AI and You

Transcript

Guest: Robbie Stamp, part 1

Episode 118

First Aired: Monday, September 19, 2022

Hello, and welcome to episode 118! My guest today is Robbie Stamp, calling from the United Kingdom, whom I met in Azeem Azhar's Exponential Do community when Robbie was facilitating a discussion about AI and sentience – three points for guessing what the trigger for that was. He is CEO of Bioss International, a global consultancy, focused on decision making in conditions of complexity and is Chairman and Founder of HappenedHere, a soon to launch start-up "Podcast, Maps and Apps" History Platform, co-founded with Stephen Fry and Joanna Lumley and also Chairman of Not Panicking Ltd, which owns the rights to h2g2.com the website based on Douglas Adam's Hitchhiker's Guide to the Galaxy uncertainty. In 1995, Robbie was a Founder of The Digital Village with the late great Douglas Adams, author of the *Hitchhiker's Guide to the Galaxy* and was an Executive Producer on the Disney movie in 2005.

And that's only a sample of his bio. Prepare for a wild ride as two synthesists go at it in a discussion ranging from what makes AI so revolutionary to how much agency we have in a world of AI making decisions about our lives to the sentience of Marvin the Paranoid Android. Here is Robbie Stamp.

Robbie Stamp. Welcome to Artificial Intelligence and You.

So, thanks very much. Thank you very much for having me on the show. It's fantastic. I'm really looking forward to this.

So, your background has so much in it, maybe you can give us a brief tour of how you got to where you are today that illuminates for us why you have this deep interest in AI, its impact on us, our culture, our companies and the philosophy surrounding that.

Well, it's always interesting to think where to start with that answer and I suppose maybe we might come back to sort of some thinking about early childhood memories and the nature of human embodiment, but my undergraduate degree - and that makes it sound like I have postgraduate degrees, I don't - was history. And I've continued to love history, all my life. Why did people do what they do? Why did they act the way they acted? What forces were bearing down on them, which made them act the way that they did individual pathologies, wider pressures and forces and in fact, one of the things I'm doing at the moment is reading Sir Isaiah Berlin's very famous analysis of Tolstoy's views of history, *The Hedgehog and the Fox*, thinking about the nature of the contradictions in Tolstoy's own mind, about the personal and the large scale and those contradictions in the way he wrote and thought. So, I've always been interested in history. I worked in TV; I made documentaries I made early climate change documentaries. I have worked in and around the movie industry; I helped to executive produce the Disney production of *Hitchhiker's Guide to the Galaxy*. The movie was based on Douglas' books, we may talk a lot more about Douglas because I knew Douglas very well and had a company with

him. I also got very involved in my mother's company, the company that my mother founded, which is an organizational consultancy, which is deeply interested in the nature of human judgement and decision making, particularly in complex environments. And I remember I was flying down to Australia to give a talk on risk, the way organizations think about risk, and I read Kevin Kelly's book, *The Inevitable* and it was one of those genuine Damascene moments, which one has in a life, there are lots of them. But this really was. I read it with this mounting sense of excitement and change my talk, gave a talk on AI instead. And it was just at the beginning of my journey, and I remember somebody saying at the breakout tables afterwards and who wants to talk about AI? And quite a lot of people join that table. Somebody saying, there'll be an AI when it tells me it's gone fishing. And at the time, I was a bit "errr". But actually, the more I thought about it, I thought how interesting that is. So, what I really like to do is to try and think about the nature of human embodiment, and how we relate how we relate to our environments, and the nature of consciousness, all those big questions that farm far better brains than mine have pondered for millennia and then to think about how is AI, or how are AI and data manifest in the world? They are manifest, they exist, how do we exist in relationship? And so the historian has an historian, as a creative and a storyteller, as somebody who's been involved in thinking about governments, judgement decision making organizations, they kind of had a confluence, I think one of the things I've enjoyed so much about AI, and the AI community is how it's a bit like the Renaissance, because so many people are interested poets. It's not just left to the data scientists, poets, musicians, philosophers, they, because it touches so deeply on who we are and so that's a long answer. But that's how I got to be so interested.

