monitor, et cetera, makes that understandable, makes it coherent. So that is one thing that I worry about. I think it might be provable that this is not computable, and I do have a first early paper on this. I think it might be provably beyond what a computer can do to ask it to figure out the meaning of a sentence in an arbitrary context that is given any context, but a specific, let's say a single sentence, what is the meaning of that sentence? Given that the context could be anything, how do we compute that? I'm not so sure. I don't think that's computable. And if that isn't computable, the Turing test is not going to go the way even if we give Turing plenty of leeway, we didn't get it done by 2000, we say 3000.

I don't think it's going to be so easy as if we allow the time interval for playing the game to be increased in light of the problem that I'm talking about because you should be able to ask questions that somehow leverage context, not just background to common sense, which again, I agree is a big problem, but I think that'll be knocked down. How about the local context of the conversation? If I'm having a conversation right now with anyone, I can assume that they know something about current events, also things that are dynamically changing that forms part of the context, that's not frozen in data yet, presumably. So I think this is going to be an opportunity for younger people who are willing to take on a clearly important area of AI. Because even if you meet with moderate success, you'll make the world, you'll give the world some amazing technology, no question about that. Even if you ultimately strike out, there's huge opportunities in this area.

Announcer:

Will the speed of quantum computers allow them to achieve true human intelligence?

Selmer Bringsjord:

If one believes that the essence of intelligence and indeed ever increasing intelligence in some kind of hierarchy from non-human animals to chimps to us, possibly aliens and supernatural beings, and God himself, if one thinks that there's some hierarchy going on here, one then has to ask what's it based on fundamentally? And one answer is speed. Speed of processing, the time it takes to do something. If one has that view, then given what's happening in the world today in connection with quantum computers and some other things that are even possibly more powerful speed-wise, then one will inevitably think, wow, this is going to be a rollercoaster. We're going to see smarter and smarter machines. If one thinks, as I do that speed has very little intrinsically to do with intelligence. You have to be fast enough to deliver and maybe it's therefore part of the ball game, but it's by no means the essence of intelligence at whatever level we're talking about in this imagined hierarchy, then I think speed is going to turn out to be quite irrelevant.

Now, the US has invested inordinate sums of taxpayer money in pursuit of quantum computing. It is now an industrial pursuit, we know this. We have the amazingly interesting chapter where Google declared quantum supremacy, the leveraging of com... And then IBM said, well, that doesn't seem any better than what you could do with our super computers. So what's going on here? This doesn't seem like a big deal. All right, so industry's in on the ballgame now, but taxpayers have invested a fortune in this. What is this going to do? This is going to produce at its best machines that can solve problems thought to be for mathematical reasons currently infeasible. Okay? Infeasible, not impossible. Infeasible because they take the problems, demand so much processing as a function of the size of the input in terms of time and space that things are not feasible, turn something that's infeasible to feasible.

But I'm talking about the mind being able to do things that are not just infeasible for an AI, they're impossible for an AI. And in this regard, Kurt Gödel, I think would have the same attitude. He'd say, well, that's astounding, or that would be astounding, that'd be great. But wait, if it's not even conceptually possible for a computing machine to solve the kind of problems that humans solve, no matter how much time is available and space and energy, then for purposes of those debates, a quantum computer is not going to change the situation. Quantum computers may change security in the world as we know it today. Basically, nullify is the fear that's been expressed for a long time. Nullify our security systems because they're really based ultimately on what's feasible and infeasible. That's why, you don't have a combo lock.

You don't have a combo lock on your lock in the locker room that has one number from one to 10, not a smart thing to do. Nine, oh, I don't know, I've got time, one through 10, you're dead. Things like that could very well be upended. And from a military perspective where time in decision-making of a machine aiding a human might be the difference between winning or losing, it's huge. Which is why the primary investment, as I understand it came certainly for a period of years out of the intelligence community research wise. But I don't think speed and quantum computing or any other such fancy incarnation is going to make a hill of beans difference compared to the human mind.

Announcer:

Some say we will be able to upload ourselves into computers one day. Is that possible?

