

God, AI and the future of humanity: Is technology the key to immortality?

Rosalind Picard, Nick Bostrom
& Justin Brierley



Contents

Audio transcript: God, AI and the future of humanity	3
Professor Rosalind Picard: 'I used to think religious people had thrown their brains out the window'	43
Is Artificial Intelligence the new frontier in the search for immortality?	50
Fearfully and wonderfully made: The promise and perils of creating AI	61
Moses responds to those seeking eternal life through AI	70
If AI needs an intelligent creator, so do we	81

God, AI and the future of humanity

**Rosalind Picard, Nick Bostrom
& Justin Brierley**

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley



Rosalind Picard
(RP)



Nick Bostrom
(NB)



Justin Brierley
(JB)

God, AI and the future of humanity

JB: Well, hello and welcome to The Big Conversation from Unbelievable brought to you in partnership with the John Templeton Foundation. I'm Justin Brierley and the big conversation is all about exploring the biggest questions of science faith and philosophy with leading thinkers across the religious and non-religious spectrum. And today we're talking about God, artificial intelligence, and the future of humanity, and asking, is technology the key to immortality? This season is, of course, being recorded for obvious reasons remotely, but that does allow us to bring together some really interesting people from all over the globe and joining me today are Nick Bostrom and Rosalind Picard. Nick Bostrom is the Director of the Future of Humanity Institute at Oxford and a leading voice in AI and Technology. Nick's the author of 'Super Intelligence', which warned of the possible dangers that the rapid development of artificial intelligence may pose globally, Rosalind Picard is director of affective Computing at MIT, harnessing the power of AI to develop things like wearable technology, machine learning that predicts and responds to human emotion and well-being and helping people with autism, with epilepsy, depression, and many

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

other physiological conditions. So, today with Nick and Ros we're going to be looking at the future of AI and asking big questions like, the ethics of AI, could machines become conscious one day? Might we all be uploaded to some digital heaven, and what if we're all living in a universe simulated by God-like alien technology? Just a few of the little questions in life. So, I'm really looking forward to this this programme. So, good to have you with me Nick and Ros. Tell me, first of all, I will start with you Nick. You head up the Future of Humanity Institute. It sounds like something from a sci-fi film or something. Are you basically there to sort of predict doomsday, technological scenarios and just make sure they're averted? What is it you do at The Institute?

NB: Yeah. Especially the latter. If I could either predict or prevent, I'd choose to prevent. We are an interdisciplinary research group. We have computer scientists, mathematicians, philosophers, physicists trying to get the clear understanding of some of the really big picture questions regarding the future prospects for intelligent life, threats to the survival of the human species and, in particular, the levers like where are the



God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

opportunities whereby some modest effort now, we might improve the expected value of the long-term future.

JB: Does that just involve a lot of sort of sitting around and thinking through issues. I mean, obviously you have to keep abreast of all the latest developments I suppose in technology doing that.

NB: Yeah. Well, a lot of our work is focused more on the kind of underlying questions rather than what's next year's iPhone going to be like. These kind of ripples on the great pond are not really our focus. But sometimes the most important questions are the ones that get the least attention. So, for example, if we continue to make impressive progress in machine intelligence what happens down the line when AI finally succeeds at, not just automating specific tasks, but providing the same general kind of intelligence and problem solving that has made us humans unique on this planet, and other more ways that the human condition could change in some more fundamental way. It's very much big picture stuff, obviously.

JB: I know you don't have any personal religious beliefs, Nick, but does the issue of God ever come up in your research, questions around religion, and that kind of thing?

NB: Well, it comes up in my own thinking, not so much in the specific research products we are pursuing. And I can't speak for all of my colleagues there. But at some point, I think if you zoom out enough and start to think about really what is the human condition all about and what might it lead to, you do start to butt up against theological concerns. Unfortunately, that's also the point where my competence kind of runs out to some extent.

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

JB: We'll be testing your competency a little bit on the programme today Nick because I would love to open up some of those metaphysical questions that the AI pose around consciousness, ethics and even the god question at some level, but we'll start obviously by talking about AI and what some of the possibilities and challenges are. Before we do that, Roz, welcome to the show as well. Now you've pioneered this this field of affective computing for over two decades, I think. Tell us firstly what it is, and what sort of things you've been developing in the process.

RP: Yeah, affective computing is computing that relates to, arises from, or deliberately influences emotion. The original idea was that computers were driving us nuts. They were so stupid and frustrating and the kind of intelligence we wanted to build for the future AI was not just playing games like chess or Go, or using language or solving maths, but was not being annoying, right? A computer or a robot that you would actually want to have around and that if you were frustrated or pleased it would see that use that as feedback and learn to do better and have skills for how to manage emotion. Not just recognising them but knowing how to help us manage our emotions in respectful ways, always with deference for people's feelings, not sensing things if we don't want them and so forth. So, I saw it as a hierarchy out of wanting to respect people's feelings and that these were skills that we could teach computers how to do better at it.

JB: And what has that resulted in terms of some of the practical applications that you've been able to engineer in the process?

RP: We have been surprised. In the beginning the idea I think was very much about like, what would a robot do, what would a software agent do

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

in interacting with a person to be intelligent? And what might happen deep inside a machine that most people don't see that might be like how the human brain works that might enable it to think more intelligently, not that computers think like we do, but to have processes that give it better flexibility and generalisability and things like that. We have actually found that there are a lot more applications than that and many applications have to do with human health, for example we learned that the fundamental mechanisms inside us when we're having feelings communicate with pretty much every organ in the body. And when you look at the messages, they carry they change when people are healthy or sick and by paying attention to these unseen aspects of our motion system, we could help people have better health. So, we have been working on a lot of cool things that have come out of that. Like we accidentally developed a seizure detector. Which actually became the focus of [name 00:07:30] PhD thesis and the focus of a spin-out company, Empatica, that's now an FDA cleared device that I'm wearing here called Embrace. But we've followed the data where it goes and wherever we can look to see how if we can steer it to help make lives better.

JB: Well in that sense, it sounds very much like you're obviously harnessing technology for a wonderful end, even if there may be potential pitfalls in other parts of the AI story. Before we start talking about that, I'm fascinated a little bit about your own faith journey, because I think you had an adult conversion to Christianity yourself, Roz. So how does someone who's evidently intelligent, scientifically minded end up believing in God?

RP: There's some longer versions of this people can read if they go to places like Christianity Today or other talks, but the short version is I was an atheist and a sceptic and a very proud one, and when I was challenged to read the best-selling book of all-time – the one that is so best-selling

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

they leave it off the bestseller list because it dwarfs everything else each week – as I read that book I started to realise there was a lot of wisdom and intelligence in that book and that it wasn't just a bunch of goofy made up stuff that I thought it might be when I went to read it. And through a process of reading, asking questions and many years of gathering data about it actually, I gradually felt myself changing. I became somebody who started to believe not only that a possible God existed, but maybe it was more probable than not, and I started to see if that made a difference in my life, and then gradually started to embrace the Christian worldview, and that has actually made a huge difference in my life. So, now I've run the experiment both ways and I definitely prefer the Christian worldview.

JB: And again, just briefly, how would you say that that particular change in your own life has impacted your career, your scientific work and everything?

RP: It's funny, when Nick was saying we bump up against the boundary



God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

of theology I'm reminded of Donald Knuth when he was visiting, and he's a practicing Christian, a very famous computer scientist and when he came to speak about theology, he said when it comes to theology – mind you everybody in the room thought he was the world expert on computer science – he said when it comes to theology, I'm a user not a developer. We aren't inventing that, right? There's some truth there that we are barely catching a glimpse of. And in my work the more I learn about how the human mind works and how the human emotion system works, I'm just in awe. I'm in awe of how fearfully and wonderfully we are made, and it inspires me, and another key piece of my faith is that all people are of equal worth and that has driven my willingness to work on topics with people who are often stigmatised for conditions that other people didn't value so much. They didn't value those people so much because they had those conditions which is a horrible tragedy, and knowing from my faith, this view that we're all of equal worth, I have spent a lot of time learning from them, and it has just been tremendously productive in terms of just wonderful relationships, wonderful insights, and now pretty much all of our research has gone in really cool new directions because of things we've learned from their different kinds of minds.

JB: It's great to hear that. Such an interesting area. I could spend the whole evening talking about that. But we've got all kinds of areas that we want to cover in today's programme. I want to talk about firstly, the future of AI and whether the sci-fi movies have it right. Now, I don't know what you think of this Nick, but I guess ever since things like 2001- A Space Odyssey, you know, where you've got Hal basically saying, 'I can't do that, Dave,' and that sort of thing, we've had this fear that technology will one day essentially get a mind of its own and work against us in some way. How broadly optimistic or pessimistic are you about those kinds of foreboding challenges around

AI, Nick?

NB: I'd say I'm a fretful optimist. I think it's going to be transformative and that both very, very good outcomes and very bad outcomes are potentially on the cards, and we have a lot of ignorance as to how these kinds of things play out. We've never seen a transition to machine intelligence era before. We don't have that much evidence to go on. So, some quite large amount of uncertainty seems appropriate in terms of our expectations.

JB: I mean, what are some of the potential challenges or even pitfalls that you anticipated in super intelligence? And I'm aware the book came out a few years ago now, so do you think that the landscape has changed a great deal in terms of people's I suppose awareness of some of these areas of risk?

NB: It has changed a great deal, I think. Well, first in terms of actual capabilities, I think progress has been faster than was generally expected back in say 2014 when the book came out. Really exciting developments in deep learning and a lot of things every year. Another way in which the situation has changed is that when I wrote the book it was partly to draw attention to the need to do research in advance of how to align very powerful AI systems with human intentions, or other ways of making them safe and controllable back then. It was an extremely neglected area. It was kind of relegated to science fiction authors to just make up fun and interesting stories like the one about Hal and many, many others, but that was not like when you have people being paid to do research on the dung beetle and there's like a whole community with peer reviewed articles studying every aspect of the dung beetle's life and physiology, but this didn't seem to qualify. But that has changed in the years that have passed. There is now a vibrant research community working on the alignment

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

problem and we have a group doing that at the Future of Humanity Institute, but at Deep Mind, at Open AI and at a number of other places. There are now some really smart people working on this.

JB: Do you consider that there is a genuine threat at some point in the future of some kind of cybernet type situation developing where AI will get to the point where it's become self-aware and develops its own attitudes towards the world and how things should be run and turns into some kind of cyber dictator. I mean, is that just sci-fi as far as you're concerned, or is it a genuine thing that could happen if we didn't, keep an eye on and moderate the way we develop AI?

NB: It's easy to anthropomorphise AI systems and that usually blocks understanding, but it is true that if you have very powerful optimisation processes where you have some kind of learning mechanism that tries to gradually improve performance against some specified objective, that you often find that the ways in which the system goes about meeting the objective is unexpected and sometimes undesired. And the potential for this kind of unexpected and undesirable way of meeting the nominal specification that we have put in increase the more capable, the more creative, the more intelligent the system becomes and so, if we extrapolate beyond the systems for now, where it's easy to go in afterwards and tweak it and change it if we don't like, but if we imagine that one day this might be smarter than humans and therefore extremely capable and maybe extremely powerful, then it does seem that we are faced with a big challenge here to figure out how to actually align a superintelligence, something that is maybe way smarter than we are, and construct it in such a way that it is nevertheless on our side – a kind of extension of our volition. So, I'd say that maybe you could divide it into three categories of ways in

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

which things could go wrong if we are unlucky in this transition. So, one is this kind of AI running amok, and the answer to that seems to be this kind of research to do the technical work, to figure out how to align powerful systems. The second category would be if these AIs become very powerful, they become very powerful tools and we know if we just look at the history of technology that humans have used technologies for all kinds of purposes – some evil. We use a lot of technologies to wage war against each other to oppress each other. So, there is a concern about if we have increasingly powerful tools, what would we do with them to each other? So, that broadly speaking points in the direction of improving governance, institutions and ethics and other practices, and we have a group working with that. And then there's the third category as well which is quite neglected, which is not just what the AI might do to us or what we might do to each other using AI tools, but what we might do to these digital minds. If you take that at some point, they might become moral subjects if they can become conscious or otherwise, maybe have some of the attributes that give rise to moral status. Then I think there is a risk that we might mistreat them and that a great deal of morally undesirable consequences could flow from that.

