2016 Conference Transcription

Date	Thursday 31 March, 2016
Session Title	Intelligence
Session Time	10:30-12:45
Moderator	Vikas Shah
Speakers	Darius Kazemi
Notes	n/a

Intro	Hello and welcome to FutureEverything's 2016 festival podcast series. Over two days, in Manchester's iconic Town Hall, we tasked designers, artists, scientists, and many more, to re-think our resources. From life, earth, and intelligence, to community and uncertainty, our speakers ask what we might need less, and more of, in our new future. Is artificial intelligence really that smart? And where are seeing robots exist? In this session on intelligence, we heard from Darius Kazemi, an award-winning artist and botmaker. He is known for making weird internet stuff, such as bots that generation random Amazon purchases, surreal metaphors, rap battle lyrics, pick-up lines, and everything in between.
Darius Kazemi	Hi everybody. My name's Darius Kazemi, and first I would like to pre-emptively apologise. I left my home in Portland, Oregon on the West Coast of the United States thirty six hours ago and experienced a cascading series of delays in my flights that resulted in me arriving here in Manchester two hours ago. So I'm running one hundred percent on coffee right now, and I guess I'm saying, be prepared. I don't want to spend too much time on my own background, so I'll keep this intro part short. I was invited to speak to you all today because I make bots. Not physical robots, but software bots; things that I like to define as computers communicating with humans via mediums designed for human to human communication. Most of my bots live on Twitter, but some live on Tumblr, and I've done SnapChat bots and Slack bots, and bots for a few other platforms you've probably never heard of. I make stuff that makes custom content for humans. If you follow the sorting bot on Twitter, it will generate a unique rhyme for you and sort you into a Hogwart's House. You stand before the sorting hat, 'my real estate is iced for [inaudible 02:22] for you are under-priced' and 'the ethic of a porcupine, the wisdom of a goose in Ravenclaw is where you go since you are quite [inaudible 02:32].' 'Two Headlines' is another one that I've made. It's a very simple algorithm. All it does

is it goes to Google news and it finds a headline from one category, like sports, and a headline for another category, like world news. And it goes to a little sidebar where it lists those topics in the headlines and it swaps the topics around. And it results in things like ‘This Town has Resisted Pelicans for Eighteen Months, but Food is Running Low’, ‘Donald Trump Joins Cast of Game of Thrones’, ‘Five Things You Should Know about San Francisco as a Service’, and the exceedingly ominous; ‘Humans to Restart by May 15th’ so we've got about a month and change, just so you know.

I also have more visual bots, like this recent work called Glitch Logos, which takes vector images of recognisable corporate brands and morphs them into something still recognisable, but wrong and sometimes disturbing. And so I've been playing with these for a while. I just love how vectors sort of glitch into these weird organic shapes. Last year I co-founded Field Training. It's a creative technology co-operative based in Portland, Oregon. I co-founded that with my wife, Courtney Stanton, and we do things similar to my personal bought work; including this bot called Madame Ava. It took over the official Twitter account for Dungeons and Dragons for a couple of months. It's actually still operating, I think, and it creates custom fortunes for fans drawn from a D&D flavoured tarot deck, and this ended up being my most successful social media campaign ever, so that was pretty cool.

Anyway, I feel like I often see gloomy messages about artificial intelligence in the news these days, and fortunately, I'm here to deliver some good news to all of you. A robot apocalypse is already here. In fact, we've been living in the robot apocalypse for a few centuries now. I'm not talking about big, scary mechanical robots that look like something out of a Terminator movie, and I'm also not even talking about ill-conceived, racist, fake teenagers or spambots, or any of the stuff that I've been building in the last few years.

The claim that I'm making is simple; the robot apocalypse has been here for centuries, we are all living in it. Every person in this room has been living in it for their entire life. And I want to back-up here for just a second and talk about the term artificial intelligence. When people talk about artificial intelligence, you'll always get a couple of wise-a*ses in the back who yell ‘hey, can you define intelligence?’ I mean, we can't even agree on how to measure intelligence in humans, so how in the world are we going to measure it in a computer? What are we evaluating our intelligence by? And this is a good point, but it's a little bit disappointing to me, because these very smart people are missing out on an opportunity to be even more of a wise-a**, is where I come in, and I say ‘forget defining intelligence, can you even define ‘artificial’?’ I'll back-up even more and point out that a lot of words get thrown around when we talk about artificial intelligence; algorithm is a big one.

Algorithms have been around for a very, very long time. In the popular press, algorithms are synonymous with big data, and technology, and surveillance, and all that sort of stuff. But algorithms are named after this guy who's a Persian mathematician, Muhammad ibn Musa al-Khwarizmi… my Arabic is bad, my Persian is a little bit better, who, sometime around the year 800, wrote a book so full of algorithms that the English word ‘algorithm’ is itself derived from a

Latinisation of his own last name. And he wasn't even the first person to use algorithms. Ancient Greeks and mathematicians in India were full of people deploying algorithms every single day. And an algorithm is, simply put, a set of self-contained instructions with a beginning and an end. All software is a set of instructions.