Selmer Bringsjord:

It might be offensive to some, but I confess I do find some of contemporary AI, especially in connection with these ideas of uploading the mind and living forever to have a disturbing religious quality. I probably shouldn't say by the way, religious quality, maybe the absence of any systematic religion is going on here. It's magic or superstition. I remember being invited to give a talk at a life extension symposium. This is a number of years ago, I'm not sure if this is in my CV, and there were some amazing theoreticians there working on the mathematics of say nanotechnology to fix what goes wrong in your brain after it's been frozen so you can be defrosted so that you can live forever because little machines are going to work their wonders.

And I remember raising my hand and saying, do any of you realize that it's infinitely more plausible for me to say I already planned to live forever because I'm a Christian, than it is for you to futs around with all this nonsense? That is if I just give you a proof for God's existence and you find it credible, I've gotten a lot further than all this mumbo jumbo you're talking about has gotten us. Now. Okay, maybe that's not fair, but come on, uploading the mind, uploading the mind. That means that which is uploaded has to meet certain, I would think, relatively rigorous, necessary conditions for being uploadable. What are those conditions? We actually have a mathematics of that which is uploadable? Instead, we basically still have Star Trek, Star Trek is fiction. We don't know what these conditions are. Who's going to be the first

one to line up, get right to the head of the line and say, gee, you know what? I'm really happy to be uploaded.

Even though you don't know what it is, you can't define it mathematically, I'm ready to be uploaded. We can't mathematically define what a cell is, okay? We can't mathematically characterize... Let's talk about the math of computer science. We can't come up with a mathematical description of a cell sufficient to allow us to manufacture one and duplicate them at will for various purposes. And yet we're going to upload something that we have zero understanding of, at least in the case of biology, whether it's synthetic or whether it's in the area of virology. I'm sitting here without a mask on because of RNA technology that's amazing. Where is that for this uploading thing? So this is all just unfortunately born of an acute desire to just live forever, and no matter what the science tells you, just try to figure it out.

And I'd be willing to pay attention if someone would lay down for me here, here's the criteria, here's the criteria for what it means for something to be uploadable or how about teleportation? Teleportation, okay, shouldn't that be formally or mathematically computationally, if not isomorphic in the same category as uploading a mind? Just teleport me to London. So as much as I love American Airlines and make the flight a lot and love it as much as I like it, just teleport me right to downtown London instead of going through Heathrow when I lay on there, which everyone knows is it's a great country, it's great, but it's a circus.

Show me how you're going to do that. And then perhaps I will pay attention to the uploading concept. Now I'm being harsh about the uploading. All right, maybe just trying to get a smarter and smarter artifact and leave the more exotic stuff. Is there a religious quality to AI of that type? That is the real optimist, just the hardcore optimist and some of these folks, many of them are my friends. Yes, AI will continue marching inexorably higher and higher toward that great peak of human intelligence and then it will go beyond it. And, sorry, Selmer, that you're so cynical, you need to be enlightened. Well, when would you be willing after the mountain has not been climbed, despite repeated attempts to say, we just at least can't do it, or we should seriously consider the possibility that we can't do it. If you never have that attitude, if you're never willing to consider that it can't be done despite failure, upon failure after trying, is that not some kind of fanaticism? Is that not superstition? Is that not a belief in magic? If it isn't, I don't know what is.

Announcer:

Is it dangerous to give AI too much credit or control?

Selmer Bringsjord:

Fortunately, I think cooler, rational minds are prevailing against the tendency, which is a foot in some quarters, to give the machine more and more control because the machine is going to exceed us and be God-like and its insights and intelligence. I don't think, my hat is off to, especially folks who deal with life and death issues, military and intelligence. They understand that this is a kooky concept, unless you bring me verification of machines that have lesser powers. Even there, I'm really concerned about giving the machine more and more control and power. But yes, there are certainly folks who don't just get the engineering high, I think that I get of doing AI to make a cool thing with genuine intelligence, even if it doesn't exceed human level. Okay? It's beyond that.

