Audio file

119045-TheJoyofAI.mp3

Transcript

I don't know about you, but I can't remember the last time I used one of these to look something up. We all know we're in the midst of an AI revolution, changing the way we do almost everything. Instead, I use this. a computer. It's packed with AI technologies, and it's linked online to so many more. Hey, Siri, where's my nearest charity shop? The closest one I see is AG UK. Does that one sound good? It does. AI used to only be the stuff of science fiction, but not anymore.

If I speak Arabic, artificial intelligence translates into English.

I still find that utterly amazing. But I want to investigate how these extraordinary AIs actually work. It doesn't think that's a dog. It thinks it's a trombone. So what on earth is going on? I'm going to explore the surprising history of trying to turn a machine into a mind.

Check me.

Check me. And I'll find out how AI can learn for itself. Yeah! And take on challenges previously thought beyond it.

So are you more interested in reading books or watching movies?

I prefer books.

Oh, a bookworm. How nice.

Some people are fearful for humanity's future in the age of AI, but I'm not so sure. Too often, this story is told as a battle between man and machine. But for me, it's about man working with machine. So what does AI actually involve, and where will it take us? There are so many things computers can do today that we call artificial intelligence. But there's still no clear definition of what artificial intelligence actually is. Perhaps that's not entirely surprising. The aim, after all, of AI is to simulate human intelligence. And human intelligence does such an amazing range of different things. We perceive and make sense of our environment. We set goals and plan how to achieve them. We use language to communicate complex ideas. And all the time, we learn from our experiences. To get computers to do any of this, to think like humans, that is, the obvious first question you have to tackle is, what actually is thinking? If we could understand how our minds work, then perhaps we can apply this to computers. Christmas, Pittsburgh, 1955. This was the moment when two American scientists not only thought about thinking, but first worked out how to mechanize it. For my money, one of science's real eureka moments. Optimism was abundant in 1950s America, and scientists truly believed that there were very few problems they wouldn't be able to solve. Herbert Simon was a political scientist. His friend, Alan Newell, was a mathematician, and they shared a fascination with the possibilities of computers. Now, in 1955, the few computers that existed in America were mostly just used for numerical calculations. But that was about to change. As Simon would later tell it, Over Christmas, Newell and I invented a thinking machine. What inspired the two men that Christmas was the new idea from cognitive science. that our thinking process is essentially a form of computation. Inside our heads, Reason, Simon and Newell, are abstract representations of realities in the outside world. And when we think, we're performing logical processes on these abstract representations. So a dog plus a cat equals a fight. And if our minds, when we think, are computing, then perhaps, reckon Simon and Newell, computers could be programmed to think like us. For their festive fun, they picked a seriously knotty logical thinking problem. Simon owned a copy of this legendary and hefty book, Principia Mathematica, by Russell and Whitehead. which uses logic over hundreds of densely packed pages to prove the theorems and axioms of mathematics. And they wondered, could they write a computer program to automate the proofs in this book?

Seibel and Newald were particularly concerned that their models actually conform to the way that the human mind operates. They had people solve problems and sort of write notes on how they were doing it, what was going through their mind while they were doing it. And they built their computer programs to actually simulate what they were perceiving as the mental process of a human solving that problem.