So almost every computer programme is an algorithm, but not all algorithms are computer programmes. For example, when you learn to do long division by hand as a child, you were carrying out an algorithm, and this is what I mean when I say there's a blurry divide between artificial and natural, like is this an artificial construct? Would you use the word ‘artificial’ to… would you define maths as a technology? Some people would, some people wouldn't. Some people say maths is inherent in nature. It's blurry, it's not very clear. Next we might ask what makes a human carrying out an algorithm different from a computer carrying out an algorithm? Well, we are slower than computers at a lot of things, but that’s beside the point. If I ask you to solve a bunch of long division, you'll probably get bored after a while and stop, or have to go eat or something like that. And a computer won't stop doing what it's told unless you make a mistake giving it the instructions; so you introduce bugs into it, or it physically breaks or the power goes out or something like that, but that’s much more easily controlled than a human getting hungry. But not all things are created equally. We have far more coercive algorithms than long division. These are things that I call social systems.

My favourite one is the corporation. It's a set of rules written down on bunches of pieces of paper, typically a combination of articles of a corporate or association and laws passed by government bodies. These sets of rules say ‘do these things in this order and make sure these numbers go up’ essentially, and yes, if you had an advantaged enough computer you could programme it to run a corporate. Many corporations are at least partially run by computers these days, but still, the vast majority of work is carried out by humans, and these humans are heavily incentivised to carry out this algorithm. Humans who fail to carry out the algorithm, or simply choose to stop participating in this are called from the pack, people who are allowed to live in society and face economic precarity at the very baseline as a punishment for not participating in this system and for not executing the algorithm that we had set up for ourselves back in 1600, and by the way, this is the charter of the Hudson Bay Company from May second 1670 under King Charles II. So that’s the source code for the Hudson Bay Company. It's like seven thousand words long and very technical.

I want to make this next point very clear, which is that when people say that we can't provide everyone social services, what they mean is the corporate AI won't let us provide everyone’s social services. And this gets back to my original point, which is that the robot apocalypse is already here and we are living in it. An out of control AI does not need big scary machines to control the world, why would it? In America, we've optimised our entire national infrastructure around the automobile, and yet I still have to take my car to the mechanic every six months just to make it not break down. I can only imagine how much breaking down and how much upkeep an actually physical robot would need to get things done. And why would an AI do this sort of thing? Why would an AI send out an army of

these things when it already has machines that are achingly slow, but have the advantage that they can carry out complex tasks, last for decades, are self-sufficient, and often docile.

Of course, I'm talking about us. Humans. You know, we can stop executing certain algorithms, but we just don't, for some reason. And why would an AI rely on clunky robots when it could coerce humans into simply building a world to suit its own needs? But you might ask; 'Darius, what you're talking about isn't artificial intelligence. Intelligence implies agency and consciousness. These machines and algorithms clearly aren't that.' And here's where we start to get into hairy philosophical questions that blow past epistemology and into the realm of ontology. What is intelligence? And I don't have time to get to that, but I'm happy to talk to any of you later on, preferably once I've gotten some sleep. But the lay definition that someone might give as well; an intelligent thing is something that's capable of solving a problem. Right? You give it a problem and it can solve it. Maybe something that's capable of beating the very best human player at a very complex game, like Go, which of course Google recently did with its AlphaGo programme. And I just want to dive into this kind of technology for a second because the important thing to realise here is that AlphaGo is a weird alien creature. We can compare it to Deep Blue, which many years ago… at this point, it's a long time ago now, was that IBM developed artificial intelligence that famously beat Gary Kasparov, the chess grandmaster, and Deep Blue worked in a way that we understand. At the baseline, what it did was what's called a 'lookahead model'. It would just say 'here's the current state of the chess board, here's all the possible next moves that could happen, and then here's all the possible next moves that can happen after that, and here's all the possible…' and you know, at some point it hits a limit where it can't calculate anymore. But the idea is if it can look ahead enough steps, then it can pick optimal chess moves to do. So that's a very human way of thinking about playing chess. Someone actually thought 'okay, this would be a good way for a computer to play chess, and we're going to programme this computer to play chess this way' and they threw a bunch of processing power at it and it beat the chess grandmaster.