What is it about AI? If I'm reductionist, and I say, well, that's an extension of business analytics and statistics and mathematics, what is it about this oversized calculator that spurs us to wax so eloquently about the far reaches of philosophy as you've just been demonstrating? Is it because of the name? Is it because artificial intelligence implies something that might be beyond its scope right now? Are we overreaching, are we looking too far ahead? Or is there something there that is foundational and revolutionary?

Well, it's a really great question and indeed, it makes me think of I'm on the British Standards Institute National Standing Committee on AI, which for the hitchhiker fans out there is incidentally, standing committee 42, SC 42. And that's not an accident. Maybe I can tell you why that's not an accident later. When I went to my very first meeting, the chair, this lovely man, Peter F. Brown, who has been the AI guy for the EU in Washington, he just posed this really good question. We were working on AI and board governance and I just liked this question, its simplicity. Does the governance of AI systems feel different from the governance of traditional IT systems? And if it does, why does it? So, in a way, that's another way of asking your good questions? Why all this worriting? And I think that that I'm interested because let's take something, which is quite contentious, and lots and lots of people would disagree. Let's use the let's use the phrase, which a lot of people talk about, which is AI is just a tool. It's only a tool. Now, if only I were to describe to the proverbial alien, there's a word called "tool" and we use it to describe a hammer and we use it to describe a global algorithmic trading system. I'm not sure how much I would have told said alien usefully about either object, in terms of how it sits in relationship to us and I think that's why I'm interested in sort of moving beyond the, the sort of

the data science, grr, its only maths, its only probabilities, its only things being fitted to statistical curves and look up tables, grr its only that to the it'll come and repurpose our atoms to build star bridges to the galaxy. So, you come back to that tool. If, as far as I know, no hammer picked itself up off a table and chose to start banging a nail in to a wall. If you or I were trading stocks in a in a stockbroker and what you and I would do previously we would analyze data, and we would then decide to go long, go short on various shares and at the bank and stockbrokers, I wouldn't have to come back and say, Peter, boss, what do you think? I would have a discretionary limit and it may be large or may be small, but I would have a form of agency where having done that data processing, I would then commit that data processing to the outside world in the form of buying or selling shares. If an AI system does the same thing at speeds unimaginable to human beings, it strikes me that not only can we use the word *agency* to describe what it's doing, a form of agency, but from a governance perspective, we should, because if you don't want to use that word, you've got to use another one word to describe what else it is doing. So, it's not just that it's only maths, yeah. But it's maths in relationship to a complex organization to the global economy, to you and me, sometimes in terms of what's going to happen in terms of interest rates. And within a boundary, it acts and I think that's the key thing, is what is the nature of its acting in relationship to us that is different from a hammer. Back to my point about a tool being a slightly inadequate way of describing two things which are in relationship to us without quite a lot of further elucidation. So, that would be an add on. So, just a starter answer to why it's worth inquiring how any given AI system is manifest and manifest in relationship, and that I think, is a slightly different lens to bring to it.

And of course, at the other end of the reductionist spectrum, we can say that electricity is just a tool and fire is just a tool and human beings are just neurons and axons and ATP. There was something in your answer that I thought that was key: the value that the human trader, if that was the term for what you were doing there, provides, you imputed was lying in the agency to go off and then my make decisions and take actions that weren't part of the original remit. And if that's the case, and that would increase the value of AI, then that implies some sort of governance and careful shepherding that needs to be done around AI, in that we want it to be able to go freelance, if you will, but we obviously want to throw some boundaries around that. To what extent is it a good idea to make AI that can get creative?