Over the holidays, Simon lined up his wife and three children, together with Newell and a bunch of grad students, and gave each of them a card like these to hold. Each card represented a step in a computer program, and effectively, they all became a real-life human computer. It worked. And before long, when coded into a real computer, it solved 38 of the theorems in this monumental book. Simon and Newell called their creation the logic theorist. Human thought itself had been simulated in what's now regarded as the very first operational, artificially intelligent computer program. Within months, the phrase artificial intelligence was adopted to describe the new field. For Simon and Newell and other pioneers, getting computers to solve logical problems was a huge breakthrough. But one particular challenge had the AI boffins gripped. This battle of wits has long been regarded as the ultimate test of reasoning power. What makes chess so challenging is that there are more ways a game can develop than there are atoms in the visible universe. Neither humans nor computers can possibly consider them all. But if a human can, nonetheless, decide what to do, How can a computer be programmed to do the same? This problem fascinated the godfather of computer science himself. In 1948, the great code breaker and mathematician Alan Turing wrote this paper, containing what's considered to be a plan for the world's first chess computer program. In it, he proposed a solution to the game of chess's seemingly infinite number of options. So today I'm going to play a game as this chess program, or as it came to be known, as Churro Champ, to demonstrate what it could do. My opponent is seasoned club player Olivia. Since no computer yet existed that could actually run Turo Champ, Turing had to use pencil and paper. Even for him, this took hours. Okay, so let's play. Good luck. Fortunately, I've got a laptop. Slightly unusual start. At the start of a game, just one move each generates 400 possible combinations of play. Two, almost 200,000. And by four moves apiece, we're into the 10s of billions. This vast multiplicity of options is called the combinatorial explosion. There she goes again. Back again. Attack your queen. No computer can calculate them all. So how to give it the intelligence to make good choices? Turing wrote down a set of rules to guide the computer's search. And rules of this kind became known as heuristics, from the ancient Greek, meaning to find or discover. For example, always consider capturing an undefended piece. The program would use these heuristics to evaluate all the possible moves and counter-moves to prune down the tree of possibilities so that it only had to go down the more promising branches. What Turing had realized in doing this was that for a given problem, programmers could codify into rules their own human knowledge of how to deal with it. Oh, see, I'm in retreat now. Then if a computer followed the rules, it could solve the problem too. Given how rudimentary Turo Champ's rules are, I find myself thinking it's remarkably effective. Check me. Still, it's no match for Olivia. There we go. Congratulations. It wasn't too hard, was it? wasn't too much. Turo Champ was an elementary chess program, but the principle that heuristics, or rules, were the way to overcome the challenge of the combinatorial explosion was a sound one. And this idea was applied far beyond chess, very successfully too, as programmers tackling a wide range of real-world problems encoded their own human knowledge into increasingly complex and varied heuristics. This approach became known as classical AI, which does many clever things. In logistics, manufacturing, construction, classical AI systems, each with a set of programmed rules, are today used to plan complex operations in highly controlled environments with maximum efficiency and economy. But most of the world isn't like an ordered production line. It's much more chaotic. How is a computer to make sense of all of this? All this movement, all this noise, all this variety? I mean, I recognise these buses. I recognise a van. I can see a taxi. I even recognise those as adverts up on the screen. We instinctively know what we're looking at. But to a computer, it's just this. A torrent of raw data, a mass of numbers without meaning. How could you possibly write the rules for a computer to make sense of all this like we do? The trouble with classical AI is that the real world is messy and complex. So it's almost impossible to write the rules for a computer to even begin to make sense of its environment, let alone apply it to a task. So even the seemingly simple problem of planning to cross a road will be beyond it. But fortunately, there's another way to go with AI. Instead of us attempting to give computers the rules, the computers learn how to make sense of the data for themselves. This approach is known as machine learning, and it's machine learning that powers most of the amazing AI tools that we use today.

Why AI has become such a big thing in the last decade is because these new techniques, which are based on learning, have become very powerful. You give the systems the ability to learn for themselves directly from raw data. And these systems learn from first principles, the structure in that data, and potentially solutions to problems. So this is a very powerful new way of thinking about intelligence.