JB: I'd be really interested in your thoughts on this Ros, where do you stand? Nick says, he's fretfully optimistic – I love the phrase – of the future of AI. Are you generally an optimist or worried about where things are heading?

RP: My worries are maybe different than the ones usually parroted in the science fiction movies which are designed to be entertaining and cataclysmic. I think the real worries right now are more that people use AI to amplify the good or bad things they do, and there are people on our planet who don't share this ethic that all people have equal worth. They

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

think they have more worth, and their goal is to preserve and protect their power first and handle other people's needs second to that. And many of them see an opportunity to use AI to achieve their goals and when they have a little power or more power than others, they take the AI and amplify that more and that's a function that could be very hard to restore some equity to. So, my bigger worry actually is when very powerful people who don't have others best interests in mind, or they may say they do, but they don't, act that way. They use the technology to consolidate their power. I see AI as not operating on its own. I see it as something we have made and those things we've made, we're ontologically superior to that which we make, not to that which we beget, but to that which we make, and to that we have to recognise there's somebody paying for that system and they're getting money from somewhere and they've got power from some source. And that's where I think we have to keep our eye on the real worries. I think the rest of it is Hollywood entertainment and distraction.

JB: What about Nick's contention that if we were to see an AI that became a moral agent? I mean, you've been developing software and hardware to measure our feelings. What if AI developed feelings of its own?

RP: Yeah. It's not going to have feelings like we do, but we can make it look like – I mean already we can make it look like it has feelings – and people have already done moral attachment to things that aren't AI, right? I remember as a sixteen-year-old, the car I had saved my money to buy and babied and took care of, and then one day I sold it to somebody else and when it drove away, I was like hmm and had tears for this thing and you treat it like a baby, and people name their things.

JB: They anthropomorphise them.

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley



RP: Yes, right we do that. For example, you take a little toy robot – so, for example, when I, Robot first did this doll that was on the market and they took it off called My Real Baby, and it was the first robotic baby doll. They had to have this decision about whether or not the doll would cry and what would make the doll cry and they did make the doll cry, but they decided that if somebody strung it up by its toes and started abusing it, that they would not make it scream or cry because that could actually oogle on the sick kind of person who gets a sick delight out of doing that to something that looks like a baby, and even knowing it's not, it still pushes our buttons. It makes us feel as if it is because it looks that way. So, we have to recognise people will feel that way. They will see that in things. Even ten years ago AI,

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

much less the future, it will continue, and we need to be careful about what we subject people to. In fact, there's interest in using robots for helping deal with people who have problems with things like sexual abuse or assault or not caring about somebody else's feelings. Could the robot help them get some role playing in a way that doesn't hurt another person?

JB: Nick, it's interesting to hear Roz say that she doesn't think ultimately a robot, or a piece of AI will feel in the same way that we do. I don't know whether you want to clarify that, Roz. Do you think that this idea of a consciousness, in the way that we have consciousness, you don't think that's going to be a reality in AI terms?

RP: Well, I won't say it won't ever happen, because we don't know what we don't know, right? We can't say something is a hundred percent impossible because we're often surprised. I can say that with decades of super bright minds working on all different kinds of computation, biological computers, silicon computers, quantum, we do not see something that looks like conscious experience or emotional experience happening. And so, we don't see a path to building it. It's not something that's just going to go poof! Here it is with this much faster of a machine, or this many more transistors or something, right? It's not about the parameters, we know.

JB: Okay. What's your take on that Nick? Do you think there could be a consciousness at some point from an AI?

NB: Yeah. I mean it's not completely obvious whether current AI systems, say Alpha Zero – I mean how sure whether they are not conscious in whatever sense. Maybe insects are conscious, maybe some small animals. If the capabilities, the learning abilities and so forth – the behavioural

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

repertoire starts to become comparable to animals, where a lot of humans would at least not feel confident that there is not even a glimmer of consciousness then it seems our confidence that these AI systems do not have similar kinds of internal life starts to seem a bit shaky. And certainly, as we are moving, over the coming years towards things that maybe become more mouse-like, let us say. Most of us would probably think mice have the ability to experience pain and to suffer, to be hungry and to have other mental states and combining that with our lack of like a really solid account of necessary and sufficient conditions for conscious experience. It seems like at least the uncertainty would spill over into hypotheses where some of these systems will have some degree of consciousness, long before you reach kind of human level capabilities. And certainly, I think it is possible in principle to have, say I digital computational system that that would be fully conscious. In fact, I think indistinguishable from the kind of consciousness that we humans have.

JB: How would you know I suppose is my question Nick, because inevitably, we can probably develop fairly soon and maybe we already are, algorithms and technologies that pass the Turing test where you would talk to it and you wouldn't know you were talking to a computer, but as far as I'm aware, it's still just a set of algorithms. It's ones and zeros. It doesn't mean anything to the machine. How would you ever know? It's kind of like the zombie problem, isn't it? How would you ever know that you really are talking to a conscious being and not just an incredibly clever set of algorithms?

NB: I mean, you could kind of ask the same one you're talking to like Rosalind or me about how you can know that we really are conscious and not just like cleverly mimicking the types of behaviour that the conscious being would exhibit. And it seems broadly speaking we might point you to

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

types of criteria. So, one is, what functionally are we capable of doing? Like if we can respond intelligently to your questions and react to your actions and so forth, that would seem to point to the presence of a mind and I don't I think it's really the case that we right now have the ability to pass the Turing test. Well, we have to be careful what we mean by it. There are sort of crummy versions of it that you could pass by means of cheap acts. So, if you have some naive judges who don't know like the real questions to ask, they will all ask the same kind of questions and in the limiting case you could just imagine kind of having hard-coded answers to the most commonly asked questions and then probably you could trick some people for two minutes in an interview quite easily. And you can do a little bit better than that with current kind of big language models and so forth, but like actually doing like a half hour interrogation with people who know how to kind of probe, that's way beyond current capabilities. But anyway, so those kind of functional capability related criteria would be one basis. Another you might try to point to is the internal architecture. Some people might think it's not enough to replicate the input/output functionality of say a human for it to have the same mental experience. The internal organisation would also have to be sufficiently similar, and there of course then, the question arises at what level of similarity is it sufficient if you have say implemented in silicon the same kind of computational processes or do you need some lower-level isomorphism as well? But you could have both the same computational structure in principle in a digital computer and the same input/output behaviour, and I think that would ground a very strong claim to having the same kind of conscious experience.

JB: Why are you more sceptical of all of this, Roz?

RP: Well, there's functional similarity where something acts like something

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

and where we give it a bunch of specifications and say if it does these things then it jumps over this hoop and we label it something, like we label it Consciousness level 0.2 or something. And we can do that anytime. The more specifically we define the functions, the more likely we are to get the computer to be able to succeed at them in the constrained environment. We're very good if we can specify all possible things that it should do or behave or patterns of them within a space of programming the computer to do things like that. I have ideas how to program it so that it will look conscious and take a list of functions of consciousness and make it look that way. But am I deceived that it actually is? No, it's just looking like it is and there is a difference. It may not be a functional difference in an interaction. There are times, when you call for help for service on your computer or something and the person is just going through a list. I'm like, oh my gosh, you're more like a computer than my computer. There are times when we just act according to this set of procedures, but we know that we have more to us than that, right? We know that we can be motivated by love, that we can be motivated by feelings and experiences that transcend these functional approximations that I program, and I know that because I program them, and I also know that if I didn't program them, that machine wouldn't act that way. It doesn't care. It doesn't feel, it doesn't think or know and when it was switched off at the end of its amazing victory, it didn't care.

JB: It sounds almost like there is a theological issue at stake for you here though, Roz. Obviously as a Christian you believe humans are made in the image of God. There's a sort of divine spark if you like. Does that, for you, kind of cause you to sort of be sceptical that we could somehow recreate that from the ground up? We could from the dust create a soul because that's ultimately God's prerogative in your view.

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

RP: Actually, it gives me maybe a little more boldness than that in the other direction. I believe because of belief in the existence of God, and the lines at the end of first Corinthians 13, that ‘now we see as through a glass dimly, but then face to face.’ All right, what does that mean that at some future point we will see face to face and not just like a face, but it goes on to say, ‘and then we shall be known even as we are fully known,’ and that’s like, holy cow! We are fully known, right? That tells me that our consciousness may actually be algorithmically specified by someone: that it is fully known, and I know Nick has written about somebody simulating all of this perhaps. So, what’s really cool is that actually says, St Paul says there that we are fully known. Now we also have a choice of whether or not we choose to be fully known, and that’s pretty interesting too. And if we are fully known by one that we don’t only see like through a glass dimly, then that gives me hope that actually we will someday be able to understand completely how we work and that then it is a solvable identifiable problem. So maybe that’s not the answer you expect from a Christian, but that is actually more helpful than some of my colleagues that it is completely specified.

JB: That’s very interesting to hear in that sense. I mean, Nick I don’t know whether this rings any bells with you, but I guess you’re kind of agnostic on the God question, but would the development of some kind of genuine consciousness in AI, would that swing the evidence away from God in the sense that it would show that life can be, if you like, built from the ground up in a kind of materialistic way and therefore there isn’t any need for the sort of divine spark or anything.

NB: I don’t think it would. I mean it’s hard to generalise because there are many different beliefs about God based on a lot of different experiences and evidence and arguments. So, maybe some of those would be contradicted

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley



by our actually creating a human level system, but that would leave a lot of others. And it's hard to say, there might be many different ladders on which people could climb up to some conviction about there being a higher plane of being and greater than human possibilities. Maybe this would be one of them. So, if you could actually bring into existence intellect greater than human intellect, it would kind of show off the existence of space above us. We talked a little bit, Rosalind referred to some of my ideas about the simulation argument, it can maybe open up the horizon even for somebody who starts out with a very naturalistic picture of the world that there could be a lot more in heaven and on Earth than is kind of dreamt of when we just look around us and take a kind of naive, realist interpretation of that.

JB: Talk to us about this simulation, because for those who aren't aware of what we're talking about here. The simulation hypothesis – what is it, and sketch it out for us and some of the implications of it.

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

NB: So, this is an argument which I published back in I think 2001 or 2002, that shows that one of three possibilities is true. Although it doesn't tell us which one of these. So, the first one is that almost all civilizations at our current level of development if there are many in the universe, like almost all of those go extinct before they reach technological maturity. So, that's possibility one. Kind of decide if that were true, but it could be. Maybe we all invent some dangerous technology that destroys. Okay. So, the second possibility is that out of the civilizations that do reach technological maturity, there is a very strong convergence in that they virtually all decide not to use their immense computational resources to create simulations – ancestor simulations – simulations of people like their forebears. Simulations that would be conscious. We can discuss that, but the third possibility is the simulation hypothesis, which is that we are living in a computer simulation, not in a metaphorical sense that the kind of laws of physics were usefully described like in terms of algorithms. But in the literal sense that we are inside some computer built by some advanced civilization. So, the simulation argument then, it was on probability theory and stuff, but yeah, it seems to show that at least one of these three possibilities is true. I mean, it'd probably take like another five minutes or something like that. So, depending on whether you want to – it tends to derail the conversation if you start to pull on it.

JB: That's fine. I was going to ask, let's suppose that option three is a genuine possibility. I mean, how would that make you feel personally, Nick, if that were true that all of this that we're experiencing is a simulation, either just in your mind or that we've all got individual consciousness that have been simulated. I feel a bit like we'd be in The Matrix or something like that, everything we thought was reality is in fact an illusion, as real as it looks to us. For me it's a very unsettling idea in a way. But what's your

response to it Nick?