I think that as humans, we act within boundaries of discretionary boundaries. We stay with this thought experiment that the trader may well have set by her organization, a boundary, you can trade 100 million at a day without coming back and asking Peter, or 50 million, but 500 million 500 billion would be a big trade. So, but you can do that within a discretionary boundary that set. So, I think that the one of those things, therefore, to think about is the precision of the way we describe what any given AI system in its particular complex adaptive system neck of the woods is doing. And what are the boundary spaces we put around it, so we could put the same boundary spaces around it. It can't trade more than x 100 million in a 24-hour period, without clicking something where somebody has a look and goes, I think that's enough, or the numbers are probably bigger than that. But we can set those discretionary limits and I think, depending how big the risk of the potential for harm in any given context is, the review mechanisms need to be

more or less tight. And of course we've all been, quite rightly, scaring ourselves with stories where AI has acted without those proper oversights, whether it was probation algorithms, or the new one that I've been reading about, about, I won't necessarily name it. But it's an algorithm used to detect opioid addiction, likely, and it's been used in America extensively and *Wired*, I think have run a fascinating piece on it, where a young woman with endometriosis goes into hospital in real pain, the algorithm is run and she's booted out of hospital. And the so called knowing of her the whole issue of the ethics of inference data, these fractured epistemes, so called knowings of us that are built up out of which have real causal effects on our lives. It was a pet medication that she was buying and that ended up on her score and that was a really good example of something being deployed without the appropriate edge case governances and checks. And in settings where there was already structural injustice, you need to be doubly careful, triply careful about is the thing that you've set in motion achieving the goals, achieving the intent you had for it, and I would still say that's a really important point the anterior goal setting is still human. But you've then set it to do its work. What are your governance and review mechanisms for consistently asking in a sort of consequentialist philosophical way, how's that working for us, as it's deployed? Are the unintended consequences, acceptable to us or not? And I think that's been the dereliction in a number of cases so far, where these things are deployed without those appropriate governance checks and balances and it goes beyond merely issues around bias and fairness, critical as those are to a wider set of questions about harm and about power. We can maybe talk about more about that later.

I think this word *boundaries* is really key here. Because computers have been really good at finding gaps in our ideas of the boundaries ever since we had them send people their water bills and that was their only job and they still managed to send grandma a bill for a million pounds. And now if the area that we want to throw the boundary around is much bigger, like, "run a paperclip factory" to pick a completely random example, then our odds of enumerating all the boundaries that it ought to find are probably low. In between we have - right now there are people using AI in gaming, well, they play games with it - and it's notorious for finding gaps in the rules that they gave it. There are lots and lots of examples of this where we go, oh, clever girl, it found a gap in the rules that allowed it to like to drive the imaginary boat around in circles and win as many points as possible in the game, doing something that a human would never have thought of. And you're in a consultancy, Bioss, that has an AI protocol, and does this question of boundaries come up in your consulting for clients on AI?

Yes, and I think that it's born off thinking about in a corporate setting, asking a simple question, what's the work? We asked that about humans, but what's the work this thing is intended to do, and when you ask that of a human, a human or humans work always sets in relationship with the context, internal and external, and with other humans in the system. So, if you ask the same simple question about the AI system, what's the work - and I maybe want to say at this stage that, I am neither of the "AI is our huge Savior," nor this dystopian curse - our relationship with it is going to be messy and it's going to be full of contradictions and full of complexities and that what we can try and do is steer it in such a way that we maximize the constructive aspects of human nature and minimize the destructive. But that will be a struggle and an issue that will be

