The coming of 'The Singularity'...or not?

LISTEN NOWDOWNLOAD AUDIO

Imagine a future where computers exceed our own intelligence; where problem solving is no longer limited by human thinking -- what then? It's a moment in technological time some call 'The Singularity'. But how much is technological reality, and how much fantasy? Science writer Mike McRae catches up with AI researchers and sci-fi writers to ponder the possibilities and probabilities.

HIDE TRANSCRIPT

Natasha Mitchell: *All in the Mind* on ABC Radio National, I'm Natasha Mitchell, good to have your ears -- and your brain while it's still something to be in awe of, because as you'll hear today, some argue we're heading towards 'the singularity' a point in time when artificial intelligence may well far rival our own.

Australian science writer Mike McRae is exploring whether the prospect is pie in the sky or a real possibility; kicking off with the very entertaining sci-fi writer Richard Morgan, author of *Altered Carbon*.

Richard Morgan: I'm a singularity iconoclast. I absolutely don't buy in to it, and I think the best description I ever heard of it was William Gibson when I was on a panel with him in France last year. he said, 'I am not buying in to what is essentially the geek rapture.' And you know the more you think about it the more you realise that's what this is, it's this sort of geek culture going on about 'we'll all be taken up into the great machine.'

Nick Bostrom: I think there certainly is a tendency among some people, but not all, to, how shall we put it, to make singularity into some kind of almost spiritual or religious idea, some form of inevitability that is guaranteed to bring peace and abundance to earth. Among the more thoughtful people who are thinking about this, however, they are aware that we are very ignorant about these questions and it is very difficult to predict what will happen. Also there are major risks as well as potential upsides with any transformation that profound.

Mike McRae: Picture a future where computers are far more intelligent than any living brain. Problem-solving is no longer limited by human thinking and progress begins to exceed the limits of our imagination. It's a world where nearly anything might be possible. But is this just another version of the elusive Utopia, a sort of heaven for geeks, or are we truly about to enter the glorious age of the digital super mind?

In physics, singularities are points in space where laws break down. In the 1980s author and mathematician Vernor Vinge borrowed the term to refer to a point in our future where all bets were off when it came to predicting new technology. Philosopher Nick Bostrom heads up the Future of Humanity Institute at Oxford University and as he explains, defining the technological singularity depends on who you ask.

Nick Bostrom: The term technological singularity has been used in at least three different ways. On the one hand it can be used to mean just the general process of technological acceleration. If you plot the development in areas like micro processors, gene sequencing or even just the general economy, there is a general sense that things seem to be accelerating, and if that gets faster and faster, at some point, people might talk of a singularity when things are moving so fast that the world is radically transformed even within a short period of time, like a year or so.

Then there's the second sense of this word, the technological singularity. It has been used to refer to a predictability horizon, some point in the future beyond which we are unable to forecast what will happen because changes at that point, as they said, will be so rapid and profound that we can't foresee what will come after. The most interesting meaning of this phrase, the technological singularity, in my view is the meaning in which it refers to an intelligent explosion, specifically the development of a greater-than-human artificial intelligence.

The idea is that once you get to a certain level of artificial intelligence, that intelligence can use its own smarts to enhance itself and to invent even smarter forms of artificial intelligence, so that within a very brief period of time you might go from something that is just slightly smarter than a human to something that is perhaps radically super-intelligent.

Mike McRae: And it's that last take of the singularity as an intelligence explosion that we're considering here, because this belief that we can artificially improve how we think has inspired countless fantasies of a future technological explosion. Well known futurist Ray Kurzweil assures us the singularity is inevitable, but surely first we need to work out how to compute the complexity of a biological brain, and is this even physically possible to do? Could we really reduce our thoughts to simple strings of code?

Computer scientist Associate Professor Marcus Hutter from the Australian National University.