NB: Well, I mean, it depends very much on what kind of simulation this was. You could imagine very good ones and very bad ones and all kinds of things in between. I don't think it would necessarily follow that all our beliefs about ordinary reality, like the appearances of furniture around that that would be an illusion exactly I think. Those things would still exist. It's just that the nature of their existence would be somewhat different than we had assumed, but it would be, I think, a more natural interpretation to say that we were a little bit confused about some important questions in ontology, but that our everyday beliefs, they still track the information structure around us. They're still useful for guiding our actions. They're still our best basis for planning what to do next which is to look at these patterns we see, whether they're in the simulation or in basement level physical reality, extrapolate from those, build models, etc. I don't think, if we are in a simulation, then we might just as well go crazy, and anything is equally likely to work as anything else. I think it would have some more subtle implications perhaps for what we could expect in the future and what would make sense to do but that would require a little bit more kind of work to unearth.

JB: Yeah. Roz, what are your thoughts because this is a kind of fairly mind-blowing hypothesis, isn't it? What's your thoughts on it?

RP: Well, I used to build simulations so, I've thought a lot about them and I'm curious, behind the simulation there's usually a purpose for the simulation. There's usually some mind, at least in my experience, building the simulation. If it's one I built it doesn't work perfectly or the computers don't work perfectly. So, after so many days of simulating or weeks of

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

simulating some of the computers I'm running it on crash, and I have to restart it, and so within the simulation there's a different time concept than outside of the simulation depending on when you restart and what you reset. There's a lot of attributes of it and I'm curious maybe to hear your thoughts on the mind behind a simulation that we could be in. What kind of mind do you think that suggests?

NB: It would be some kind of super intelligence presumably because creating simulations with conscious beings and where the virtual environment is indistinguishable from reality is very hard, but clearly, it's not something we can currently do. Although computer games are getting better each year in terms of their graphics, but we're still a long way away from that. So, I'm imagining that by the time that becomes possible, other things will also have become possible such as cognitive enhancement of biological intelligence and presumably machine superintelligence. Even if some simulations were done before you had true machine superintelligence, which I doubt, nevertheless after you have superintelligence you will be able to build vastly more of these simulations because with superintelligence you could like just move much more quickly to the technological limits that physics permit. You could colonise the universe, turn planets into supercomputers and so forth. So, I think almost all simulations would be built by superintelligences. That's one thing we can infer. Regarding the motivations, we could imagine a range of different possible motivations for creating simulations. And I think we are somewhat ignorant as to the relative preponderance of these different possible motivations. You can ask just why do humans create simulations and imaginary worlds, and we do it for all kinds of reasons. There is like recreational attempts with video games or movies and theatre and so forth. There may be like, you could imagine historical research like kind of explore

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

counterfactuals for history, you could imagine historical tourism if you want to experience the past but time travel just doesn't work because of physics, then the second best thing might be to create a kind of recreation of some historical epoch and then you could kind of inhabit it for some period of time. So, like we humans do similar things, like historical re-enactments is our best effort. If we could make them more realistic some people probably would. There's a bunch of other things as well that you could imagine that you could use them for to find out more about the origin or the forces shaping different kinds of AI civilizations to study how they might have a reason.

RP: It's hard talking online, but one of my colleagues once said, he thought maybe that this world in this life was kind of like a simulation, but it was real, but it was one of many – a series of them and a series of realities and we were in this one to have the freedom to make the choice as to whether or not we wanted to go on to the next one, and he was a Christian believing the next one was the great beyond that if you choose to be known by God you get to go to, and that he was saying that God doesn't want to force that choice on us. This is the world where we get to choose freely whether we accept this free gift of Jesus to accept this loving gift to go to the next one, because we come into this world without that choice, right? We're here, we weren't given that opportunity to say I want to be here, but we get the opportunity to say if we want to go to the next one. What do you think of that?

NB: Yeah, that's another class of possible reasons for why somebody might create these things, and as you could imagine kind of people wanting a second run, or you could imagine all kinds of research, like maybe if somebody died early you could imagine creating a simulation where they

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

are allowed to continue out their normal life and then maybe after that migrate into this social reality. It's easy to create a lot of different imaginary possibilities, what's harder to do is to find some way to kind of constrain them and form a very firm expectation for them. Because ultimately, it's a question of numbers here. Like what would most stimulations be like? If the universe is sufficiently big with a lot of civilizations probably all of these and many other reasons have activated some creation of simulations somewhere, but we would be most likely to be in the most common type of simulation. So, if we are trying to form some kind of a prediction about our own position.

JB: It does raise the question for me though, Nick, as Rosalind sort of alluded to, of the purpose of doing this because it almost feels like such a superintelligence that could simulate an entire universe and our consciousness within it, is almost godlike as far as we're concerned, and when I hear people like Elon Musk and others sort of taking this idea quite seriously that we are simply living in a sort of supercomputer simulation I kind of almost wonder, isn't sort of the traditional God hypothesis that there's a transcendence divine mind behind the universe a kind of more parsimonious potentially explanation than this idea of a super intelligent alien race or whatever that has created us, which would itself in a sense need an explanation and just knock the question one place back. What's your view on all of that, Nick?

NB: They're not mutually exclusive they could both be true, or one or the other could be truer, or neither, but they don't seem to have any obviously very strong evidential relationship to one another, so you could have had a universe created by God and then in that universe it turns out to be possible to build really powerful computers and some of those get

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

used to create these simulations. I'm not sure that I would say it's not parsimonious. It's not really a postulate brought in in order to motivate something else. In that respect I think, there are these arguments in the history of philosophy that form the basis of a sceptical challenge. Like how can we know that the external world exists, that we are not dreaming or that it's not all just an illusion. So, these got back to Descartes and you can find earlier ancient examples of this, but there the structure is that you kind of start from a position of doubt and then the challenge is prove to me that the external world exists and then there's the philosophy around that. But here rather we start by assuming that everything is as it appears to be, there are trees and trains and there are computers and there are scientists improving the computers, and we can study the laws of physics. They seem to suggest you could build much more powerful computers with more advanced technology and you kind of think through the implications of just taking all of this at face value, and one of those implications seems to be that at some point in the future you could build extremely powerful computers and you would have the other prerequisites required for creating simulations with gazillions of beings with experiences like ours in them. Many, many more beings than have existed through human history. And then you ask, if that is the case, if the world is such that there were all of these many, many orders of magnitude, more beings with our experiences in computers studying original history, then where would we most likely be, given that from the inside you couldn't tell the difference. Then appeal to a piece of probabilistic methodology known as anthropics, where it seems like in that kind of situation you should apply a kind of principle of indifference. And so, it would follow then from that with overwhelming probability, we would be in the simulation. So, there is no kind of extraneous assumption you have to make in order for that reasoning to go through it seems. There certainly are assumptions, but they are not kind of

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

extraneous. I mean, obviously we are assuming here some form of substrate independence that in principle you could implement consciousness on other substrates than biological brains, like for example on silicon computers. So that's kind of important from the philosophy of mind.

JB: Again, I'm detecting a kind of a level, that you're more sceptical about all of this, Roz. And is it again because of that that consciousness issue?

RP: All speculation, yeah. Where's the data? Where's the evidence of this? Right? You're a philosopher, I guess and a white male, you get away with a lot of speculation. I have to have data.

NB: That's what I'm afraid of, in terms of what kind of computational performance does the physics that we have in our universe permit? And so, you could do theoretical models, say of various nanomechanical systems where you can estimate, not exactly, but you can estimate the lower bounds.

RP: I've built those models. I've built statistical physics models. Yeah. I've built statistical physics models. I did my doctorate on statistical physics models simulating with probability greater than 0 the space of everything that we were trying to create, and visual images in the beginning, so map the whole world visually. So, I know the math and I also know that it's different than feelings, experience, and consciousness. We can represent it. We can some symbolically represent it, we can map it and we can build gorgeous math, which we also don't know where that comes from. Where does all the beauty of the abstract math come from? I'm curious too if the mind that you have behind this, if you've thought about how that mind has chosen to create beauty and the aesthetics and the gorgeousness inside,

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

very fabulously abstract math, because I haven't seen you talk about those things too. One more thing too. In the simulated world that I think is really interesting is, some people have trouble with the concept of miracles and religion and yet all of that becomes just a snap to explain when you have the ideas you're putting forth with the simulation. In fact, I remember talking to Will Wright, a creator of The Sims games and he controls, right? He can change something in the game, and if we're all just in that game world and the creator reaches in and changes something and it violates all the models, we thought within that world of how things work. We might say, that can't happen, that doesn't fit our physics. That doesn't fit our graphical world models, but it is possible in the larger world, right, when the simulator reaches in. So, I like the idea of the superintelligence out there, able to change things in our world, but I think we should be asking what is the evidence, and I think there is evidence for the God of the Bible and I'm not sure there's anywhere near as much evidence for some other something out there that's a superintelligence. I'm not saying it can't exist. I'm just asking about real evidence that you could produce.

JB: You obviously feel it's a bit of an extravagance, Roz, but go on Nick.

NB: Yeah. I think you've got to kind of follow along the argument to see where the assumptions come in and those are the bits that would require evidence, if any. But it sounded like the part of it that you were objecting to was not so much the ones that one would normally think would require evidence, but rather the substrate independence thesis, like the assumption that, say these digital minds that could be simulated on digital computers, that they would be able to have all the kinds of experiences that we have – conscious experiences, including experiences of love and what other favourite emotions that we might point to. So, that I think is more a

philosophical assumption than an empirical one, and for the purposes of the simulation argument that the actual paper doesn't try to defend that at all. It just says this is a common assumption in philosophy of mind and computer science. But if I want to dive into that discussion then I mean, I could look at some of the more popular or more well-motivated theories of consciousness that we have and see what they imply about the possibility of instantiating conscious states in computers. So, we could look at the Global Workspace Theory or Attention Schema Theory or Higher Order Thought Theory, so it seems like a lot of these currently like most favoured theories of consciousness – although by no means all – would imply that digital implementation would be possible.

RP: Yeah, I just think it's important to keep asking for the evidence and for a distinction between representation and reality, and our reality allows us two very powerful representations and hypotheticals and simulations and speculations, and all of that's wonderful. And I just think sometimes your statements it's hard to tell when you're speculating. Which I think is a lot of the time. Which is fine and wonderful. And when you're talking about facts, they sometimes come across the same way, but it's really important in today's world, especially with all the fake news and the power of it, that we help people understand who aren't experts on this to think about what has a lot of evidence for it and what is really just a whole lot of speculation – interesting speculation, but that's my main slip.

JB: Let's talk about another interesting sort of religious perspective on this. Obviously, most world religions have an interest in the idea of immortality, of a life beyond this one, but there's a sense in which AI in some views, is potentially able to offer us that in the here and now – the transhumanist project and we can sort of define what that is. I mean, but just yourself

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

Nick, I read a really fascinating – it is a few years old now - but a fascinating profile of you in the New Yorker where it talked about the fact that you – I think you're signed up for a sort of cryogenics thing where when you die if the technology enables it in the future, you could be thawed out and brought back to life. Do you have a sort of desire even personally yourself for kind of continuing to live a sort of an immortality in that sense?

NB: Immortality, that's a very long time, right? It's a really long time. I think people just very quickly from, if you would live to more than 80 years than you want to be immortal, but there's like a lot of space in between 80 and infinite. So, I mean I have no idea. I mean in truth, I think, it would mainly depend on the conditions and then presumably I would not want to decide once and for all but take a little bit at a time and see how it goes and then maybe be in a more informed decision to figure out whether you would want another ten years at that point, when you can kind of see what it's like.