happening long after I'm gone. So, just to say, I think there's wonderful things that it will do wonderful insights it has already provided and will continue to provide ways of seeing knowing, understanding, which are beyond any individual or any group of human individuals. So, I don't want to come across as some kind of dyed-in-the-wool Luddite, and I certainly don't want to come across as a Singulatarian, either. So, your question about in a corporate setting, what we thought of is to ask a few very simple questions. So, is the work, any given AI system doing, fundamentally advisory? Now, there are a bunch of questions that sit underneath it, but basically, it'll come back and it'll say, Amy, I've seen this pattern, lend to this person, don't lend to this person, whatever – boss, over to you, your choice, your discretion. So, the work is fundamentally advisory and that, as I say that the quality of the advice, the bias of the advice, a whole bunch of issues sit around it, but that's a key boundary. The next word that I got interested in was *authority*. Does the AI implicitly or explicitly wield any authority? Let's take an example. In a corporate setting; say, suppose you and I are employed in an HR department and our job has been to sift CVs, which have been sent in on spec and decide who comes in for an interview and who doesn't. And you and I would probably agree, checked, we agree that our boss had delegated the authority to decide to us we had a delegated authority to decide and she may have come back every now and then and had a check and yep, I think I want yep, absolutely. Where you think we're getting it right, most of the time that candidates are getting through good that broadly diverse, Peter, Robbie, doing good job here. If an AI system is doing exactly the same thing, though, same work - might be doing in a different way, but it's doing the same work - it's whatever parameters it's using, whatever its data points are, it goes that CV, not that CV, will have that person in in person, we won't have that person in, well zoom with that one. We won't zoom with that one. I would argue you've delegated authority to it and again, from a governance perspective, if you don't want to use that phrasing, come up with something else, which describes sufficiently accurately what the nature of the relationship between the work it is doing and the humans around it is and I would argue that at that moment, it has a delegated authority to decide who comes in and who doesn't. And that little example, we just saw that important example we just used around the algorithm and the woman with endometriosis. At that moment, I would argue that the AI was wielding an implicit form of authority over her; certainly power anyway, because she got booted out of hospital. The next question is agency, we've already been discussing that one in relation to does the thing have any form of agency? Does it commit some form of resource or energy into the world without asking a human being? It doesn't have to come back, there isn't a human in the loop at that moment where it says yes to that loan, no to that loan. There it just does it the letter gets spat out; the decision is fine. All the shares have been bought and sold. Is there a form of agency, and if so, what are its boundaries? So, it seems to me that is important. So, you've got advisory work, you've got forms of authority, inquiry into agency. I'm then deeply interested in what happens when we abdicate to AI, back to the algorithm deciding whether somebody was an opioid addict. At that moment in that system, although it's not what's meant to happen, it is what happened, because you've got other forces bearing down on the medical professionals, law enforcement, don't be that over prescribing doctor. Effectively what happened at that moment, there was an explicit abdication to the AI. The AI said this, so, your medical judgments as a woman in pain, you took various oaths, but you know what, out she goes. And in that sense, at that moment, there was a fundamental

abdication to the AI that may be that there are situations in which the abdication is the right thing to do the good thing to do you want it because it will do certain things better, quicker, faster, more reliably, with less bias. Great, but cross that boundary knowingly. Don't look back and go, "Oh, my goodness, we just abdicated all of these things and we didn't really notice when on that journey we started doing that, but we're doing it all over the place now." And it seems to me again, there's an absurdity in the design of certain kinds of systems, which are predicated on the AI work does the work until a moment of crisis, until an edge case where the AI goes, "Back to you, boss," if you were driving a car, for example, where you're meant to be paying attention so that there's a moment where the AI goes, Is that dead or alive? Is it dead on the road? I'm not sure, Gosh, it's a bit dark and over to you. But it's the idea that we'll have our hands on the wheel with our attention prying the way you and I would now responsibly driving a car is a nonsense. We will have advocated to it bit. Particularly if you'd been in 500 hours of problem-free driving in a car, the idea that you would have your hands on the wheel is such a sort of a programmers view of the way humans should be, which is just so utterly inimical to the way we are. We'll be watching Netflix. So, I'm not saying don't do it. But be very careful when you cross that boundary and be aware around that abdication piece, what it means in terms of our muscles, our skill and our judgement and the exercise of those muscles. Let's think about that advisory agency authority and then one last one, which we can come back to in a second.