Let me show you how machine learning beats classical AI at dealing with complex data, with an example that won't get me run over. How to cope with spam. It's hard to be sure, but perhaps 400 billion spam emails are sent every day. That's something like 8 out of 10 of all emails. Without spam filters, we'd all utterly drown in junk. Those incredible one-time-only offers. Those performance enhancements. So how do you ward off what you don't want, but let in emails that you need, the non-spam, or to give it its correct, very technical term, the ham. The classical AI approach to this would be to come up with a set of rules. For example, you choose specific spammy words, and if they come up in an email's title, it gets zapped. But here's where the rules of classical AI hit their limits. Some of those spam words can also be ham words. And so emails you do want to read get junk too. And then what about this? How do you begin to write rules to catch all these? But machine learning can find patterns in all the e-mail data. to tell ham and spam apart. It first needs what's called training data. Lots of it. A heap of what you already know is ham. These are all from my inbox. And a load of what is most definitely spam. Then it can start hunting and all this stuff for the mathematical patterns. For instance, First, it identifies the most common words in each pile. For a professor of physics called Jim, there are plenty of these in the hand pile. They're very different from the words that are strictly spam only. None of these in my inbox, thank you very much. But it's because of how words appear in both piles that machine learning really comes into its own. It crunches the training data, looking for the patterns in how all these words combine, and works out how the ham are all subtly different from the spam. So, ham or spam? Today's filters are so good, we hardly get any spam at all. Though, if you are in the regular business of doing great deals on Viagra, then you'd better still check your spam bin. When it comes to simulating many of the things humans do, machine learning outdoes classical AI time and again. The reason why may lie in the workings of the brain itself. While classical AI attempts to mirror our conscious, rational thinking, machine learning may better reflect the enormous power of our subconscious minds.

We are only conscious of a small amount of what the brain does. When you open your eyes and you see a world, it happens effortlessly. We're aware of the outcome. We're aware of seeing the world, but we're not aware of the process. We're not aware of what it takes under the hood to generate this inner universe that we effortlessly experience.

In 1988, a computer scientist and roboticist called Hans Moravec, who was fascinated with the workings of the human brain, pointed out that from a human perspective, progress in artificial intelligence seemed paradoxical. You see, the things that seem difficult for our brains to cope with, things that require a lot of conscious mental effort, like chess, were proving to be relatively simple for AI. Meanwhile, the things that our brains seem to find a cinch, that we do unconsciously, like making sense of what we see, what we hear, our environment. So my ability to see where the camera is, or to hold this brain gently without dropping it, were proving to be the toughest challenges for computer programs. This became known as Moravec's paradox. Moravec reckoned it was all to do with our brain's evolution. Here's how Moravec very eloquently put it. Encoded in the large, highly evolved sensory and motor portions of the human brain is a billion years of experience about the nature of the world and how to survive in it. We're all prodigious Olympians in perceptual and motor areas. So good that we make the difficult look easy. Abstract thought, though, is a new trick, perhaps less than 100,000 years old. We've not yet mastered it. It's not all that intrinsically difficult, it just seems so when we do it. It's A neat and very convincing explanation, but it also highlights an enormously fruitful shift that was taking place in artificial intelligence. From attempts to build computer programs that mirror what our conscious minds seem to do, to ones that replicate how our brains themselves are physically structured. These remarkable and very powerful machine learning systems are called artificial neural networks. They're inspired by how real brains respond to the world. This, you'll agree, is a dog. that because your brain just fizzed with electrical and chemical signals, making their way from the eye back and up through dense layers of neurons, each one a single cell. And depending on the combined strength of the signals coming in, it either does or doesn't fire. Your brain contains something like 90 billion of these neurons. and they're networked together, often with thousands of connections each. That's at least 100 trillion connections in total. And it's this vast neural network in our brain which is brought into play to spot that spot here is something that indeed should bark.

It makes sense to try to mimic the brain to some degree. The question is, how closely do you do it? Of course, in flight, people did not build aeroplanes that had flapping wings. Rather, they understood the principles of flight, and so there are some shared features between aeroplanes and birds, but they're not direct copies. I think the same thing applies to neural networks, that it's not about replicating every last detail of a human brain or an animal brain, but trying to identify the principles by which brains work.

An artificial neural network is a virtual creation of computer software rather than a BLOB of real brain tissue. But when it's presented with our dog, actually a picture of our dog, or to be even more precise, the pixel information from a picture of our dog, the virtual neurons pass signals through the network so it too can tell what it's looking at. But first, just like dogs, and indeed spam filters, artificial neural networks must learn what to do by being trained. For this, we'll need to show it lots of Spot's friends. Each time we tell the network what it's looking at, it tweaks its connections to better recognise doggy pixel patterns. And it can learn about other things too.