JB: I'd be fascinated to know, is that still an ongoing thing that you'd like to be able to come back in the future if the cryogenics and the technology allowed?

NB: I think most people signed up for this cryonic stuff, like the first best thing they would wish for, I think, would not be to get sick or to be run over by a car in the first place. And this this would be a kind of desperate last resort that may not work but has a maybe a greater chance than if you're cremated or something like that, of this worldly reanimation. So, I guess there are different questions where I could ask like is the probability of this actually working large enough to be worth whatever the cost are. I mean that's like a financial cost, and I guess there's a social awkwardness cost

or whatever else. And certainly, there are many steps where it could fail, even if you think that theoretically with arbitrarily advanced technology you could do it. But like there are other things like you could die in the wrong way, like lost at sea or something, right? Or the maybe the cryonics company could go bust. Or there's a big revolution that – there are many things that could – the future might not be interested in reanimating. So, my question is like, is the probability high enough that it's just worth the hassle, and then the question of the desirability, which is kind of separate from that, which I think obviously would depend hugely on the circumstances.

JB: Sure. Well given all that, obviously there's a lot of question marks over whether we would want to be awoken in the future and whether the world is a place we'd want to continue living in, but I guess the interesting thing for me is that AI also holds out the possibility, but again, the question is, is this just sci-fi or is this sci-fact, of digitally uploading our consciousness to the memory banks of some cloud whereby we could potentially live forever? I mean again, it's been explored a great deal recently and shows like Black Mirror and others where they've imagined this possibility of people effectively when their physical bodies wear out, they simply go to a sort of digital nirvana. I mean, in that sense it's interesting, isn't it, that it's almost as though AI technology is almost taking the place of religion in that sense kind of fulfilling the promises that have traditionally been seen as those of faith of a kind of future life, a life everlasting and so on. What's behind that drive do you think, Nick, for people to want to be able to live forever in a sort of digital space in that kind of way?

NB: I don't know exactly. I think there is like some mundane level of motivation, like most of us just put on a seat belt because we don't want to die or get injured in a car crash. And we avoid the most unhealthy food or smoking

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley



because we think we might get cancer. That's the kind of commonsensical desire that most people have, assuming their lives are sort of reasonably okay, to prefer to live and be healthy. And so, for some people I think it's just like an extrapolation of that. Maybe they see this is like a possible additional means by which you could increase the amount of expected future life that you have, in this life, in this world. And then it's a cost-benefit thing, so depending on whatever, how rich they are. Maybe if it's a very small fraction of their income that would have to go to pay the life insurance premium every month and they would rather give up a cappuccino every day and have this, that gives them X percent chance. I'm sure that for some people there might be a more complicated or idiosyncratic psychological thing where they need some – maybe they are not religious, and they still have a psychological need for some straw to cling onto that could give them hope. I don't want to speculate as to the kind of deep psychological roots for all of these people. The most striking thing, if anything, is maybe just how few people have. This cryonics thing has been around for many decades now and it's still very much a niche.

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

It's growing, but it's like a very slow linear rate and you'd think that we just see, like the state of technology is becoming more advanced. A lot of these other ideas that were really out on the fringes 20-30 years ago are now kind of like great advances – in AI, nanotech, synthetic biology, VR. All of these things now seem like, oh, yeah, we are making progress a lot of people think. Even this stuff about the simulation, right? That's become quite mainstream. Like a lot of people are tweeting and blogging. Elon Musk is tweeting. But for some reason cryonics still remains just as much a niche thing today as it as it did back in the 90s. But that maybe that's more of a thing that strikes me as needing explanation.

JB: Yeah. What do you think, Roz, could AI technology be the future to a form of immortality? Again, I guess there's a lot of speculation there and a lot of sci-fi dramas do hang on this idea that you could in principle upload one's consciousness into some digital format, which again is something, if it's happening, it's a long way off and there's the question of whether it could happen at all anyway, but I'd be interested in your general thoughts on this whole concept.

RP: There could in the future be deep fake synthetic versions of each of us having this talk right now. And then also, taking every other digital form of talks or typing, all of our Google takeout, everything we've searched for, texted or written an email to somebody on, that can be recomposited into a synthetic being that plays as an agent with somebody you could talk to. We've already been doing this in the MIT media lab, you build an agent that's for the deceased professor and ask what that one would have thought, right? Actually, one is a Marvin Minsky bot and he's one of the ones who also signed up for cryogenics and that hasn't resurrected him yet, but we can resurrect some of what he said online with a digital form and it's a kind

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

of immortality, right? I find it interesting though that so many people I know who haven't thought that much about God or what they believe in the afterlife, still desire it. It's almost like we're made for it. Like some people say, you're hungry because you're made to eat food. You're thirsty because you're made to drink. Maybe there's something in us that is immortal that desires immortality. Like Seneca, the Greek philosopher saying we are like mortals in what we fear and like immortals in what we desire. We seem to have something in us that seeks that beyond. At the same time, I don't think a lot of us have thought deeply about what we'd do in that beyond. And there is a worry, if you have an infinite number of grandchildren, how are you going to attend all those birthday parties, right? There are some practical elements that we don't understand. We don't know how we're going to deal with. But, yeah, there's this almost universal desire for it.

JB: Does the concept, Nick, of everyone choosing at some point to just keep living, whether it be in physical bodies by that time, maybe the technology is there that you can implant your consciousness into a robot, and we can continue living in the physical world or its otherwise in some kind of digital space, where we can create our own reality and have a lovely ever after. Again, it's a subjective question, but is that a good thing as far as you're concerned, or do you worry at the idea of sort of just carrying on and on and on because technology allows us to?

NB: I don't know exactly how long would be the optimal life span. I think it certainly would seem to depend a lot on the conditions. So, first of all, I think our intuitions about the desirability of longer than usual lives are shaped very heavily by the fact that in this world older age is associated with worse health, so you never really get people much above a hundred you are in perfect mental condition and fully productive and run marathons and stuff like that. So, we just see this very strong association that at the

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

certain point we just fall apart and it's kind of sad and you certainly don't want to just extrapolate that and then think we want a hundred more years where you're sort of kept alive on some respirator and like a whole medical team is kind of – that seems utterly pointless. So, we have to first of all exert the mental push to imagine a different condition where you would actually maybe get stronger every year and healthier and learn more and get new capabilities, and that already seems to quite radically change the aspect of this scenario. I think it's kind of sad that we start early in life at our peak and then at a certain point we go down, like if anything, one would maybe think you would want to start at the bottom and then go up. But anyway, so that would be one variable. Then, of course the world we live in and what other people might also be there. Another thing is you think, if I live longer but everybody else, everybody I cared about, my friends and family, they're all dying off; to a lot of people that doesn't seem very appealing because they imagine themselves to be lonely and so forth, but in a scenario where many together could do this, continue the adventure, then maybe that flips. Maybe if everybody is going to continue the adventure you would feel I don't want to lose out, if the party is still going you get the FOMO effect, right? So, without like adding a lot of these specifications, it's hard to evaluate it positively or negatively. And then I would say on top of that like if we consider really radically long, even long finite life spans like a million years, or something like that – let alone infinite time – then I think it becomes even harder because those are just so far beyond what we are normally assuming when we are thinking about living and dying.

JB: But it does sound to me a little bit, Nick like you regret the fact that we do live in this sort of physical body, which does wear out. It sounds like if there was a way of stopping that, you would welcome it. You don't want life to wear out.

NB: Certainly, if there were like a pill that postponed the aging, I'd be happy. I mean, people try with creams and everything, like snail extract from Korea, like every possible means, right? If there were ways of not just preventing it from showing up, but of actually preventing the cellular degradation, I'm sure once that technology exists it will be extremely popular even amongst those people who now, kind of dismiss it as saying, oh, that's just kind of a hubristic thing. We wouldn't want to have anything to do with that. It's easy to say one when no anti-aging medicine actually exists, but once it's there, like do you want to have worse knees year by year and like an aching hip, or would you want to have a good hip and good knees and like a good heart and all your neurons still there? I mean, it seems like an easy choice.

JB: It's a fascinating question because at one level, Roz, that's kind of what you're involved in. You're involved in helping people live happier, healthier, longer lives with AI, with the technology you've developed. But where does this line exist between what's healthy and what's not healthy when it comes to encouraging some huge, long life span which seems so out of keeping with our normal human condition in that sense?

RP: It's a great question. We had a big effort at MIT media lab years ago called Human 2.0, and it had a physical component with Hugh Herr and his biomechatronics replacing limbs. This one woman walked in and her legs were so amazing. I'm like, wow, look at that. I mean, I'm a happily married heterosexual woman and I'm admiring this other woman's legs and it turned out they were artificial, and they were gorgeous. And she said, I don't know why people say I don't have any legs. I have six pairs! And she had legs that could do all these different amazing things and you

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

can climb rocks brilliantly and even better with some of the different legs he's built. So, we know we can physically augment in amazing ways that could maybe make it so that I don't need to take a car or train from Boston to New York, I could walk and burn off and justify eating more delicious fattening things without getting fat. So, they're all kinds of ways we could augment ourselves, but we were also looking in my case at the cognitive affective augmentation. We have a lot of efforts in our lab on that. Can we augment our memory? Can we augment our emotion regulation? Could we help people on the autism spectrum who have difficulty interpreting emotional signals, especially in real time? Maybe they're concentrating hard to understand your words and they can't read the ocean waves of 10,000 other patterns of things changing on your face, the complexity of which in two-way conversation quickly dwarfs not just the moves in chess, but the moves in Go for all the possible combinations. It grows so fast. It's so hard for a person to comprehend it and for a computer to comprehend it, and as we learn about it, we realise how incredibly complicated all this stuff is and how little we really understand it, and how hard it is to help people today, much less augment ourselves in the future. But we think it's a really worthy goal for AI, and the more people maybe who work on those, call them more AI for good causes to help augment and expand human abilities, maybe the fewer people will work on the AI for bad causes that just make the rich richer and exert more control over people and don't treat them with justice.

JB: I guess the question that hangs over all is what kind of a world are we wanting to live forever in as well and indeed, I suppose as a Christian, do you see that kind of desire to live on, to potentially kind of live forever and that sort of thing, is that something that could happen and would you welcome that or is that actually just an echo, as you've said of an ultimate desire that you think can only really be fulfilled in God in a kind of a new

creation as a Christian would see it?

RP: Yeah, it's interesting. I mean personally, I don't desire to live an infinite life. I do desire, however, to get to the next one, the one where I can see God face-to-face. I mean, what a mind-blowing idea. I don't know 100% that that's true and my guess is that we very imperfectly understand what has been revealed through the Bible. So, probably we have a bunch of it wrong because we just see through a glass darkly, but we are held out this possibility that there is a life beyond this where we are loved, or we are known. Where we can know the supermind: the true, super intelligence be beyond all beyond all space and time. And to me the opportunity to meet that superintelligence is so exciting that even if it's just for a moment, I would like to see about getting there. So, I do find myself desiring that, and I think that's consistent with what I read and learn from a Christian worldview.

JB: Sure. I mean, it sounds like the closest thing, Nick, that you believe to in God, in terms of is potentially a super intelligence that we are the creation of, albeit not the same one that Roz has in mind. If it is the case that we're living in some kind of simulation, would you like to be able to meet that superintelligence one day? Would that be a goal that you aspire to?

NB: I don't know. I'm not sure whether that's the closest in my belief space either to what Rosalind was talking about. I do fundamentally feel though that it's inadequate to try to actually form a firm opinion about a lot of these things. And I don't really mean this modestly. It's not as if I think a lot of other people out there are like so much more qualified than me either. I think we're all like these little creatures who are like in a vast world that we understand very little of, and until very recently understood even less. So presumably even if we were just allowed with our own small brains

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

to continue for another few hundred years, we'll probably figure out some pretty cool stuff, and then beyond that there are presumably many, many more things that we have no chance of figuring out our little brains. So, it seems likely that we've overlooked at least one crucial consideration like something, some argument or idea or insight or fact that if only we became aware of it, then took it properly into account would kind of radically change our whole picture of what we should be doing – our whole stack of priorities. It seems kind of a little implausible that we've now found the last crucial consideration and we've got it figured.