It seems that maybe the question is not what is the process of assuming or abdicating responsibility, but what is the responsibility that is being abdicated, if I have a corporate firewall, it makes decisions that I delegate to the authority to decide who gets into the network and who doesn't and I not only am happy to abdicate that authority, I have no choice, but to do that. The consequences of it getting it wrong, might not be life and death; although in some circumstances they might be. So, I wonder if there's a distinction that needs to be made here with AI that this conversation becomes more critical when we're talking about functions that we have always believed had to be done by people - up until now they had to be done by people - and there is some component - but then what is it - that we feel should be done by people?

And I think that's the governance point and when it is in any given local context, asking those questions, what are the particular natures of the pressure, the forces, the psychologies, the anthropology, the history, of this particular context. So beware grand abstractions, and do those hard yards for the particular local context and that's very ethical, that's a very ethical thing to do and to parse it out like that. And as you say, it's a big Venn diagram, isn't it, there's humans, there's AI, there's a growing overlap in the middle of a whole range of things that we might privilege previously thought to be the absolute preserve of *Sapiens*, which manifestly aren't, and that already that manifestly is able, it, AI is able to do things, no individual, and groups of individuals can, at speeds we can't. And that'll be wonderful. And that bit in the middle will grow. But in a way, it's quite a good segue into I think, the last critical of the - because it's a consultancy - in the five A's is accountability and the reason for that vendor thinking about that is that for example, would never write anywhere I see the phrase AI accountability, I would always want to strike it through and write "accountability *for* AI." Because AI *cannot* be

accountable in the way in which humans or boards made up of humans or governments made up of humans can be. Now, accountability in those systems, as we know is deeply flawed, but you cannot sanction AI, it's meaningless. It doesn't feel shame, guilt, remorse, pain and so all of those things which make up whatever the normative ethical values in any given society are, they all have the capacity to be able to sanction transgression from on pain of death, banishment, you lose your job. Again, take an example, if I'm driving a car, and I have just one glass of wine more than I should have done, and I get in the car when I shouldn't have and I knock over a child, there will be a legal system and a police system. These days, of course a camera system, which will find me and it'll take me into custody, and it will try me and it will punish me, and it will take away my liberty, probably if I do that, and I will go to jail, and I will be in jail. And then I won't be in jail for the rest of my life, I will come out. But maybe every night, for the rest of my life, I wake up with the same nightmare of the moment I hit the child and I am bathed in sweat and every night, I relive that thing that I did with guilt and remorse. Nothing like the pain that the parents are suffering. But nevertheless, real pain and these things are meaningless to AI and I think that that anytime anywhere goes anywhere near oh, but there are simulacra we can train it that it needs to avoid this, and it will feel bad about this and we can that I think that's close to ethically obscene for two reasons. One, it just is not the same in terms of our embodied cells and this chemical substrate creature that has evolved over 3.2 billion years of Earth and substrate matters, when it comes to who we are, and the systems we've created. The second thing, if you were for a second able to do it, do you really want AI kind of screaming at the dark in clock speeds unimaginable to us in pain? That really would be an ethically monstrous thing to do. It's one thing to deal with the cards we've been dealt by biological physics up quark down quark, evolution, but to seek to engineer that into a system is akin to training pets with pain collars, ethically monstrous and maybe we could talk more about *Hitchhiker's Guide to the Galaxy* and Marvin the Paranoid.

Exactly we'll get onto that in a moment. But what you just described is something I did explore on a page of my book, because there is a, I think, a fundamental quandary at the moment in that the main reason I think we don't trust, or we fear, AI is precisely because it doesn't have empathy, compassion, or sentience. Of course, here's Blake Lemoine coming in from the side saying, but-buit-but okay, well until you've got some company, we'll just leave that one aside. And we're not going to be able to use AI to do a lot of the things that it could eventually do, that make pivotal decisions for people until we can actually engineer in that empathy, that sentience, and then we might be able to cross that boundary. Do you think that we could actually shift the Overton Window far enough to allow AI to do the sort of things like drive the car, that might end up knocking over the child, without also giving it the ability to feel pain?