Marcus Hutter: OK, the very short answer is yes, not everybody would agree to it but a significant fraction of our researchers and physicists -- if you come from the physics side and look at physics, physical theories which explain all kinds of things, you know, the whole universe or parts of it, all these theories are computable. Now look at living organisms: nobody has yet found the difference between life substance and normal physical processes. And if you extrapolate then the human brain should also be just a physical process and this is then most likely computable. If you look at the details how the brain works and compared to how a computer works the details are very different. I mean the human brain is based on neurones which are massively parallel and modern computers are mostly serial with high speed and a few processes only. In principle this doesn't matter.

Mike McRae: And this doesn't matter in principle because of a theory called the Church-Turing Thesis: put crudely, it suggests that all physical processes should be computable.

Marcus Hutter: So that's in principle. In practice it makes a difference, mainly speed, if you tried to simulate, say, the processes in the human brain on a sequential standard computer it will be very slow, so it is not the wisest thing to use the standard computer, it would be better to have a special computer which simulates directive neurones, if you take the neurone approach.

Mike McRae: So theoretically the work done by our mush of grey matter could be achieved with the right amount of wires and logic gates, but that's if we can come to grips with how the brain itself works first. Nigel Dobson-Keeffe is a research scientist with the Defence Science and Technology Organisation in Australia. He's sceptical that the growth in computer processing, or CPU power, alone will get us there.

Nigel Dobson-Keefe: There doesn't really appear to be any good solid evidence that the limitation in why we're not succeeding in that area is purely due to just straight grunt CP power and memory and all these sorts of things. It's probably more fundamental than that in the mechanisms by which we're trying to in one case simulate how humans behave. It's just something we haven't done, got some of the key points of it at this point in time. And that's not surprising, because at this point in time there's many aspects of human behaviour that we simply don't understand either.

Mike McRae: Many people seem reluctant to compare our ability to think with, say, a computer's ability to process information. In your experience could this impact on how we interact with super-intelligent machines in the future.

Nigel Dobson-Keeffe: When we use the words 'think' and 'process' they are actually quite different ways, they are two different consciousness actually go about things. When we talk about a machine we talk about a process, generally they are very linear in their thinking, they use algorithms and they use data structures that are quite hierarchal. Whereas thinking in humans, especially things like creative thinking, is a much more random and free flowing process. And the problem here

is when we design our machines because they don't actually reflect how humans think, this leads to a very challenging area which is how do we actually get the interaction of machines and humans to be efficient and useful from the human's point of view and also useful from the machine's point of view with the task it's meant to do.

Mike McRae: What is it that fundamentally makes the difference between how a computer works on that level and how a human thinks? In other words is it possible to create a computer that could actually mimic the thought patterns of people?

Nigel Dobson-Keeffe: The best way of comparing is probably by outcomes. Can the machine behave like a human as opposed to does it do it in exactly the same way? So we don't really care how a machine would learn to understand other humans and be able to take into account other humans, what they're thinking and how they're behaving, as long as at the end of the day you have a machine that would know when it's appropriate to communicate when you don't interrupt. And those are the sorts of things humans learn and if you wanted a machine that could interact very well with humans it would need that sort of ability, otherwise, like most machines we'd just find them annoying, you know, you're in the middle of trying to do some work and the email goes off. So that situation awareness that humans are very good at, and that's one of the things that humans do very well, because we have the flexibility of thinking, we can make a very quick decision based on less than idea information.

Nick Bostrom: I think there's a big difference between computers today who do very well on some specific tasks like calculation or playing chess, or managing huge amounts of data, from the form of general intelligence that we humans have; the general ability to solve novel problems across a wide range of domains. That's something that computers don't have yet. But if and when computers attain that general intelligence, they will be able to apply it to all sorts of things including the design of artificial intelligence.

So at some point the process of developing better artificial intelligence will be driven not by human programmers and human ingenuity but by artificial intelligence itself. And at that point you could get the kind of feedback process where a smarter artificial intelligence creates artificial intelligence that's smarter still, and that new intelligence will be able within an even shorter time to create an even more powerful intelligence and you could then, at least as far as the reasoning goes, get an intelligence explosion.