It's got loads and loads of adjustable numbers. And it's, when I say loads, I don't mean hundreds, I might mean millions. So we expose it to a load of data. We show a load of cat images, a load of dog images, and we tell it which is which. and it adjusts its numbers so that when we show it new cat images and new dog images, it correctly says what's in them.

Now it's trained, the neurons in the network's inner layers first detect the simplest shapes. They then identify combinations of these shapes, doggy features. Then combinations of combinations. The more layers, the better these networks do. But remarkably, even the scientists who build these networks don't really understand how they come up with the answers.

Neural networks are radically different from, I think, any previous kind of technology. Previously, some complicated device, some complicated clock, whatever, Someone had put it together and they had a, they knew how it worked and they'd known why that piece was in there and why that joined to that. With neural networks, you can understand some of what it's doing, but then there's a load of other stuff and you have a look at it and it's frankly mysterious. You can't make any sense of it, so we don't know what it's doing there.

Whatever's going on under the bonnet... With neural networks, AI can now make much better sense of the messy real world. And with this breakthrough, its potential has increased enormously. AI is now booming. Whether it's optimizing harvests, interpreting medical images, grading students, Detecting financial opportunities. Neural networks are mastering new tasks in all parts of our lives. Take transport. And the AI application we're often told is just around the corner. Driverless cars. This British company is busy developing and running this tech in a range of different vehicles on real roads, not just test tracks. Oxpotica took me for a spin. As we drive, I begin to realize just how much this kind of AI is going to revolutionize the way we live. It's actually remarkable how safe I feel. you very quickly trust that it knows what it's doing.

Every fraction of a second, the car runs simulations of what the world might look like. And it simulates lots of possible outcomes. Well, if I drove that way, what would this look like? If I drove that way, what would this look like? And it generates... Thousands of simulations, 50 times a second. So continually updating, what if I did this, what if I did this? Then evaluating them and choosing the best one. So it tries to do the least worst thing the whole time.

Feeding into these simulations is a continuous stream of data from the car's onboard lasers and cameras. The laser data gives it a 3D model of everything around it. Any object that's moving or might move is located and tracked. Then with camera data, it identifies what these objects are, and so how they might behave. The AI in the car doesn't need to communicate with any other computer. It's entirely self-contained. And all this is only possible thanks to neural network systems that learn from their driving experiences. All this comes together on my drive as the car negotiates a sudden moment of high drama on the highway. Oh, very good. So here's the classic driverless car situation. A woman crossing zebra crossing who stepped out. just as we were coming up to the crossing, and it stopped. Are we going to get to the point where I've got a driverless car, therefore I'm going to have a nap? That is completely safe, I'll leave it entirely up to the car.

We will. I'm absolutely sure we will. I think the vehicles that you can sit in and they will drive you around parts of a city, part of an airport, campus, that's coming quite soon. The vehicle that has the same functionality as your car does now, that can get you from anywhere to anywhere, any weather, any time of day without having any difficulties and total confidence that you're going to get there. And you can buy that from a forecourt and you don't need a steering wheel and you don't even need a driving licence. In fact, it may not have any windows Long. time away. But I don't think there's anything that's unattainable about humans driving that a machine can do. To that argument, I think you would have to, it hits, and there's something that's not computable about driving, and that doesn't seem particularly realistic to me.

With AI muscling in on ever more of what we do ourselves, it's no wonder many worry about how the AI revolution might change our lives. Revolutions make people nervous, especially when they're not the ones in control. Probably the biggest fear is that AI might take people's jobs, and they might never find work again. One of the concerns of AI is that it is leading to this huge technological revolution that is going to affect society. Yes. I mean, we can't stop it, we can't mitigate against it.

And nor can we deny that there's change coming. But I think we can now look ahead and go, new jobs are coming in the way that new jobs came because of computers. And think how many people have jobs that are now only doable because they have a computer, have become possible or were invented because you have computers. So I can't deny that there's transformations coming, but I'm, if you like, almost pathologically positive that it's going to make us healthier and wealthier and enhance our capabilities and change jobs in the way that computing did as well.