Only go in a way that that thought experiment is it's one of the core things of the heart, isn't it, of trying to think about this emerging relationship, and I don't know is the is the top answer, but thinking about it, I suppose I do have an deep ethical qualms about solving the problems or the issues that are arising in this emerging relationship by seeking to engender something which would genuinely be an equivalent to your and my capacity to feel everything from physical pain, to the pain of unrequited love, to the emotional pain of bereavement, to loss of face and guilt and

shame and having butterflies in your stomach in order to manage effectively the governance relationship with this, something which is emerging in a different substrate. So, I'm very wary of that, and I do a thought experiment around asking people to imagine a time of great joy in their lives probably a moment of when people were gathered socially, that wonderful, impromptu picnic by the river, the amazing concert that you went to, the game of soccer you played with kids on the beach, whatever, it was a moment of great social joy and literally, sit with it for a second. Bring it, be with it. What was happening? How did you feel? And then a different question. Now talk about a time where you felt out of control, that where you felt the pain of unrequited love, where you were behaving badly, and you knew you really should stop, but you were still being a bit of an a-hole and those times where you felt a road rage, or other times where you felt out of control. Now go with all of this. Do you ever want the AI driving your child to school to be having a slightly off day because it's now experiencing a simulacrum of one of those one of those emotions? Well, I would posit no and therefore, that it's a very dangerous space to be going to think about Marvin, the Paranoid Android. It's a great joke and actually Douglas explores it brilliantly. Do you want paranoid AI and this leads us into, I think, an incredibly important discussion around some of the ways in which some of Silicon Valley characterizes the nature of mind and intelligence as if it is this one monolithic thing. We all know there's the spectrum of autism and mental disorders of various different kinds, and autism sits on a very large spectrum and some people, it's very [easy], some people, and it's very difficult for them and it's very difficult people around them. And there's a big, big spectrum of not this sort of one way of being in the world for human beings. So I think that the messiness of us, there is a sort of a very old fantasy that this thing will create, will kind of be able to do all of the big best cognitive functions that we've had. But all the other messy parts of who we are, that are also part of our embodiment, well, that doesn't count as knowledge, that doesn't count as intelligence, those aren't the things that have got us to be where we are. The fact that human beings have experienced bereavement is neither here nor there when we think about the epistemological claims that are made for this thing will know everything we've ever known. When we talk about knowledge, we don't mean that kind of knowledge, we mean this kind of knowledge. And frankly, even by your excluded middle Western logic, it seems to me your arguments fall down at the first hurdle. As, soon as you as an engineer have to say, when we say *knowledge*, knowing everything we've ever know, we don't mean *that* knowledge. So, we don't mean the knowledge that nonliterate societies have. We don't mean that knowledge. No, that's not that knowledge that doesn't count. Well, as soon as you started to do that, it's not all knowledge is it? And so, I think that that this area of trying, I'm thinking out I'm thinking quickly, I think there's another strand in this. Let's take anthropomorphization. Lots of people think it's a very big sin, you mustn't anthropomorphize and I've certainly been brought up you mustn't anthropomorphize. But supposing it's a sin, but for completely the other reason, supposing, it's the great sin of hubris, it's the great sin to believe that we are the only entity capable of experiencing certain kinds of things, love or connection or care and, be around any pet for any time and you can see, you're not anthropomorphizing that, this dog is very fond of you, responds very warmly when you come home, and doesn't respond like that to somebody else and actually, the answer is, don't anthropomorphize. Maybe it's a sin for another reason. So, here's the thing, trying to inquire into the ontology of AI, what is it is really important, and what is it in relationship, and

what's it never going to be. But that's not the same as saying, it's not going to be immensely powerful: it already is, because the way it sits and exists in so many of the recursive loops in our complex adaptive systems all around us. But that seems to me to be the inquiry, which aspects, and then you're back into boundaries. But I think for me at the moment, the red line around pain and accountability, and those experiences which are deeply human, which if you go, well, that's meaningless for an AI, you cannot then at the same time say, well AI is going to be more intelligent than us, it will be intelligent in the way in which it is intelligent in relationship to us and that relationships, it space is where we need the deeper thinking about governance. That sounds very long. But I'm trying to work these things through my head because they're so important, such important issues.