Mike McRae: There's no doubt that today's computers can already outdo us when it comes to certain mental tricks. But the intelligence of the singularity is far more complicated. We're not talking about just crunching big numbers here, we're looking towards a system that might contemplate, dream and understand the universe. Now that's quite a big step.

We heard from artificial intelligence scientist Professor Noel Sharkey from the University of Sheffield on the show recently. So does he think we'll develop such a digital intelligence?

Noel Sharkey: Well the short answer to that is no, but I'm not certain about that, I can't be certain, but it's all based on the notion of computational intelligence. That means that you can recreate a human intelligence using computers, and people forget that this is actually an assumption, it's never been tested. So that's the assumption that's being worked on with the idea of an artificial intelligence. I mean it could be that we're certainly physical and our minds come from the physical realm but whether or not the computations is another matter, it could be very chemical and not replicable really.

But the other side of artificial intelligence is this idea that you know we didn't need to look at the bird to build an aeroplane so maybe we can build a machine that's intelligent without it being animal intelligence. And there's a couple of things wrong with that: one is that we did look at the bird, Leonardo looked at the bird and we studied the aerodynamics of the bird in order to get flight -- that's the first thing. But the other thing is on the same idea you need to look at human intelligence in order to get machine intelligence.

Now with machine intelligence it tends to be single functional, so for instance the world's greatest chess player was Deep Blue, it beat the world champion, and that was an extraordinary achievement because I was working in AI right back when we talked about the notion that there was a horizon effect in artificial intelligence as far as chess was concerned. So it might beat the odd master but it would never get to grand master level and here it was beating the world champion which was absolutely incredible. But it was an engineering achievement in as much as the machine was able to search millions, upon millions, upon millions of moves ahead, which a human can't do. And this really isn't intelligence, this is brute force, it's like arm wrestling with a digger, with one of those mechanical diggers and then saying it's stronger than you.

But the thing is it's single functional. If Kasparov had got up and left the table the robot wouldn't have cared less. If you'd taken all the chess pieces away the robot would have carried on moving, sorry it wasn't a robot it was a computer, but it didn't care less whether it won or not, it had no motivation, no desire, it didn't have a sort of sense of 'I've got to strive to be the worlds best chess player and learn very hard and then enter the competition and get excited'. So it's a kind of very different kind of thing than a human.

Mike McRae: Professor Noel Sharkey. Even if intelligent software could play a key role in discovering something new, it's hard to believe a computer program might ever be awarded a Nobel prize.

The line between tool and colleague may seem trivial, but as a social species, the way we interact with intelligent technology is just as important as the technology itself. And just how important is it for us to recognise our own humanity in an intelligent program? Nigel Dobson-Keeffe.

Nigel Dobson-Keeffe: We have started looking at adding the aspects of emotion, in some way fake emotion, to some of what we call virtual advisers where we have these software avatars on the computer screen that looks like a human, sounds like a human and moves slightly like a human, has lip synchronisation and stuff. Adding emotional cues to this such as happy and sad can be done, and has been done, although at this point in time we don't really know how important that is in a situation of say having an adviser tell you some information that might not necessarily be very emotional, it's telling you some facts. You could surmise in that situation that adding the emotional aspect to it probably doesn't help much.

In a situation that is potentially emotional, adding those features to a machine might well be useful but humans are very good at judging emotion and the honesty of the emotion as well, and so that could be a very challenging task to try and have a machine that is convincingly emotional. Being a social creature, we have evolved to place very high reliance on the ability to understand the emotions of our fellow humans. That's something that humans actually are very, very good at, you know we're quite poor at doing things that machines are quite good at such as repetitive tasks and long mathematical things, but when it comes down to emotional aspects humans are quite good at that and so that could be a very challenging task. There's probably more scope for having advancement in other areas of the machine intelligence before we actually get there which might in itself pose some problems. Because if we have actually advances in how they think, or take into account complex situations but we haven't developed any emotional aspects to it, you could potentially have situations where machines just don't come across as very convincing, or seem very cold and calculating in which case humans might not trust the machines, in which case the machine might even be right but it doesn't really matter because the humans aren't very trusting of them, they'd rather have their advice from another human.