You know, you look at civilization around us, that's all a product of intelligence. And I think of AI as a powerful tool, perhaps the most powerful tool of all, that will allow us to reach the full potential of humanity.

We're still a long way off this brave new world. And to get there, we'll need even cleverer AI. But that's what Demis Hassabis and his colleagues are dreaming up at DeepMind, the Blue Sky AI Research Division of a leading search engine provider. Here they're trying to develop neural network systems that can learn to do anything without any human intervention. You know what their mission statement is? Solve intelligence and then use it to solve everything else. That's ambitious, you'll agree. And they're going about it in a rather intriguing way.

The idea is that we first test and develop AI algorithms so that they can master games. But then our hope is, if we do that in a general enough way, they'll be able to be used in the real world for serious problems.

And it turns out they've got a real thing here for retro Atari classics when it comes to testing what an AI could learn to do for itself. Presumably, you're having to teach your AI the rules of the game so that it can learn how to play.

No, we don't. It learns really only from its experience. All it's seeing is those pixels and whether or not its score increased or not. And then trying to solve the puzzle of, well, my score got better then. What was the action that I took? And that's really just done through a learning algorithm that changes all of the millions of connections in this neural network. to say, let's reinforce this action, or let's not reinforce this other action. So while we could program up some rules that said, here's the brick, here's the ball, here's the paddle, and here's how you move it, we don't do any of that. We simply let the algorithm learn on its own.

So how quickly does it learn and improve?

So after about 300 games, we see that we can get to human-level performance. But the nice thing about an AI algorithm is we can just let it run. And so we let it keep on training for a few more 100 games, and then we see that it does get to superhuman performance.

Well, that-- Let's take a look at that.

I want to see that. Sure. So at the beginning of the game, it's moving back and forth. It's hitting the ball back. But as the game progresses, then the ball is going to move faster and faster. This is where humans stop being able to return. But the algorithm discovered a really interesting strategy, and we weren't expecting to see this. We had no idea. So it was really exciting to see what it's doing now, which is what we call tunneling. It has managed to systematically hit the ball only to one side, and that means that it breaks through to the top, bounces around, maximum reward, less risk of dying, of losing the game.

So that's a strategy that it figured out for itself, because it could see that would... give it a huge advantage.

It managed to discover this absolutely on its own.

Variants of this AI, neural networks learning entirely by themselves, have gone on to reach superhuman level on over 40 different Atari games. What's remarkable isn't just the AI learning so rapidly and successfully, it's how it discovers its own strategies for success. But could a neural network AI even discover things that we don't know of? In 2016, DeepMind's programmers created an AI system that taught itself to master the ancient game of Go. In Go, players battle for control of territory. And although the rules are simple, it's nonetheless an enormously complex game, where players need to rely on their intuitive sense of pattern.

Whereas chess might be 50% about intuition and 50% about calculation, Go is more like 90% intuition, 10% calculation.

DeepMind built a neural network system called AlphaGo. Trained by playing millions of games against itself, it was able to capture the intuitive, almost unconscious pattern recognition ability that human Go players have. Confidence of AlphaGo's powers, DeepMind challenged rather greatest Go players in the world, Lee Sedol, to a very public five-game tournament. Nobody outside of Deep Mind thought that he would lose a single one. In the end, AlphaGo beat him four games to one. But the most significant moment came in game two, when AlphaGo played a move no human player would have even considered.

That's a very, that's a very surprising move. I thought, I thought it was a mistake. At that point, we didn't know, was it just a, you know, a useless move, or was it actually a brilliant move? So coming on top of a fourth limestone is really unusual. And in fact, Lisa Dole, when confronted with move 37, his jaw dropped visibly, and he thought for like 20 minutes. So, at the very least, we knew this was a shocking move.

Remarkably, not one of the humans watching understood why AlphaGo had done what it did.