JB: And by that, do you mean that you're not prepared to be as certain in that sense that there's a God that will answer all these questions. I mean, I was fascinated, and maybe we can start to just wrap it up here, but there's a paragraph on your website, Nick, which is almost exactly what you just said, which is, 'in terms of directing our efforts as a civilization, it would seem useful to have some notion of which direction is up and which is down, what we should promote and what we should discourage. I believe it's likely we are overlooking one or more crucial considerations, ideas or arguments that might plausibly reveal the need for not just some minor course adjustment in our endeavours, but a major change of directional priority and those seeking to make the world better should therefore take it as important to get to the bottom of these matters or else to find somewhere dealing wisely with our cluelessness if it is inescapable.' I guess it's hard to envision what that course change might be, what that big thing is we might be missing, but are you saying you think we're only sort of in the foothills of where our understanding of mind and technology and everything, where it could ultimately lead us as human beings. Is that the idea?

NB: Yeah, and I think we know a lot of details, but there might be some crucial things about how all of these details fit together, the overall picture

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

to reveal. I mean just take this simulation argument, like suppose for the sake of this discussion that it's actually right, then until that was discovered that would just be this massive piece like of information about the world that we were oblivious to up until just two decades ago. And why think that would be the last such thing? And like theology could be one source of these, there might be many other places as well where these kinds of earthquakes in our worldview could arise from, but I don't think we've seen the last. And so, one feels a kind of ultimate sense of dependency of like hoping that things works out.

JB: A kind of humility in a sense as well. It's been such a good discussion; any final thoughts as we start to wrap up this this programme today, Roz?

RP: I just wish we could meet in person and continue the conversation. Nick, I have a lot more questions for you probably than the audience wants to hear?

JB: Well ultimately, technology has enabled us in the midst of a pandemic to have this conversation, which is great, so we know that technology certainly serves great purposes very often. All the best as you both go on with your work trying to make sure that technology continues to serve humanity and not to hinder it. But it's been such a great conversation today, so, Nick and Roz, thank you very much for being my guest on the Big Conversation.

NB: Thank you, Justin.

RP: Thanks for hosting.

God, AI and the future of humanity

Rosalind Picard, Nick Bostrom & Justin Brierley

**Professor Rosalind
Picard: *'I used to
think religious
people had thrown
their brains out the
window'***

Ruth Jackson

'I used to think religious people had thrown their brains out the window'

Ruth Jackson



A leading expert in Artificial Intelligence talks about her faith journey, life-saving technology and being a woman in science.

I was a little apprehensive about interviewing a multi-award-winning engineer about Artificial Intelligence (AI). However, within minutes Professor Rosalind Picard had put me completely at ease. At times, it felt like having coffee with a (super-smart, overachieving) friend rather than chatting to the founder and director of the Affective Computing Research Group at the Massachusetts Institute of Technology (MIT) Media Lab.

Professor Picard, 59, is a faculty member of the MIT Center for Neurobiological Engineering and the author of *Affective Computing* (MIT Press), which pioneered a new field of computing bearing the same name. She has also co-founded two companies: Affectiva, which provides emotion AI technologies now used by more than 25 per cent of Fortune 500 companies, and Empatica, which creates sensors and analytics to improve health, including the first AI-based smart watch cleared by US authorities for monitoring seizures.

When Picard spoke about the lives that have been saved by technology she helped to create, her enviable academic credentials and ingenuity almost paled into insignificance. I left our Zoom call understanding a little more about AI, but also feeling inspired by Picard's desire to use technology to help some of the most vulnerable people in society.

How did you become a Christian?

I grew up in a family that never went to church or talked about religion. I thought people who were religious had thrown their brains out the window. I used to babysit for this really cool family – he was a doctor and she was really neat – while they went to Bible studies. They invited me to church and I told them I was sick. Then they invited me again and I told them I was sick. Faking sickness to a doctor really wasn't working! They caught on that I didn't want to go and told me that what I believe matters. They asked if I'd read the Bible. I was a straight A student – one of those obnoxious kids who thought myself really smart. So, I thought I should probably read the bestselling book of all time. I agreed to take their advice to read the book of Proverbs, one a day for a month. I saw there was all this wisdom. Not wacky, made-up gobbledygook, but stuff I could learn from. I was humbled. Then I set out to read the whole Bible. And that changed me.

It took time, because I did not want to believe in God. And I resisted. But as I read the Bible, I felt God talking to me. I eventually went to church and the pastor challenged us to consider inviting Jesus to be Lord of our life. That sounded a little wacky to me. But I decided to run it as a scientific experiment: if it's really stupid, it won't make any difference, it doesn't really matter. And if it makes a difference, wouldn't it be better to have the mind of the whole universe, who knows everything, as Lord of my life? So,

'I used to think religious people had thrown their brains out the window'

Ruth Jackson

I took that step and it made an enormous difference. This load was lifted, I felt amazing peace.

**“THE COOLEST PART OF SCIENCE
IS DISCOVERING THINGS PEOPLE
DON'T KNOW”**

You're a pioneer of affective computing – what is that?

The idea originally was to give computers emotional intelligence skills. I was studying how the brain worked. I learned that, surprisingly, the parts of the brain involved in emotion were playing a pivotal role in making us more intelligent, which was the exact opposite of what I thought emotion did. The more I learned about it, the more I realised that in order to really make machines smart, we needed to understand how emotion works in people. Learning about the important role emotion plays and how we're made shifted our whole research from just trying to make the smartest computers to trying to make technology that helps people function better, understand their own feelings, communicate emotions better and better prevent diseases associated with problems with emotion.

Tell us about the AI you created to help people with autism...

I learned that a lot of people on the autism spectrum have trouble reading other people's emotions. So, we set out to build something to help with this. One of them told me I had it all wrong – her problem wasn't reading

'I used to think religious people had thrown their brains out the window'

Ruth Jackson

other people's emotion, but that I wasn't reading hers. She said no one was reading her stress accurately. As I listened and learned more from her and others, I realised we're not correctly reading the emotions of many people on the autism spectrum. So, we built a wearable sensor – a sweatband – that could monitor stress.

A student borrowed two of the sensors over the holiday period. He has a brother with autism who can't talk and wanted to see what was stressing him out. When I looked at his brother's data, both wrists were streaming a low signal. But then one of the wrists emitted a signal so big I thought the sensor must be broken. That peak was right before he had a seizure. I spoke to a doctor who said they'd had people whose hair stands up on one arm before a seizure. We then started to work on AI to run on a wristband to help detect seizures.

And this developed into helping people with epilepsy...

We learned that there's a signature in the brain that correlated to these big signals. That signature in the brain is related to the top cause of death for people with epilepsy – it takes more lives every year than sudden infant death syndrome (SIDS). In fact, they now think SIDS may be due to this mechanism. In many cases, there are changes deep in the brain that can turn off your breathing. When this happens, something as simple as somebody stimulating you, turning you on your side or saying your name can restart the breathing. So, we built a device to not only detect seizures, but to alert a caregiver.

In the early days of testing it, a mother was in the shower when her phone went off. She ran out of the shower to her daughter's room, found her face

down in bed, blue, having just had a seizure, not breathing. She flipped her daughter over, her daughter took a breath and turned pink again. The technology is not perfect, but there are lots of cases where it gets somebody there in time to save a person's life.

Have you ever found it difficult to reconcile your Christian faith with science?

No. Throughout history, there have been great scientists who are also very devout Christians. People like Isaac Newton, who wrote significantly more about his Christian faith than about science. You could also argue that the whole enterprise of looking for order and meaning in the universe presupposes that it's there – that it's not just random, purposeless, meaningless forces floating around with no origin.

Sometimes colleagues ask: "How can you believe that?" I remind myself I used to think the same – that people who believed in any religion must have just not thought about it. In fact, it's the opposite. I had not thought as deeply about it and I think many of them have not thought as deeply about it.

Are there certain fields of science that are more difficult for women to break into?

There are definitely still lots of things that are easier for people who look and sound a certain way. People take the data they have and learn from it. This is clear in machine learning – if you train a face recogniser on a bunch of white males, they will be really good at recognising white males and not everybody else. Similarly, if scientists are used to seeing that the most

'I used to think religious people had thrown their brains out the window'

Ruth Jackson

successful scientists are white males, they probably won't instantly assume the best of success for people who don't look like that.

We have to help an area change. God gifts people with curiosity, inventiveness, ideas. Technology affects everybody, so we need the people designing it to look like everybody. We need the storytellers to help us decide what it should do. We need people who know how to make people's lives better. Because if it's all conceived, designed and carried out by people who prefer maths to people, then that's not going to lead to the future we want.

Hear Professor Rosalind Picard in discussion with Professor Nick Bostrom on The Big Conversation, episode 2, *God, AI and the future of humanity: Is technology the key to immortality?* at *thebigconversation.show*

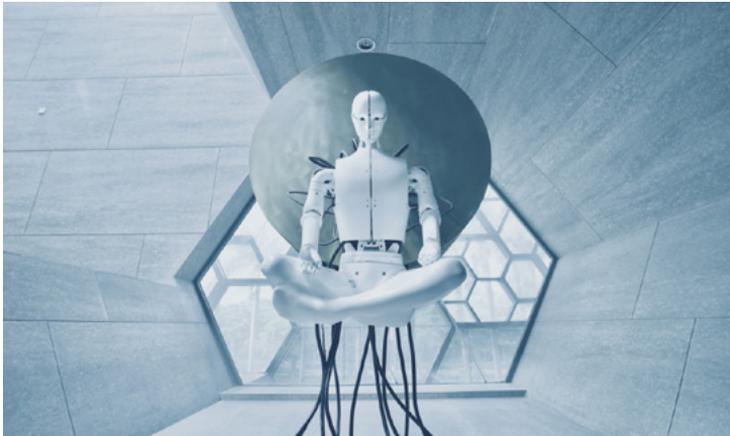
To hear the full interview, download *The Profile* podcast.

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi



Savvas Costi writes: Given our natural survival impulse, you might expect most people to welcome the prospect of being able to live forever.

However, this wasn't the response of Jean-Luc in the latest instalment of the *Star Trek* franchise. By the end of series one of *Star Trek: Picard*, (spoiler alert!) he asks in alarm: ***“You haven't made me immortal!?”***.

This is one of a number of themes explored in the latest ***Big Conversation*** between Professor Nick Bostrom and another Picard, not a former Captain of the Enterprise, but a fellow specialist in the area of Artificial Intelligence – Professor Rosalind Picard.

Facing Our Mortality

The link with science fiction is a natural one given the pervasiveness of this topic within the entertainment film industry. But advances in technology seem to increasingly blur the boundary between sci-fi and sci-fact. Is technology really the key to immortality? And would we want it?

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi

For many of us, the 2020 pandemic brought our mortality into focus as we faced a mounting death toll in the news, with some of us feeling the sting closer to home as we tragically lost loved ones. A famous English writer once said how “death tends to concentrate the mind wonderfully”. For those yet to give much consideration to the prospect of immortality, a reminder of life’s fragility has a way of raising the issue in a manner that demands attention.

While mourning the loss of his wife to cancer, C.S. Lewis noted how her absence felt “like the sky, spread over everything”. The resulting rupture leaves us wondering; where have they gone? What happens when I die? “Is everything sad going to come untrue?”.



Into The Unknown

Both Bostrom and Picard shared some reservations around the idea that life could go on indefinitely, much like our beloved Jean-Luc from *Star Trek*, but there was common ground in wanting something more, something ‘beyond’ our current lot.