[***Terminator 3* trailer**: It is time (explosion) there is a new Terminator, the TX.]

Natasha Mitchell: On ABC Radio National's *All in the Mind*, going global on Radio Australia and online, I'm Natasha Mitchell and today Australian science writer and communicator Mike McRae is considering whether we'll ever reach what some dub 'the singularity', a time when artificial intelligence

exceeds our own, and technological progress could explode exponentially as a result.

[***Terminator 3*** **trailer**: It's faster, more intelligent and more powerful. ... Oh my God. It has been programmed to destroy other cybernetic organisms. It was sent back through time for one purpose only, to kill us all.]

Mike McRae: Our fascination with creating artificial intelligence is matched only by our fear of possibly succeeding. It's far from a recent concern -- Jewish folklore tells stories of the golem; a living statue created to serve its master, yet instead murders him in cold blood. And so how careful should we be in our approach to bringing on the singularity? Oxford philosopher, Professor Nick Bostrom.

Nick Bostrom: If and when we become capable of creating a machine that is or has the potential to become truly super intelligent, bringing into existence the first seed AI, or this first machine that might reach super intelligence, will be one of the, probably the most dangerous thing that the human species has ever done. Intelligence is the most powerful force in the universe, it's what makes humans such dominant creatures on earth, it's certainly not our strong muscles, our sharp claws or teeth, it's clearly our brains. Similarly a machine that had general intelligence that surpassed human intelligence would also be extremely powerful. It would therefore be critically important to make sure that the first such super intelligent machine was human-friendly, that it had some motivational system that ensured that it would seek only outcomes that were good for human beings.

Mike McRae: Nick Bostrom, who heads up the Future of Humanity Institute at Oxford University.

Regardless of whether the singularity will occur within our lifetime or not, it remains an inspiration for many science-fiction writers. Together with Vernor Vinge, renowned authors such as Greg Egan, William Gibson, Greg Bear and Isaac Asimov have all used it as a theme.

Reading from *Altered Carbon*: Human life has no value. Haven't you learned that yet Takeshi with all you've seen? It has no value, intrinsic to itself. Machines cost money to build. Raw materials cost money to extract. But people? You can always get some more people. They reproduce like cancer cells, whether you want them to or not. They are abundant Takeshi. Why should they be valuable? Do you know that it costs us less to recruit and use up a real snuff whore than it does to set up and run a virtual equivalent format? Real human flesh is cheaper than a machine. It's the axiomatic truth of our times.

Mike McRae: In his award-winning novel *Altered Carbon* British author Richard Morgan explores the implications of using computer hardware to record and store an individual's mind. His character Takeshi is killed and re-sleeved in a new body. It's a potentially dystopian scenario as are so many sci-fi stories starring super intelligent computers.

What do you think it is about the super-intelligent computers that suit such a depressing vision of our future?

Richard Morgan: A very good question. I think the funny thing is, yes I think Gibson and myself both get the sort of same rap for writing dystopian futures as it were, but the truth of the matter is I don't really think, I wouldn't want to speak for Gibson but I look at the future I've written in *Altered carbon* and to me it doesn't seem dystopian, it seems pretty much just a current extension of the way things are now.

If that's dystopian then I guess that the present we are living in now is also dystopian. And I think the issue of the technology is something that Gibson certainly did par excellence, it certainly was a big influence on me, is the idea that technology is not going to solve our problems because the problems are human. So all that will happen is the technology will be used divisively: some people will benefit, some people will lose out. There will be an over-class who get to the benefits of the technology, there will be an underclass that suffer from it if you like. So yeah, the sleeving technology in *Altered Carbon* if you're rich it's great you never have to die, you can zip from body to body and live forever. If you're poor you become a resource, you know, your body itself doesn't really belong to you, it can be taken away from you and given to somebody else and you will suffer and the technology will not save you.