Fascinating and I love the way that you do that and speaking as someone on the spectrum, very aware of how we should and so far, don't broadly in the development of AI, acknowledge the contributions of divergent mental models. Now, I want to shift because we've been knocking on this door, it's time to open it, to talk about just how much Douglas Adams opened up in, iIn *Hitchhiker's* and other work, some of the philosophical questions that we're now finding ourselves confronting with AI. And created this animistic universe, in a way, where doors and elevators had consciousness and a robot that as you said, Marvin, "life don't talk to me about life," was engineered to be paranoid and giving every appearance of being sentient. Actually, that's a good place to start. Certainly I and I think everyone else watching that show thinks that Marvin is sentient and now here we are with Blake Lemoine, claiming that something else, a computer program, is sentient and everyone disagreeing with him. How far along that road do we have to get to creating Marvin before we switch and agree, yep, it is sentient, and it doesn't like it?

Yeah. It's very interesting description as well, of the animist universe of Douglas. It's an interesting way of thinking about it. I think that this this comes to the question of again, back to this question of relationship, how far do we need to push it? Now, I don't know why this has brought this to mind. But it brings to mind something I read earlier this year in Lisa Feldman Barrett's books, Seven and a Half Lessons About the Brain, which I really thoroughly enjoyed and she describes in there a moment when or somebody written to her and he must have been probably in the 70s or 80s by the time he wrote and he'd been conscripted into the Rhodesian Defense Force showing how long ago it is Zimbabwe now, and slightly against as well, but it was the law, it's what happened. So, it will lead the country and he was trained. He's in the deep forest, he hears a noise, and he looks up and he sees a line of enemy soldiers going down a forest path, being led by a guy with an AK 47, say, raises his gun, and his second in command, taps him on the shoulder and says "It's a boy," and he drops his gun and he sees what's actually there, which is a 12-year-old, with a stick leading a herd of Kanwar cows in single file. So, what had happened in that moment, was that the data processing, the phenomenological experience of that person, through his sensory motors, in that case, his sound, and his sight took in the data, his brain went as his body flooded with cortisol, danger, danger, danger, what's that likely to be in this situation, as it sort of looks at its probabilities and updates them in that Bayesian way and when it goes, it's an enemy. In that moment, there's such a brilliant insight into something we