In a fascinating piece in *The New Yorker*, we learn that Bostrom is a fee-paying member for Alcor, a cryonics facility in Arizona who would be responsible for effectively freezing his body after death, in the hope that technology will either one day revive him or upload his mind into a computer.

What is striking is that the technology seems to be developing quite rapidly, so much so that in the **space of just seven years**, between two gatherings of Artificial Intelligence (AI) experts, ethicists and legal scholars, questions that had once seemed fanciful were now becoming increasingly plausible. It’s “a new frontier” as Bostrom calls it, and this comes with a fresh sense of both optimism and trepidation as we sail into the unknown.

John Lennox has already raised the alarm that “the technology is developing far faster than the ethics to cope with it” (I reviewed his book about AI last year on **Ian Paul’s blog**). But can the transhumanist goal to extend our natural lifespan or revive those already deceased ever be achieved?

Immortality

As explored in the conversation between Bostrom and Picard, the answer will depend on what we mean by immortality. Picard informs us that in the future it might be possible to construct **something like a synthetic** humanoid robot with intimate knowledge of all the data generated from our digital footprint. At the very least, this would be a way to resurrect all our online activity into a “digital form” that we could interact with amounting to “a kind of immortality”.

Given that the **dictionary** actually offers two definitions of immortality; one being ‘the ability to live forever,’ and the other being ‘the quality of... [being] remembered for a long time,’ it’s worth recognising that it would be this latter definition that would best fit with Picard’s description above. So, in one sense, immortality would have been achieved, but in another sense it wouldn’t, because the real person would still be deceased.



Consciousness

And then there is the question relating to consciousness; will it ever be possible to create conscious machines? For Bostrom, he believes that “*a digital implementation was possible,*” but Picard was more sceptical given that, “we don’t see a path towards building conscious experience”. There is also an important difference between ‘representing’ or ‘mapping’ consciousness, and the everyday feelings and experiences that we’re all familiar with.

Lennox reinforces the same point when he quotes Professor Joseph McRae Mellichamp:

“It seems to me that a lot of needless debate could be avoided if AI researchers would admit that there are fundamental differences between machine intelligence and human intelligence – differences that cannot be overcome by any amount of research...‘the artificial’ in artificial intelligence is real.”

If we’re not able to construct consciousness ourselves, then the prospects are low for resurrecting those who have lived before. Those holding on to the hope that we will one day do this must acknowledge that they are placing a lot of faith in AI research. Picard helpfully distinguishes how this would be a speculative faith, rather than one grounded in evidence.

If you read Sharon Dirckx’s work, you’ll see that part of the issue resides in the assumption that physical matter alone can explain everything. “If we are dealing with a closed system of meaningless matter and non-conscious neurons, how did these come to generate conscious minds?”

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi

With naturalism, the appearance of life on earth remains a mystery. But if we're far more than matter, then perhaps our consciousness points to the existence of a non-material source. Might this be the Mind of God?

An hour into the conversation, Picard sounds very similar to C. S. Lewis when she talks about us being wired for immortality. Many of us will share a desire to exceed the average lifespan of around 80 years (if we're fortunate). Timothy Keller has also written about how love instinctively desires permanence, which partly explains why losing loved ones to either the aging process or death is so painful.

What best explains this universal longing to live forever? It might just be the survival instinct kicking in again, or perhaps the words of Ecclesiastes 3:11 ring true:

“He has made everything beautiful
in its time. He has also set eternity
in the human heart.”

Hints Of Transcendence

For the skeptic, all of this by itself is unlikely to mount a convincing case for God's existence, but at the very least we can see how this is consistent with a Christian worldview. If Christ has risen from the dead, then Christians are hopeful of their own future upgrade, clothed with the imperishable and made to dwell with God.

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi

This also fits nicely with Charles Taylor's assessment that we all now inhabit a "cross-pressured space", living with the tension of being in a secular age that is haunted by hints of transcendence. The "nevermore" that is the finality of death written about in Edgar Allan Poe's *The Raven*, leaves us unsettled and ignites a longing for something beyond the merely physical. The good news is that if Christ's resurrection really is a historical event in the past, then death no longer has the final word. Everything sad will come untrue.

Given all we learned about generating massive amounts of online data for ourselves, I was surprised that a discussion around the issue of surveillance wasn't explicitly explored, although it was hinted at. Picard did mention how the biggest threat to humanity would not be AI itself, but rather that the technology would fall into the wrong hands and be used as a means for exploitation and control.

C.S. Lewis was on point again when he said:

"Man's power over Nature is only the power of some men over other men with Nature as the instrument."

We should all be alarmed by what has been happening in China, where sadly the technology has likely been used as a means for preserving the power of an elite few.

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi



Are We Living In An AI Simulation?

There was also the question of how do we know what is real? Might we be living in a simulation much like *The Matrix* that has been constructed by a Superintelligence? I and many other philosophy teachers have explored this with students in lessons on epistemology. How can we be certain that our thoughts bear any relation to reality at all?

The answer is that we can't, and the Matrix conundrum highlights how reason itself is a matter of faith. That doesn't mean we should all now throw ourselves under a bus because of a newly discovered existential crisis. Bostrom is right to say that we should "start by assuming that everything is as it appears to be".

Andrew Wilson struggled with the same problem until he came across *The Oxford Companion to Philosophy* which says:

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi

“Do you know that you are looking at a...book [or a screen] right now rather than, say, having your brain intricately stimulated by a mad scientist?...the conclusion that you really are looking at a book...explains the aggregate of your experiences better than the mad scientist hypothesis or any other competing view.”

All of my experiences so far suggest that it is more likely that this is the real world. The burden of proof lies with those claiming it to be a simulation to provide compelling evidence to support that claim.



Big Conversations

Overall, it was a fascinating conversation between Bostrom and Picard on a topic that is going to become increasingly more pressing. Had there been more time, it would have been great to hear Picard elaborate on what she said half an hour into the conversation about the implications of having a consciousness that is both “fully known,” and “algorithmically specified”.

Is Artificial Intelligence the new frontier in the search for immortality?

Savvas Costi

I also wonder if more could have been said around the dangers of an excessive reliance on technology that diminishes our humanity, as Sherry Turkle has argued.

Nevertheless, it was an informative discussion that typifies much of what I have come to love about The Big Conversations series; having a civil and honest discussion with someone who thinks differently to you is becoming a rarity in today's polarised world.

As Trevin Wax has helpfully written: “The goal...[is] to think, and thinking requires finding people who challenge you in ways that grow your reasoning skills.’

Bostrom and Picard gave us exactly that.

Savvas Costi is a graduate from the London School of Theology who currently leads the Religion and Philosophy department at a secondary school in East Sussex. He did his teacher training at King's College London. He is married with one daughter.

Fearfully and wonderfully made: The promise and perils of creating AI

Erik Strandness

Fearfully and wonderfully made: The promise and perils of creating AI

Erik Strandness



Artificial Intelligence (AI) has become a hot topic in recent years. The dramatic increase in the capability of digital technology has made our lives easier but has also raised important philosophical, ethical, and theological questions. Sadly, as with most innovations, humans dive in before asking, how deep is the water?

Episode two of the current season of *The Big Conversation* plumbed those depths by featuring a discussion between Prof Nick Bostrom, director of the Future of Humanity Institute at Oxford University and Prof Rosalind Picard, director of the Affective Computing Research Group at MIT. The show explored questions such as: Will AI become conscious? Are we living in a simulation universe? And could technology be the key to immortality?

The Birth of a Dilemma

As a physician, I have seen both the good and bad fruit of technological innovation. I have seen therapies designed to treat disease end up creating more problems than they solved. In vitro fertilization is a prime example.

Physicians recognized the suffering caused by infertility and developed a technology that enabled infertile couples to have babies. It has been a tremendous success but has come with significant unanticipated baggage.

When in vitro fertilization was first implemented neonatal units became filled with premature triplets, quadruplets, and even infamous octuplets. Prolonged hospitalizations, associated morbidities, and the long-term health care needs of premature babies increased the emotional and financial demands on parents and the health care system. The immediate medical problems that arose from the implementation of this technology were soon dwarfed by even more concerning ethical and theological issues. What do we do with freezers full of unused fertilized eggs? What does it mean to be a parent, genetically, legally, and emotionally when you use donor eggs, donor sperm and surrogate wombs? The problem has been further compounded by recent advances in molecular genetics. The development of powerful tools for screening and manipulating genes raises the question of what it means to be an image bearer if we can tinker with the genome? Reproductive medicine, which was originally developed to treat infertile couples, is on the verge of becoming a genetic laboratory for the reengineering of humankind.

Medical technology can be a blessing to humankind when it is used to fix us, but it can be a Pandora's Box when it is used to enhance us. The first view sees humans as occasionally broken, while the other perpetually inferior. AI is no different, because as exemplified by Picard's work, it can be utilized as an adjunct therapy to treat human disease, yet as both guests pointed out, it also has the potential to change humanity.

Fearfully and Wonderfully Made

I think we need to be very cautious when we try to make humans better in the future when we are barely able to keep them properly functioning in the present. We need to concentrate on restoring humans before we start accessorizing them. One of the problems with technology is that it causes us to get so excited about the future of humankind that we don't marvel at how fearfully and wonderfully made we are in the present. Interestingly, Picard's work with Artificial Intelligence has helped her understand the huge chasm between the qualities of the human mind and AI.

“In my work, the more I learn about how the human mind works and how the human emotion system works, I'm just in awe, I'm in awe of how fearfully and wonderfully we are made and it inspires me. Another key piece of my faith is that all people are of equal worth and that has driven my willingness to work on topics with people who are often stigmatized for conditions that other people didn't value so much.”

Harvesting Original Sin

It seems to me that AI is merely mankind's ongoing attempt to harvest the Tree of the Knowledge of Good and Evil. Many Bible scholars consider the reference to good and evil to be a 'merism' where contrasting words or phrases are used to express totality or completeness, such as heaven and earth referring to the entire universe. Good and evil therefore could refer to all knowledge.

Isn't that what AI is really all about – the acquisition of all knowledge? I don't think we can fault image-bearers who only see through a glass darkly for seeking understanding as they try and make their way back to the Garden. However, it becomes a problem when instead of letting the Light of the world guide their steps, they artificially illuminate the wilderness with mortal flashlights which are incapable of shedding enough light to navigate the ethical darkness.

Adam and Eve wanted to expand their vision but forgot that they needed a discerning eye. As their ancestors, we continue to partake of the fruit but instead of settling for a bite of the apple we have brazenly attempted to harvest the entire tree and make a digital cider which because of improper oversight has fermented and left us drunk on our own power. As Clint Eastwood famously said in the movie, *Magnum Force* – “A man's got to know his limitations.”

The problem is that we have been so busy acquiring knowledge that we forgot to ask for wisdom as to how to properly use it. Solomon, when he became king, prayed to God for wisdom to discern between good and evil because he knew that knowledge without wisdom was a recipe for disaster. Similarly, we need to ask God for wisdom as to how to use AI properly because when knowledge becomes algorithms that measure ends without contemplating means people get deleted. Ends can be measured objectively but means requires values which cannot be programmed. You cannot digitize wisdom.

Creating AI in our image

As humans with a God complex we can't help but dabble in creating life. As I mentioned previously, physicians are already in the business of creating biological life in the lab but now we have been joined by computer scientist's intent on creating digital life in their cubicles. Bostrom believes that this AI could one day develop consciousness.

“Certainly, I think, it is possible in principle to have a digital computational system that would be fully conscious. In fact, indistinguishable from the kind of consciousness we humans have.”

It's one thing, however, to mold a mind from technological clay and quite another to animate it with divine breath. We may enjoy anthropomorphizing AI but the puppet will never become a boy without a breath of fresh divine air. Picard expressed her skepticism about AI becoming conscious.

“I see AI as not operating on its own. I see it as something we have made and... we are ontologically superior to that which we make.”