It turned out to be decisive in that game. About 100 moves later, a battle in another part of the board ended up perfectly connecting up with the piece that was played on move 37. Lisa Doll commented afterwards that when he saw that move, he realized that this was a different type of machine, that it wasn't just regurgitating human knowledge or memorizing positions. In some sense, it was actually creating new ideas. Oh, he resigned. He's done. OK. Wow. Wow. Yeah.

The AI had made a genuine discovery, one with profound implications.

It showed that these types of learning systems can actually come up with a new idea that hadn't been searched or thought about before by humans. And what's amazing is that can happen in Go, which we've played for thousands of years. Then how much potential has this kind of system got in other areas like science and medicine? I think with these powerful tools, we're going to enter a golden era of scientific discovery.

And yet... A computer that can outgun the top human with strategies it's intuited by itself is unnerving. And it begs a big question. What if, one day, scientists manage to create an AI that rivals or exceeds the full range of what human intelligence can do? The idea of a computer that not only outstrips our intelligence, but that also slips dangerously out of our control is a staple of science fiction. For his film, 2001, A Space Odyssey, which was made in the late 1960s, director Stanley Kubrick created one of the most chilling realizations of this idea ever seen, the HAL 9000 supercomputer. In the film, HAL, in its own words, foolproof and incapable of error, starts acting in unexpected and disturbing ways. The astronauts on board its spaceship make plans to deactivate it. And when it finds out, it attempts to kill them all, and very nearly succeeds. HAL wasn't malevolent, just remorselessly logical. The astronauts would have stopped it from completing its mission, and so, of course, they had to be eliminated. Hal wasn't entirely dreamt up by Kubrick and co-writer Arthur C. Clarke. It was also inspired by the work of a British computer scientist called Jack Good, who was a veteran of Alan Turing's code-breaking effort at Bletchley Park during World War II. Jack Good had laid out a startling vision of the future of artificial intelligence. in an essay entitled Speculations Concerning the First Ultra-Intelligent Machine. Let an ultra-intelligent machine be defined as a machine that can far surpass all the intellectual activities of any man, however clever. Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines. There would then unquestionably be an intelligence explosion. and the intelligence of man will be left far behind. This intelligence explosion identified by Jack Good might well have been for the benefit of all humankind, but what must have grabbed Kubrick's attention was the sting in the tail. The first ultra-intelligent machine is the last invention that man need ever make, provided that the machine is dosed on enough to tell us how to keep it under control.

King Midas said, I want everything I touch to turn to gold, and the gods gave him exactly what he asked for. So his food turned to gold, his water turned to gold, his wine turned to gold, his daughter turned to gold. We do not know how to say precisely what we want. And if you have a super intelligent machine, it's kind of like a god. It will find some way of giving you your objective. in ways that you didn't expect. And so we've got to figure out a way that guarantees that we retain control forever over things that are much more intelligent than us.

A super intelligent AI is an alarming thought. But in reality, it's not coming anytime soon. So we've plenty of time to work out how to control one. The AI behind so many of today's amazing breakthroughs is still fundamentally limited. It can find patterns in complex data, often better than we can. But it can't yet convert these into the kind of meaningful conceptual thinking that's so crucial to our intelligence. Let me show you, using, yes, our furry four-legged friends again. I've had a state-of-the-art neural network installed on this tablet. It's been trained to identify over 1,000 different kinds of animals and objects, using over a million examples. Now, it hasn't been trained on these pictures. It's seeing them for the first time. First up, a classic portrait of a dog. Right, not only has it recognised it as a dog, but it's pretty certain it's a Brittany Spaniel. Well done, network. Right, Let's try it on something slightly harder, because this isn't a classic portrait of a dog. Can't even see its face clearly. So let's see how well it does. It's pretty sure it's a Whippet. Now I'm pretty sure that's a Staffy, but still very good. And again, this one is not a classic portrait. Could be a dingo. There's even 3% chance it's a lion. But still not bad. So, three out of three for the neural network. But, and it's a big but, it doesn't have any real understanding of what it's looking at. Let me show you what I mean with these three pictures. Now, they look pretty identical to the first three. Right, so picture number one... It's 100% sure that's a tabby. It thinks this dog is a cat. OK. Picture number two. And this one, it sure is a baboon. Right, picture number three. And it doesn't think that's a dog, it thinks it's a trombone. So what on earth is going on? Well, let me tell you. see, each of these three pictures has been altered ever so slightly by adding a few pixels. On the left is the original. On the right, with the additional pixels chosen specifically to fool it. The neural network doesn't see that overall it's still a dog. It only responds to pixel patterns without understanding what they all add up to.