But Go was considered an unsolvable problem, or like a very hard problem. People were projecting that AI would solve Go in 2030 or something, and it happened in 2016. AlphaGo is based on this deep learning idea. It's the same kind of stuff that drives image classification algorithms, and the whole deep dream stuff is kind of like a deep learning thing pushed in reverse to create aesthetic outcomes. Broadly speaking, these things are called recurrent neural networks, and they're a special kind of neural network that has some interesting feedback mechanism built into it. AlphaGo was fed hundreds of thousands of Go games adding up to about thirty million different moves from different levels of Go player. And what AlphaGo does with these moves, is it converts them into vectors, it converts them into three hundred dimensional vectors, it converts them into layers of three hundred dimensional vectors. AlphaGo doesn't know how the game of Go is played, doesn't know the rules of Go, all we did was we gave it a bunch of Go matches and said 'draw your own conclusions about this system.' The three hundred dimensional vector is actually described as what is

called the hidden state of a recurrent neural network. And it's called a hidden state because we have no f*cking clue what the AI is doing.

There's a great blog post by Andrej Karpathy called ‘The Unreasonable Effectiveness of Recurrent Neural Networks’. I encourage you to look it up, and my next few slides are kind of pulled from that. In it, he tries to visualise how recurrent neural networks work. He'll pick a single, three hundred dimensional vector set, which is kind of like a neuron in a brain, and he'll visualise when that neuron is lit up. Sometimes this makes sense, and here's two examples. So this is a neural network that’s trained on text and text generation, and you can see there that the text that is between quotes is red and really active, and so okay, that’s a neuron that becomes active when text is in between quotes. Alright, so that’s our quote neuron, that makes sense. And then here's a neuron that gets active towards the end of a line. Okay, so that's the end of a line neuron. This is starting to make sense. But sometimes it makes no sense. Here's a typical example, like what the hell is this? What kind of parsing is this doing here? You could stare at this for a while and not understand what is going on. It seems almost random. And here's a great quote I'm just going to read out loud. Karpathy says ‘of course a lot of these conclusions are slightly hand-wavy as the hidden state of the R&N is a huge high dimensional and larger [inaudible 16:19] representation. We can see that, in addition to a large portion of cells that do not do anything interpretable, about five percent of them turn out to have learned quite interesting and interpretable algorithms.’ AKA, we don’t understand ninety five percent of what a recursive neural network is doing.

Interestingly, this is actually similar to the ratios of DNA that we can sequence in a genome, but we have no clue what it does either. I thought that was just interesting, it's just a correlation, but it seemed a little bit poetic to me in both genetics and recurrent neural networks. We have about an equal amount of not knowing what the hell is going on. These AIs can do things that we sort of recognise as human, but they're not modelled the way that humans think, which of course makes sense, because we don’t understand.

We don’t have a complete model of human cognition and consciousness and problem-solving, so of course, how could we programme a computer to be like us if we don’t even know what model we're trying to replicate? And you see, this when they interview the top level Go players that get beaten by AlphaGo, where they'll look at the moves that AlphaGo makes and they'll have no idea why AlphaGo did what it had chosen to do. And the humans actually talk about learning new strategies from AlphaGo, because AlphaGo is playing in this extremely alien way.

When we are building artificial intelligences, whether they're corporations or recurrent neural networks, we are building alien intelligences. They are among us and this has aesthetic ramifications as well as practical ones, although I think aesthetics is [inaudible 18:19]. We need to stop coding these things to try and be our friends, to try and talk like us. I think it's really important when you're designing an artificial intelligence that’s supposed to talk to a human, that you exaggerate the alienness of this intelligence. Because otherwise you're lying. You're building something that is not honest. So whenever I build Twitter bots,

	now my bots are mostly not based on recurrent neural networks, they're not based on deep learning or anything. I've done a little bit of deep learning stuff on my own, but when I create my bots, I don't put a human face on them. I don't give them a cute name. I always put the word 'bot' in the title of the Twitter account. You know it's the sorting bot, and it's just this hat that talks to you. Or two headlines. Two headlines doesn't have a personality. Its profile image is a printing press, which is a machine, and it's just called 'two headlines', and if you go to the profile it says it's a bot, and I try to forefront the botness of these things. And again, it's about confronting people with alien intelligences, so that they are prepared for what they are actually about to encounter. If we think back to the whole Microsoft Tay thing from last week, I think people came to that bot with extremely loaded expectations, and a lot of work could have been done. Just a little bit of framing work could have been done to make the output of the bot less insulting to people, in addition to all the other safeguards you could have put in; word filters and things like that. So whenever I'm talking to clients about building our bots, I'm always encouraging them to make these entities as alien as possible. And I think we have to exaggerate their alienness, otherwise we fool ourselves into thinking that they're more like us than they really are, and I think that can have, well, unforeseen consequences that might be interesting, but might be really terrifying. So thanks, that's all I've got for today.

Outro	We hope you enjoyed Darius's talk and thanks for listening. You can hear the rest of the talks from 2016 at futureeverything.org/2016podcasts.

[Transcription ends]