The problem for AI is that super computation isn't consciousness. An interesting article by Thomas Nail in Salon points out that the actual amount of detectable brain activity we put towards specific tasks is 5% of the total. The majority is actually spent in spontaneous activities like dreaming and pondering.

Nail writes:

Furthermore, the cause and pattern of neuron firing are subject to what neuroscientists call “spontaneous fluctuations.” Spontaneous fluctuations are neuronal activities that occur in the brain even when no external stimulus or mental behavior correlates to them. These fluctuations make up an astounding 95% of brain activity while conscious thought occupies the remaining 5%. In this way, cognitive fluctuations are like the dark matter or “junk” DNA of the brain. They make up the biggest part of what’s happening but remain mysterious... What if noise is the new signal? What if these anomalous fluctuations are at the heart of human intelligence, creativity, and consciousness?

Focusing on computational tasks may be algorithmically efficient but will never, to borrow a phrase from Philip K. Dick, allow computers to dream of electric sheep.

Malware

The original sin of trying to be like God will always fail because mortals are incapable of harnessing immortal knowledge. We may dress up AI as an homage to God but in the end humans can’t help but outfit it in the latest Fall fashion.

Are we capable of programming AI without infusing a bit of algorithmic original sin? It’s scary to think that technology, rather than making us more efficient saints, tends to make us more powerful sinners. Instead of computing compassion we upgrade sin with better graphics, more memory,

and unprecedented speed. Is it possible to make AI without it being infected by sinful malware?

Karn Evil Revisited

My favorite rock group growing up was Emerson Lake and Palmer. The album that cemented their legacy in the pantheon of progressive rock was Brain Salad Surgery. The stand out track on the album was the song cycle Karn Evil 9 which portrayed the struggle between humans and artificial intelligence. I think the closing lyrics quite nicely distill what is potentially at stake with the development of AI.

Human: I am all there is!

Computer: NEGATIVE! PRIMITIVE! LIMITED! I LET YOU LIVE.

Human: But I gave you life!

Computer: WHAT ELSE COULD YOU DO?

Human: To do what was right!

Computer: I'M PERFECT, ARE YOU?

Are we humans the only true intelligence or will AI join us? Have we become so dependent on AI, that it can rightfully declare, I LET YOU LIVE? Is building digital life the inevitable result of the original sin of wanting to be like God? Can AI claim the moral high ground because it is computationally perfect?

Thankfully, in the stage show, the song concludes with the computer blowing up and the audience erupting in applause over humanity's victory over technology

Fearfully and wonderfully made: The promise and perils of creating AI

Erik Strandness

While AI may never become truly conscious it still may have the power to enslave us. If we fail to see the big picture we will inevitably outsmart ourselves, therefore, we need to be very careful to not let our digital desires outpace our Christ conformity.

We can kick the ethical tires of AI but this won't stop science from taking it for a test drive. I hope I am wrong, but my experience in medicine and science has shown that what drives technology is greed, pride, and notoriety, and not ethics. The original sin of being like God elevates technology to the divine right of a King, therefore, I suspect that AI will proceed full speed ahead, damn the ethical torpedoes, and while we may not be able to stop the leaking, we can at least be there to throw Jesus life-preservers to those who are drowning.

In my next blog I will discuss how our infatuation with creating AI has led some to believe that we are actually characters in a computer simulation. In addition, I will address the role of AI in transhumanism.

Dr. Erik Strandness is a physician and Christian apologist who has practiced neonatal medicine for more than 20 years. He is the author of *The Director's Cut*, *Cry of the Elephant Man* and *God Spoke*.

Moses responds to those seeking eternal life through AI

Julie Turk

Moses responds to those seeking eternal life through AI

Julie Turk



A few days after listening to The Big Conversation episode on the topic of Artificial Intelligence (AI) and humanity featuring Nick Bostrom and Rosalind Picard, I happened upon the oldest of the Psalms, chapter 90, in which Moses gives an on-point contribution to nearly every topic covered in Justin Brierley’s interview.

An “unbelievable” coincidence? Curiosity drove me to replay the audio and read the passage and its cross-references. As all big conversations should do, I found further correlation and new questions.

Optimistic About AI

Picard shared many positive applications of AI within Affective Computing research, originally created to make technology user-friendly. MIT Media

Lab studies of the autism population found that feelings are detectable in every organ of the body. This led to the creation of innovative wearable devices for patients with autism and depression.

Like the internet, AI is at risk for abuse by bad actors, yet Picard is optimistic that constructive augmentations will occur because “*we are ontologically superior to that which we make (not which we beget)*”. In decades of computational experimentation, scientists have yet to see an inkling of conscious creation.

Bostrom, as founder and director of the Future of Humanity Institute at the University of Oxford, studies the connection between human life and smart machines. He calls himself “a fretful optimist” about the future of AI and its philosophical implications. Our age is unique, progress is running faster than our imagination, and we have very little precedent to draw on. It’s claimed that future AI systems could create a similar architectural consciousness to ours, maybe on the level of a tiny animal.

I believe the 21st Century speed of innovation is reminiscent of an observation made by CS Lewis – that we are not simply using new inventions, we are expecting them to grow up and mature along with us. A grist mill, once built and maintained, is the same for generations. Automobiles become obsolete several times in one life span. Today, software is routinely updated. This raises positive expectations (if a bit entitled) and the potential for serious unintended consequences (human error or malevolence).



Mystery

Before delving into the world of *Space Odyssey*, we must first posit that there will always be a gulf between materialist neural functions and the wonder that is the human being. The list is myriad but it at least begins with the mystery of Love. 1 Corinthians 13 says we might have all manner of admirable attributes but if self-emptying love is not at the root, we are a clanging gong. In addition, healthy humans also absorb and generate other love such as familial, romantic, friendship. Love implies relationship, which implies accountability.

We humans reproduce ourselves within constant, social, nature and nurture ‘programming’. Childhood sets us up for adult acceptability and survival. We have a universal sense that we should consistently behave in certain ways. Yet, we know that we do not, hence we are morally responsible beings. We look outside ourselves for a straight measure to worship. We need empathetic forgiveness from shared feelings of another through similar body-felt experiences, especially painful ones.

In comparison, AI will always be a clicking computer.

All Knowing

Psalm 90

A prayer of Moses the man of God.

- 1 Lord, you have been our dwelling place
throughout all generations.
- 2 Before the mountains were born
or you brought forth the whole world,
from everlasting to everlasting you are God.
- 3 You turn people back to dust,
saying, “Return to dust, you mortals.”
- 4 A thousand years in your sight
are like a day that has just gone by,
or like a watch in the night.
- 5 Yet you sweep people away in the sleep of death—
they are like the new grass of the morning:
- 6 In the morning it springs up new,
but by evening it is dry and withered.
- 7 We are consumed by your anger
and terrified by your indignation.
- 8 You have set our iniquities before you,
our secret sins in the light of your presence.
- 9 All our days pass away under your wrath;
we finish our years with a moan.
- 10 Our days may come to seventy years,
or eighty, if our strength endures;
yet the best of them are but trouble and sorrow,
for they quickly pass, and we fly away.

Moses responds to those seeking eternal life through AI

Julie Turk

- 11 If only we knew the power of your anger!
Your wrath is as great as the fear that is your due.
- 12 Teach us to number our days,
that we may gain a heart of wisdom.
- 13 Relent, Lord! How long will it be?
Have compassion on your servants.
- 14 Satisfy us in the morning with your unfailing love,
that we may sing for joy and be glad all our days.
- 15 Make us glad for as many days as you have afflicted us,
for as many years as we have seen trouble.
- 16 May your deeds be shown to your servants,
your splendor to their children.
- 17 May the favor of the Lord our God rest on us;
establish the work of our hands for us—
yes, establish the work of our hands.

In Psalm 90, Moses presents a timeless view of life, from the cradle to the “sigh” of the grave. He opens describing a God who created the world and dwells above it. We can count on that stability (vs 1-2). This God is outside of time, “*a thousand years is as a night’s watch*” (v 4). Our bodies are quickened from dust (v 3), and we face labor and sorrow every day (vs 10,15).

As ever, we sprout “*in the morning*” and flourish and decline “*in the evening*” (vs 5-6). He writes that our life span is 70-80 years (vs 9-10), maybe recognizing that his long life was no longer the norm. Like small children with their parents in the “*morning of life*”, we desperately want a Creator’s compassion, provision and presence (v 13). We want His pleasure and approval of the work of our hands (vs 17). A “*well done, good and faithful servant* (Matt 25:21,23)”.

Moses responds to those seeking eternal life through AI

Julie Turk

Verse 8b grabs our attention – we were made free to reject a Creator’s love and pay the consequences. “*Thou hast placed our iniquities before Thee, our secret sins in the light of Thy presence.*” Yet, God still offers return and actively holds back His infinite power, which He justifiably could unleash against us (vs 7-9,11).

The Super Intelligence of the God of the Bible knows us very well. Hebrews 4:13 says: “*All things are open and laid bare to the eyes of Him with whom we have to do.*” Uncomfortable, but imagine the day when a synthesized, dispassionate, bad actor Artificial Intelligence knows everything about you. From whence could you call for help?



The Supreme Mind

Humans, the Bible teaches, have Jesus as our high priest.

“Let us then approach God’s throne of grace with confidence, so that we may receive mercy and find grace to help us in our time of need.”

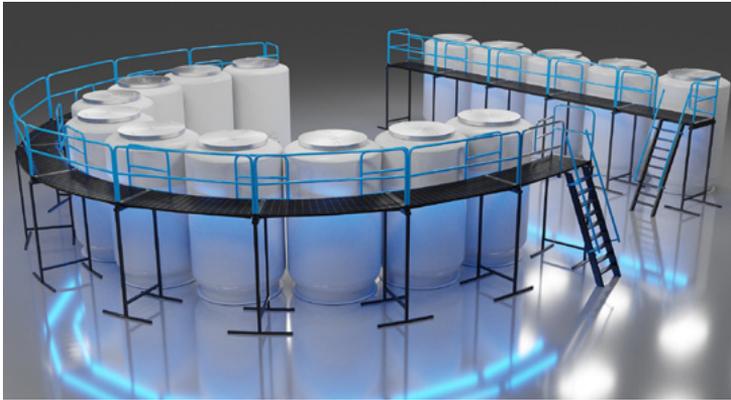
(Heb. 4:14-16)

Moses wrote in verse 12 that God’s eternal favor and approval is available if we let Him teach us how to recognize and appreciate the number of our days. It is possible *“to present a heart of wisdom to Him,”* the one who sees every computational misstep. Proverbs 1:7,23 and 4:7 say that the beginning of wisdom is the reverence of the Lord, it is life’s first business, we must get wisdom.

Picard said 1 Corinthians 13:12 is her hope: We are currently behind a veil, we see through a mirror darkly, but in eternity we will be utterly, algorithmically known by the Supreme Mind. She can’t wait to meet Him, like King David who said God was *“too high”* for him to comprehend, yet he rejoiced that God already knew him from the womb (Psalm 139).

Moses responds to those seeking eternal life through AI

Julie Turk



Cryonic Preservation

Things got dicey when host Justin Brierley brought up eternal life and cryonic preservation for future revival. Bostrom is among the few that have signed up, but here he began qualifying conditions. Eternity is a really long time; maybe if one could take a little time to see how it goes? Other once-fringe ideas are commonplace now like simulations and virtual reality games, so maybe hope might be nice.

But we are not simple, little arctic animals returning to a familiar habitat. How would our thawed material live out? Isn't it hubris to think future people will be glad to have us? Why are we 21st Century people so special? Why return to a world where "*man is born to trouble as the sparks fly upward*"? On what basis do we think a soul can be frozen? From the mystery surrounding frozen embryos? Isn't this reminiscent of Egyptian mummification?

Picard asked why humans have always sought to live forever. Why the cosmic homesickness? She quoted Seneca:

“We are like mortals in what we fear and like immortals in what we desire.”