It's really, not consciously I think, it is just an attitude that I

had, and it's a reaction against that very weird technophilia that you still find sometimes especially with the sort of libertarian wing of science-fiction that believes that technology is just going to come and turn us into really nice people -- and that ain't going to happen.

Mike McRae: If we went then on fiction alone, it seems clear that any attempt to create a self-aware computer results in some sort of disaster or at the very least a moral lesson for us not to play God. Where do you think that fear comes from?

Richard Morgan: I think what it is it's an iconic myth, you know the Frankenstein myth if you like. I guess Mary Shelley was the first one in the modern age to articulate it, but it goes way back beyond that, you can look into mythology and you can see the same thing, it's this idea of setting loose powers that ought not to be out of the box. And you see that, it's the Promethean thing, it's Pandora's Box and it's just this idea that some things are best left alone, you've got to be careful how you handle stuff, it's the basic lessons in life -- aren't they?

We're in Australia here so let's look at the sort of Aboriginal context, if you like, I guess the Aborigines have this thing about you only burn in the burning period because God knows, if you burn later than the burning period the entire country goes up in smoke. So that I guess is a kind of Frankenstein myth, isn't it, they've internalised in their culture this sense there is a time to set fire to stuff and you bloody well don't do it except at that time. So I think all that happened with Shelley and then everything that came afterwards, that that got tacked onto technology and technology became the environment in which we tell the story of you've got to be careful.

So I don't think there is any big mystery but of course the thing about these iconic myths is that they speak to something in our makeup, in the way we are as human beings. That's why they're so tenacious, that's why you keep getting references back to this story. So it's actually quite hard to tell a story about creating artificial life or something of that sort that has any kind of happy ending. You know we're not wired to listen to that story; we are wired to hear a story of that sort that ends in tears.

Mike McRae: How do you see intelligent technology impacting on social interactions further down into the future?

Richard Morgan: What I think is very clear is we're seeing a simulacrum of intelligence emerging in our machinery, in the sense that we are now able to design machines that appear to behave intelligently. We are getting to the point where you've got music systems that will notice what you listen to most and then put that into playlists so you get your favourite songs listed. Marketing devices where you can make a look at what people are focussed to on the internet and then provide them with advertisements directing them to things that they might like and so forth. And this looks like intelligence, because it's sort of reactive and it webs into the way that we behave, especially when we're in cyber space.

Whether that is intelligence I don't know because first there's a very big philosophical issue about exactly what intelligence is and how far down the sort of creature ladder do you have to go before you find a creature that is an animal but isn't actually intelligent. Is consciousness the issue, and which animals have consciousness; because again we don't really have any answers for that yet. So yeah, I don't think there is any doubt that our machine systems are going to get more and more sophisticated and they will start in many ways to resemble personalities.

The best one I've seen so far is this psychological program they've got for a computer that basically provides psychiatric counselling to people because it's programmed to sort of push response. It responds to most things with a question and it gets the person to react and they've tested this on people and people can sit down and have a conversation with a computer and they are quite happy to, even when they know that it's a computer, they are still quite happy to because the computer has been designed; it's able to fake human interaction. I don't think that makes it intelligent, I think that's anthropomorphising if you like and in the same way that

you've been swimming a lot in the last couple of months and fish can sometimes look amazingly human and there's expressions on fish faces and you go, 'Man, he's looking really lugubrious.' 'It's a fucking fish, man.'

You know similarly just because something looks like something it doesn't mean it is and I think we'll get a lot of stuff that looks like intelligence. Whether we're going to have actual artificial intelligence, I do not know. My sense is it's going to be a long time coming and not least because we don't really understand our own intelligence enough to say let's build something like that. That's it -- you know Vernor Vinge, he thinks it's coming, he's talking 15 years at most and he's a computer scientist. But then again he's also a libertarian and anyone who's libertarian, I've got to question their hold on the physical realities of life in other things as well. You know I think he's an intensely optimistic man and I'm not ... so (laughter).