On the one hand, the networks perform incredibly well. You can show them an image they've never seen before and they'll get it right. But on the other hand, it's incredibly fragile. I can just tweak any image and now it gets it wrong. So it really raises a challenge. And you think, well, I'm not sure I understand how this thing works at all. if it can be knocked off so easily.

We can be fooled by optical illusions ourselves, of course. But with a neural network, it's very hard to understand what makes it decide it's looking at a dog, or a cat, or a baboon, or a trombone, whether it's right or wrong. The network is undoubtedly, in some sense, intelligent. But at the same time, there's no understanding of concepts there. It doesn't actually know what a dog is, let alone anything else. which is why it can be fooled with just a few pixels. And that's where all AI is today, capable of finding patterns in data with astonishing detail and sensitivity, but with no real understanding of what those patterns actually mean. Given this, the hardest challenge of all for AI must surely be that uniquely human ability that relies on our understanding of concepts. Language. So how well can an artificial intelligence today simulate real human-to-human conversation? I've come to a top robotics lab in Edinburgh, where they're working on how to make an AI that can behave like an assistant, companion, or even a friend. I'll be talking to a cutting-edge conversational artificial intelligence, or chatbot, called Alana. Usually, Alana's just a disembodied voice, but today, especially for me, Alana will be routed through a rather cute robot.

Hello, good morning. How are you doing? Right, let's get to know one another. So, Jim, are you more into sci-fi or fantasy?

I prefer sci-fi.

Awesome. Personally, I love Star Wars. But back to you, what is a sci-fi movie that you enjoy?

My favorite sci-fi movie is Blade Runner.

There are two Blade Runner movies, Blade Runner and Blade Runner 2049.

I prefer the original. What is going on there, then, when it's thinking about different responses and different options?

So, yeah, what's really happening is that there's several different programs running in parallel, and they're all listening to what you're saying, and they're all volunteering things that might be good things to say next. Inside Alana, we have a persona bot, which has a kind of personality with favorite movies and stuff like that in it. We have a Wikipedia bot, we have a news bot, and we have a rapport bot, and what that's trying to do is to find out about you and your personal interests.

So it's almost like a whole committee of programs, all... Yeah, that's absolutely right.

So it's got, it's like a whole bunch of different voices all sort of competing to say what they want to say. And then what we have is a ranking function, which is trained using machine learning over thousands and thousands and thousands of previous conversations that we've had. What it's trying to do is choose the one response which is most likely to lead to a best conversation in the future.

The aim of Alana is to handle general chit-chat, like an AI you could enjoy a pint with. It's ambitious, and Alana's ability to generate a conversation really is impressive.

I think Roger Federer is awesome, so I am wondering...

But it isn't quite there yet.

Okay, since you like sports, what is a sport that you like to watch?

I prefer football.

Great. Which player do you like?

Any player who plays for my favourite team, which is Leeds United.

Okay, I heard you talk about Leeds United. I think that Leeds United is a 1974 television play shown on October 31st, 1974 on BBC. The film is based on an unofficial strike in Leeds in February 1970. Shall I go on?

I'm really not sure whether I want Alana to go on here or not.

This is actually an interesting example of an AI problem in, figuring out what is the user actually talking about. Didn't really have enough common sense knowledge to realise that when you say Leeds United, that's really normally talking about a football team. So it turns out, you know... There are many possible things you could be talking about when you say Leeds United. So what it did is it looked in Wikipedia for anything called Leeds United.