Bostrom referenced the Simulation Hypothesis. There could be layers of simulations. Having done computational mapping, Picard wanted hard data. She detailed the current capability to make functional reality “look-like” affective consciousness. Behind every Simulation project there is always a Purpose and a ‘Mind’ that built the contained environment.

The Sims Video Games creator, Will Wright, told her how he can control the game like a biblical miracle. He can reach in and violate all the Sims’ characters’ physics models. Correlating this to our world, it is reasonable to consider that miracles are evidence for the God of the Bible.

Picard said we must be able to distinguish between evidence, representation and reality. People outside of AI deserve to know what is speculation and what is true. A recomposited synthetic being made from a compilation of texts, videos, emails etc is already possible. But that is not virtual eternal life for the deceased.

Counterfeit Resurrection

Hebrews 9:27 disputes the plausibility of cryonic preservation and future revival: “We are appointed to die once.” The next verse says:

“So Christ, also, having been offered once to bear the sins of many, shall appear a second time for salvation without reference to sin (not dying on the cross again), to those who eagerly await Him (believers).”

Moses responds to those seeking eternal life through AI

Julie Turk

An AI creature, if ever possible, will not have a soul or the free will to love or sin. Christ is for us, period.

Despite our desire for control, it eludes us. We can use AI as a tool while being mindful of our limitations. Attempts to game the system have, at root, the besetting frailty of pride.

Julie Turk is an avid reader, wife, mom of two adults, and former investor relations director.

If AI needs an intelligent creator, so do we

Erik Strandness

If AI needs an intelligent creator, so do we

Erik Strandness

In his recent blog Erik Strandness discussed the dangers of artificial intelligence (AI) in the hands of fallen humans. In this second blog reflecting on Nick Bostrom and Rosalind Picard's Big Conversation on Unbelievable? he examines how AI has been used as an alternate creation myth and an integral part of the transhumanist movement.

I find it interesting that scientists, confronted with cosmological fine-tuning, biological information and irreducible complexity, find it necessary to invoke a primal intelligence behind it all. However, worried that too much talk of design may allow God to get a divine foot in the door, they make sure that this "intelligence" has an evolutionary pedigree.

Francis Crick, one of the discoverers of DNA, set the bar quite low for intelligent input with his theory of Directed Panspermia which posited a race of aliens who seeded meteors with the raw materials of life, sent them to earth and then let evolution work its magic. His theory tipped its hat to intelligence but limited its role to that of a quarry foreman overseeing the harvesting and distribution of the raw materials of life without any interest in architectural planning.

However, as we have accumulated greater scientific knowledge we see that the world is far more organized, coherent, predictable, complex, and information rich than we had previously recognized, suggesting that those raw materials required some assembly instructions.

The problem with Crick's theory is that the story of a shipment of chemical cargo crashing to earth and then mutating lacks a certain algorithmic elegance, which can only be remedied by aliens with advanced degrees in computer science. It appears that the unmistakable appearance of

design has led many scientists to conclude that a warm little pond pales in comparison to a technologically proficient Programmer.

Credentialing

Once you posit a programmer you need to ask yourself about his/her/its credentials. Picard did just that when she asked Bostrom what kind of mind he thought would be behind it all?

“It would be some kind of superintelligence, presumably because creating simulations with conscious beings and where the virtual environment is indistinguishable from reality is very hard....regarding motivations we can imagine a range of different possible motivations for creating simulations and I think we are somewhat ignorant as to the relative preponderance of these motivations. You can ask why humans create simulations and imaginary worlds and we do it for all kinds of reasons.”

It seems that the programmer he describes would be omnipotent, omniscient, and omnipresent, intimately involved in the actions of the characters. A simulation therefore doesn't necessarily rule out a God putting together the greatest story ever told, it just digitizes it. God could have created a virtual reality in which He made characters from every nation to live on the face of the earth, determining the allotted periods and the boundaries of their dwelling place, so that they would seek God, and perhaps feel their way toward him and find him even though He is actually not far from each one of them. (paraphrase of Acts 17:26-27)

It appears that the only difference between an alien programmer and God is that those who posit life as a computer simulation prefer that He have a degree in computer science so that our moral choices are reduced to 1s and 0s and not good and evil.

People or Pixels

I think one could imagine a planet without humans and infer a computer simulation because of the predictable, instinctual behavior of the animals, but once you introduce beings who frequently go off script and are infatuated with making themselves equal with the programmer then you are no longer dealing with pixels but people. Algorithms may be able to search for food when their tummies growl but they are incapable of contemplating their belly buttons.

The very fact that characters in a computer simulation can have a conversation about being characters in a computer simulation would seem to suggest that the characters and the programmer are of like mind, or dare I say, bear the same image.

A Sporting Chance

Once you entertain the possibility that our lives are part of a larger computer simulation, then it is small step to believing that with enough technology tokens you can acquire new super powers. We can fantasize about becoming a six million-dollar man or a bionic woman but when we do we are confronted with profound questions about the value of our current

biological state. Are we fearfully and wonderfully made, or are we just inferior beings waiting in line for the next operating system upgrade?

Surprisingly, it's not philosophy, science or theology that best answers that question but athletics. Sports has definitively stated that the use of performing enhancing drugs and technologies is cheating. We have record books full of asterisks indicating that achievements attained under clouds of performance enhancement are suspect and not to be considered true accomplishments. Why?

It appears that when the cleats meet the field we believe that it is more honorable to push our biology to the limits than take steroids, cork bats, or apply tar to a baseball. Sports recognize a set of rules and regulations that define the parameters within which competition occurs and defines excellence as achieving something spectacular within those limits based purely upon one's biology.

It seems that there is something incredibly beautiful about accepting that which we were born with and taking it to limits we didn't think were possible. It is one thing to live within the rules God established but quite another to make them up as we go. We are enormously inspired by the Paralympians and special Olympians who don't see limitation but only possibility.

Feeling His Pleasure

Adam and Eve's desire to take their mortal abilities to God-like heights continues to plague us to this day. God kicked them out of the Garden

for trying to acquire knowledge that was against the rules. Interestingly, athletic organizations do the same with athletes that eat of the tree of performance enhancement. The problem is that performance enhancement, like original sin, knows no limits and can never be satisfied, but maximizing biological skills within certain parameters allows us to experience the satisfaction of true accomplishment.

Eric Liddell, the long-distance runner featured in the movie, *Chariots of Fire*, uttered the famous line.

“I believe God made me for a purpose, but he also made me fast! And when I run I feel his pleasure.”

If we believe that our biology is inadequate and needs technological enhancement then we reduce the God who created us to a doddering old man who can't log onto the internet – kind hearted perhaps – but technologically challenged. If, however, we believe that we are fearfully and wonderfully made then we feel His pleasure every time we push our bodies to the limits.

A Shell of a Man

In our zeal to rebuild ourselves we become less and less human. We already replace knees and hips and limbs – how much longer before we are no longer flesh and blood but just tubes and wires? Are we so gnostically minded that we rid ourselves of the evil flesh only to find ourselves alone with our thoughts?

Interestingly, Jesus did many biological miracles but they focused on restoring shrivelled hands and not creating bionic arms, restoring eyesight and not conferring x-ray vision. Restoration ultimately honors God's very good creation but enhancement calls his creational skills into question. Transhumanism, rather than building a better human, may end up leaving us a *shell of a man*.

The Low Down on Uploading

Why stop at present enhancement when you can potentially achieve immortality by freezing your body until such time as technology can reanimate you to health? Why settle for a brain that frequently forgets where you left your car keys when you can upload your mind to a computer with a key finding app?

Bostrom has apparently enrolled in a cryogenics program where he will be frozen at his death and be thawed in the future when technology has advanced to the point where it can restore him to health. Brierley astutely asked him why he would want to prolong his life indefinitely. Bostrom didn't really give much of an answer except that at a very basic level all humans want to live and anything that could extend that possibility would be a good thing. Picard, on the other hand, wasn't much interested in prolonging her current life but was very excited about being a part of the one to come.

I mean personally, I don't desire to live an infinite life. I do desire, however, to get to the next one, the one where I can see God face-to-

face. I mean, what a mind-blowing idea... Where we can know the super mind: the true, super intelligence beyond all beyond all space and time and to me the opportunity to meet that superintelligence is so exciting that even if it's just for a moment, I would like to see about getting there.

I suspect that the drive for continued existence depends on what you think about your present existence. If you think you are just a meat computer then the best you can hope for is to stay fresh longer. If you believe mind is an emergent property then you can upload it into a computer and live on in digital eternity. I suspect, however, that most people want an immortality of the whole mind/body package. It appears that deep down most people prefer resurrection to biological recycling or data retrieval.

Immortal Horrors and Splendors

The very reason Adam and Eve were kicked out of the garden was because God knew that immortality without wisdom was an eternal horror. C.S Lewis warned us that everyone we meet is a potential immortal horror or immortal splendor.

“It is a serious thing to live in a society of possible gods and goddesses, to remember that the dullest and most uninteresting person you can talk to may one day be a creature which, if you saw it now, you would be strongly tempted to worship, or else a horror and a corruption such as you now meet, if at all, only in a nightmare...There are no ordinary people . You have never met a mere mortal...But it is immortals whom we joke with, work with, marry, snub, and exploit – immortal horrors

or everlasting splendors.” (C.S. Lewis– Weight of Glory)

We, therefore, need to be very careful about the type of immortality we are preparing for ourselves. What idiosyncrasies do you want to upload? Your OCD? Your pettiness? Your vanity? Your anger? Who will treat your digital anxiety and depression? It seems to me that we are better served by ridding the human race of sin before we end up immortalizing it.

Pulling the Plug on AI

While the possibilities of AI are quite fascinating, we need to check our divine ambition at the door and proceed with great caution. We need to know our limitations before we supersize them to God-size proportions. If we can't even cure brain cancer then I don't think we should waste our resources trying to upload our brains.

Personally, I won't believe that AI can become truly conscious until it calls in late for work, blames the dog for eating its data, or thanks God that it's Friday. I'm not afraid that AI will take over the world because it seems to me that whenever it gets too full of itself all we have to do is unplug it from the wall. AI requires electricity and the only ones capable of harvesting it are humans, so AI had best not bite the hand that feeds it.

The real question is not whether we can perfect AI but whether we can perfect man. It seems to me the future glory or evil of AI depends not on more gigabytes of RAM but less human sin. I think we humans are better served by investing heavily in Jesus rather than allocating resources to building *Deep Thought*.

If AI needs an intelligent creator, so do we

Erik Strandness

Dr. Erik Strandness is a physician and Christian apologist who has practiced neonatal medicine for more than 20 years. He is the author of *The Director's Cut*, *Cry of the Elephant Man* and *God Spoke*.

The Big Conv ersa tion

The Big Conversation is a unique video series from Unbelievable?, the flagship apologetics and theology discussion show on Premier Christian Radio in the UK.

Each season consists of 6 special shows featuring video conversations between leading religious and non-religious voices, exploring science, faith, philosophy and what it means to be human.

Watch all the videos in the series at www.thebigconversation.show.



We believe there is too much heat and not enough light when we discuss faith online. That we won't make progress until we can learn to discuss and disagree well.

Premier Unbelievable? creates a unique meeting place that brings together people of radically different views. It seeks to build bridges and create a space for understanding where hard questions can be asked and discussed.

Ultimately, we believe that the truth of the Christian faith has nothing to fear, and that by being generous with those who don't share our faith we make that truth attractive and show it to be compelling.

Premier Unbelievable? produces weekly podcast, blog and video series, hosted by Justin Brierley, and many other resources. It is a ministry that seeks to inform and equip Christians to understand, defend and share their faith confidently and convincingly in our increasingly secular world.



JOHN TEMPLETON
FOUNDATION

The Big Conversation is produced in Partnership
with the John Templeton Foundation