Nick Bostrom: But recognising this human fallibility and the difficulty of getting these things right doesn't mean that one can rule out therefore any scenario that sounds really radical. Our ignorance means that we must assign some non-trivial probability even to these very radical possibilities. We don't know when there will be a singularity or even if there will ever be one. We also don't know that there will not be one, and we have to be able somehow to entertain both possibilities.

Richard Morgan: It's just a variation on a very old religious undercurrent in the way human beings behave, this idea of sort of moving on to a better place. And the thing you find about the singularity enthusiasts is they are all convinced it's going to be great. But I don't know about you, I'm fairly healthy, and I'm not interested in leaving my body behind, at least not now. And if someone was to say to me OK your options are you upload and lose your body, or we can clone you a new body that looks pretty much like the one you're in now, or even one that doesn't look pretty much but is healthy and functional, I think I'd go for the clone option every time. I have no desire to exist inside a machine matrix, it might be fun as an arcade game, it might be fun to do it for a bit but I wouldn't want to live there.

Mike McRae: For as long as we've been aware of our ability to make machines that can think, creating human-like intelligence has seemed just out of reach. The thing is, although our brain can be described as a biological machine, it is one of such complexity that we are still struggling to describe how it functions. Knowing when or even if we'll make a better model is anybody's guess.

For now the technological singularity might say more about the human condition than it does about what awaits us in the near future. Yet just as writing and computing have greased the wheels of progress in the past, it can't be dismissed that unforeseen inventions might indeed help us in the future.

Natasha Mitchell: Mike McRae there. Canberra-based science communicator and writer. And I've put Mike's extended interview with the very entertaining sci-fi writer Richard Morgan up on my *All in the Mind* blog, where I'd really love your comments too. Do you think the prospects of the singularity are inevitable or totally fanciful? Get to that, more info, the audio and transcript all via the *All in the Mind* website at abc.net.au/rn/allinthemind

I'm Natasha Mitchell, thanks to co-producer Anita Barraud and studio engineer Russell Thompson. Look forward to your organic intelligence engaging with the show next week -- bye for now.

Guests

Professor Nick Bostrom
Director of the Future of Humanity Institute
Oxford University
http://www.fhi.ox.ac.uk/staff_info/research_staff/nick_bostrom

Associate Professor Marcus Hutter
Research School of Information Sciences and Engineering

Australian National University
Canberra
http://people.cecs.anu.edu.au/user/470

Professor Noel Sharkey
Professor of Artificial Intelligence and Robotics
University of Sheffield
UK
http://www.dcs.shef.ac.uk/~noel/

Nigel Dobson-Keeffe
Cognitive Scientist
Australian Defence Science and Technology Organisation
Australian Government Department of Defence
http://www.dsto.defence.gov.au/

Richard K. Morgan
Science fiction author
http://www.richardkmorgan.com/

Further Information

All in the Mind blog with Natasha Mitchell for your comments and discussion

All in the Mind: 1/09/09: Do you read me HAL? Robot wars, moral machines and silicon that cares - Part 1

All in the Mind: 8/09/09: Do you read me HAL? Robot wars, moral machines and silicon that cares - Part 2

Furture of Humanity Institute, Oxford University

The Singularity Summit
Gathering in October 2009

The Singularity is Near - the movie (scheduled for release in late 2009)

The Singularity Institute for Artificial Intelligence
Hosts of the Singularity Summit

Ray Kurzweil
Futurist

Essay: The Coming Technological Singularity
Written by Vernor Vinge. First published in 1993.

Singularity 101 with Vernor Vinge
Published in hplus magazine, April 22, 2009.

The Church-Turing Thesis
Stanford Encyclopedia of Philosophy