But it highlights what we as humans would regard as quite a simple problem, just how complex it is to teach a machine. Well, I'll forgive Alana for that one. Let's see if I carry on.

Okay, what is a video game that you played recently?

Well, I haven't played a video game For about 20 years, the last one I played was Gran Turismo.

Regarding Gran Turismo, I heard on the radio that Gran Turismo Sport is now scheduled to be released October 17th on PlayStation 4. Should I tell you more?

No, thank you, Elana, but that's very interesting.

You are very kind.

So what you actually said is you haven't played the game for 20 years. So a human would really infer from that you don't really want to know lots about this game and this kind of stuff. And this is the kind of thing that makes language research so incredibly fascinating, but also incredibly challenging, because there's a lot more going on than just the words that you said. So we are actually pretty far away from having the kind of conversations you see in science fiction films.

Speaking of sci-fi...

Elana, if you don't mind, I want to talk to Oliver for a moment.

You can ask me about my favorites if you like. What things do you like?

No, I'm talking to Oliver now. What is actually going on there for Oliver?

I saw this on the news, had a headline, Star Wars, the last channel.

It wasn't the best chat I've ever had, but free-flowing conversation like this is still a real achievement for AI.

Shall I say some more about this?

So where does AI go from here? Getting to the next breakthrough may be inspired by studying not so much what adult humans do as infants. Somewhere between 18 months and two years old, children start doing something remarkable. Show them how to do something just a few times, often even just once, and they start practicing it for themselves. This is called one-shot learning. For computer scientists who have to train even the most sophisticated AIs on hundreds or thousands of examples before they learn anything, this is like the holy grail. Anyone who works in artificial intelligence will appreciate just how advanced these little humans really are. They navigate a complex 3D world. They grasp basic physics like gravity and inertia. They formulate plans and carry them out.

I think you can see how nascent AI is even still today, even with all of the successes it's had, because when you see all the amazing things that a toddler learns, our AI systems are nowhere near the capabilities even of a, you know, a two-year-old.

The foundation of their amazing capabilities is how much they've learned as babies. Since birth, they've continually explored and experimented, drinking in information every second they're awake. These little children learn directly from data and experience, rather like computers do with machine learning and artificial neural networks. But they also understand the world with abstract concepts. It's the combination of the two. the way their learning seamlessly produces the concepts, and the way the concepts then direct their learning that makes them like the most amazing computers you can imagine. It's this combination that AI researchers are one day hoping to crack. AI is developing fast, no longer just relying on programmers telling it the rules, It's learning to do amazing things by itself, faster and sometimes even better than we can. What's more, it's started to discover ways of doing things we didn't know about. But it's not yet advanced enough to really learn or think like we do. Still, if it could one day rival all our abilities, I wonder if it might become like us in another way too. Could an artificial intelligence ever have real emotions? Could it be happy, sad, or jealous? Could it be social, or feel friendship, even love? In short, could it become conscious? Now, I don't believe there's any magic pixie dust that we have to sprinkle over the gray matter in our heads to bring about consciousness. There's nothing our brains do that couldn't, in principle, be replicated. And if AI does one day become conscious, we will also have to treat it well. Not because if we didn't, it might rise up and destroy us, but more profoundly because it will be the right thing to do. Perhaps one day, we'll even feel it would be cruel to switch a computer off. We need to use AI wisely. And that goes for now as well as in the future. But if we do, I think humanity has little to fear and a huge amount to gain. I feel inspired by what AI can already do today, and I believe that through AI, we'll greatly extend our own capacities, changing our lives in ways we can't yet imagine. The evolution of machines that think must surely be one of the greatest developments in human history.

On the topic of books, I love, Do Androids Dream of Electric Sheep?

Yes, that's one of my favorites too. Philip K. ****.

That's not appropriate.

Alana, do you know any jokes?

A restaurant nearby had a sign in the window which said, we serve breakfast at any time. So I ordered French toast in the Renaissance.