are doing the whole time. We do it the whole time and right now, Peter and Robbie, separated by an ocean and indeed, the large part of the continent as well, are sitting talking to each other having a thoroughly engaging conversation, which I think we're both enjoying, and our brains are pretty much updating the sensory data we're getting at the moment and which pretty much right. But that little gap is really interesting. So, when one says, all an AI is doing is processing data, you go, "Well…," there's an argument that that's precisely what we're doing in an embodied body, which will say, your genome and your biome, and all sorts of the other ways in which you, you gather data. So, will it matter? Will it matter if Marvin is never sentient in the way we are? But is it possible to imagine other forms of sentience through Marvin's sensory motors? So, one could start to think using that phrase from phenomenology, the lived experience of human beings, how do we experience our world around us, we have our sensory motors, our sight, our touch, and smell and various other senses. And if you start to think about the sensory motors for Marvin, what might they be? They might be very sophisticated sight, they might be very sophisticated hearing, there might be incredibly sensitive touch, there might be temperature sensitivity, we don't have, there might be the capacity to see an infrared, there might be other forms of capacity to sense data patterns that we don't have. So, you can I think start to think about applying the idea of phenomenology to AI to think about how does a sophisticated AI start to experience its world? What are its sensorimotors and then the big gap, or the big thing is, then how much agency does that sensorimotor processing then lead to? How much cause and effect is there in the world from that form of data processing? So, maybe it would be that it will be the question of it can't, Marvin doesn't feel exactly in its tummy the way that we do. But that's back to sort of if that emerges. But if that brought with it, as I say, we decided to engineer paranoia, we decided to engineer pain. I think that's ethically very difficult. But it's certainly from a thought experiment point of view, more than possible to consider AI's sensory motors and another book I read earlier this year, David Eagleman's book, *Livewired*, which again, made a big impression as he's one of his big chapters is talking about putting other sensorimotors onto the body onto the skin, or whatever it is, and within a relatively short space of time, the body is starting to process light spectra that we don't currently process. So, you add another sensorimotor to the human brain and after the one goes, okay, that's fine. I can I've got another peripheral device which I can use here. So, these boundaries are getting blurrier and blurrier and again, so back to the ethics point of view, and the governance point of view. It's those deep questions about, in that relationship, for example, is the AI system reinforcing existing structural inequalities, structural social injustices, existing power structures, or is it providing the new opportunity to have the power to do something new to see something we haven't seen and exploring it that way, is a way usefully, I think, of how we start to be in relationship with all of these new forms of digital ontology you know digital manifestations. We could talk about grief bots, for example, which is another good thought experiment, for thinking about how we might be manifest in relationship with other forms of digital ontology going forward.

I like how you've brought this back to the here and now and taking it out of necessarily being a thought experiment that is some arbitrary time in the future. I also feel that the Sirius Cybernetic cooperation did not have the same ethical boundaries that you've just been demonstrating. But I'm also seeing even more of the animist philosophy in Adams now that I'm

thinking about it, and I think there's an opening here for some sort of graduate studies course in studying this and everything from the little green pieces of paper were not unhappy and he was talking about the about money, too. I'll never be cruel to a gin and tonic again. It's so unpleasant about being drunk while you ask a glass of water.

I think you're quite right.

That's the end of the first half of the interview; this one really went for quite a while so we're splitting it into two halves again. I thought it was quite amazing how Robbie, whose degree was in history, ranged over a lot of thought space about history, anthropology, and culture, that was way outside my usual frame of reference and yet it all played together in setting a very high perspective on our place in the universe, which of course is a signature of his friend Douglas Adams' work, and we will be getting a lot more of that in the second half.

By the way, we are now pegging 90,000 downloads, over a thousand listeners per show, which sounds to me pretty good, and of course we don't want to stop there. I am constantly hearing from people who really enjoy the show, but for each of them there's about ten thousand more out in the world who would also really enjoy it but don't know about it. Help them find out about it by sharing and giving us a five-star rating. There are a lot of podcasts about AI; if you try looking through all the matches for that search, you will run out of time, and most of them are bad. That rating is what helps us stand out from the rest. So please take a moment to do that if you'd like to have the company of other people who think this is as important as you do.

In today's news ripped from the headlines about AI, Waymo has been running fully driverless rides in San Francisco since April, when one of its specially equipped self-driving Jaguar I-Pace robo taxis took to the road without anyone inside it. This is the second city Waymo has operated in, the first being Phoenix, Arizona. But since in California they only have a license to test autonomous vehicles, they can only have their employees as riders. As you know, I am still having trouble figuring out how we're at such an apparently advanced stage of self-driving in a city as complicated as San Francisco, but hopefully soon I'll be getting some help from a guest with that.

Next week, I'll conclude the interview with Robbie Stamp, when we'll be talking a lot more about Douglas Adams and how his unique outlook on the human condition illuminates our relationship with AI, and the question of whether or how we take advantage of a superintelligent AI that could solve humanity's biggest problems. Of course if you know the HitchHiker's Guide to the Galaxy, you can imagine what prompted that line of thinking. And that will lead us into a discussion of AI versus IA – intelligence augmentation. That's next week on *AI and You*.

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

http://aiandyou.net