It's like we want that human level and then by golly, we want what's beyond it, because that will be some kind of supernatural agent to really maybe clear things up and we'll be these creatures that have given birth to these wonderful things in the universe. If I'm not mistaken, I think Seligman, the little God of positive psychology actually considers advances a secular eschatology that is directly in line with this, even though people don't think of him as an AI person and he's not. But that's basically the vision that

we give rise subsequently to creatures by our hard work that are greater and greater than us, and we should be happy now that we're playing this role, this is absurd.

Announcer:

Are you a materialist? Why or why not?

Selmer Bringsjord:

I'm not a materialist or physicalist for reasons first that have nothing to do with the human mind or a spiritual realm in anything like the traditional spiritual sense. Maybe that C. S. Lewis would use that. I remember he says at one point, look, they're simplifying grotesquely, actually. There are two kinds of people in the world, those who might walk out on a sunny day and say, ah, wonderful, another sunny day, I like the weather and everything there is just physical. And then the other type of person who walks out and says, oh, this is a wonderful glorious day, and there's so much more that I'm detecting than the merely physical because there's stuff non-physical behind the scenes. I remember reading that and thinking, all right, that's really cool, really great. But many people already knew that before going outside because they just thought about mathematics.

So it's really hard to find an irrational number, but I believe irrational numbers are perfectly real. It's really hard to find an algorithm. I mean, there are famous algorithms, quick sort for just sorting a bunch of stuff when they're jumbled up and you know how to order them. What's an efficient way to do that? There's an amazing algorithm from Tony Hoare that still stands to this day. Where is that algorithm? Well, what do you mean? Where is it? It was on the page over there, you can take a look at, it's on Wikipedia, but are you sure? Because when I looked at Wikipedia and looked at the book, I saw two different inscriptions. So neither one of those can be the algorithm. Where's the algorithm? Where is any significant mathematical object? So for me, materialism is right inside my house with my eyes closed, thinking about basic questions regarding what these things are absurd.

And then if we get there, we can consider what kind of entity it would take to appreciate that immaterial reality. And we might find that in the C. S. Lewis sense or what he intended, that there must be mines that are able to access immaterial ponder, and perhaps mines that are themselves immaterial, which is of course, what Descartes said. Now, there are a few people in that category and professional philosophy these days, but I would really love to see a debate materialize in 2021 or 2022 with Descartes coming back. Maybe we throw in Turing for good AI connections and we have Descartes and ask Descartes, you see you were wrong. And he's like, what are you talking about? You see I'm wrong. Nothing has changed. Everything's exactly the same as what it was. In fact, all my predictions have come true with regard to AI and the mind, and you still don't have any degree of confidence, rational confidence, that mathematical objects are physical.

And the kicker would be if it was Descartes, let's say Turing. Okay? And maybe there are aliens, that's interesting to consider. If the aliens show up and join the discussion, what are we going to say if they got here by using technology in their ships that exploited the very same underlying mathematics that we have discovered, we can no longer say, oh, we just made the math up. It's like a game. We create the concepts and the symbols and then we kind of use it. No, and I mean that seems like what would happen, just hypothetically. If they show up in a ship and the ship got here flying really fast, how exactly do they do that without using the math that I'm pointing to in the start of this line of inquiry? So, all right, I'm not a materialist, I'm with Descartes and I'm with Gödel himself.

Gödel said, look, I think the brain is to how when..., I think the brain is probably a finite digital computer in biological clothes. Great. So the thing inside my cranium is a finite digital computer. What am I? If you answer the question well, you are that, and since that is a small, purely physical thing, you're a small,

purely physical thing. And Gödel would say, would you mind please establishing that I'm identical with either my brain or at least maybe my central nervous system? 'Cause I'd like to know how that works. And that's proved to be a rather tough thing to do.

Announcer:

This has been Mind Matters News. Explore more at mindmatters.ai. That's mindmatters.ai. Mind Matters News is directed and edited by Austin Egbert. The opinions expressed on this program are solely those of the speakers.