I want to understand the big questions, the really big ones, that you normally go into philosophy or physics if you're interested in them. I thought building AI would be the fastest route to answer some of those questions. Why did you think that? Well, when I was at a university in the United States, I had a friend who was doing a PhD in physics, and he was doing a PhD in physics, and he was doing a PhD in physics, and he was doing a PhD in physics, and he was doing a PhD in physics, well, I guess when I was a kid, my favorite subject was physics, and I was interested in all the big questions, fundamental nature of reality, what is consciousness, all the big ones. And usually you go into physics if you're interested in that. But I read a lot of the great physicists, some of my all-time scientific heroes like Feynman and so on, and I realized in the last 20, 30 years, we haven't made much progress in understanding some of these fundamental laws. So I thought, why not build the ultimate tool to help us, which is artificial intelligence, and at the same time, we could also maybe better understand ourselves and the brain better by doing that, too. So not only was it an incredible tool, it was also useful for some of the big questions itself. CAWTHORNE WILLIAMS, JR.: Super interesting. So obviously AI can do so many things, but I think for this conversation, I'd love to focus in on this theme of what it might do to unlock the really big questions, the giant scientific breakthroughs, because it's been such a theme driving you and your company. Yeah. So, I mean, one of the big things AI can do and I've always thought about is, we're getting, you know, even back 20, 30 years ago, the beginning of the internet era and computer era, the amount of data that was being produced and also scientific data, just too much for the human mind to comprehend in many cases. And I think one of the uses of AI is to find patterns and insights in huge amounts of data and then surface that to the human scientist to make sense of and make new hypotheses and conjectures. So it seems to me very compatible with the scientific method. CAWTHORNE WILLIAMS, JR.: Right. But gameplay has played a huge role in your own journey in figuring this thing out. Who is this young lad on the left there? Who is that? So that was me, I think it must have been about nine years old. I'm captaining the England under-11 team and we're playing in a four-nations tournament, that's why we're all in red. I think we're playing France,

Scotland and Wales, I think it was. CAWTHORNE WILLIAMS, JR.: That is so weird. Because that happened to me too. CAWTHORNE WILLIAMS, JR.: Sure. CAWTHORNE WILLIAMS, JR.: In my dreams. CAWTHORNE WILLIAMS, JR.: Right. CAWTHORNE WILLIAMS, JR.: I mean, this is, OK. And it wasn't just chess, you loved all kinds of games. CAWTHORNE WILLIAMS, JR.: I loved all kinds of games, yeah. CAWTHORNE WILLIAMS, JR.: And when you launched DeepMind, having it tackle gameplay, why? CAWTHORNE WILLIAMS, JR.: Well, look, games actually got me into AI in the first place, because while we were doing things like, we used to go on training camps with the England team and so on, and actually back then, I guess it was in the mid-'80s, we would use the very early chess computers, if you remember them, to train against, as well as playing against each other. And they were big lumps of plastic, you know, physical boards that you used to, some of you remember, used to actually press the squares down, and little LED lights came on. And I remember actually not just thinking about the chess style, I was actually just fascinated by the fact that this lump of plastic, someone had programmed it to be smart and actually play chess to a really high standard, and I was just amazed by that. And that got me thinking about thinking, and how does the brain come up with these thought processes, these ideas, and then maybe how we could mimic that with computers. So, yeah, it's been a whole theme for my whole life, really. But you raised all this money to launch DeepMind, and pretty soon, you were using it to do, for example, this. So, I mean, this is an odd use of it. What was going on here? Well, we started off with games at the beginning of DeepMind, this is back in 2010, so this is about 10 years ago, it was our first big breakthrough, because we started off with classic Atari games from the 1970s, the simplest kind of computer games there are out there. And one of the reasons we use games is they're very convenient to test out your ideas and your algorithms, they're really fast to test. And also, as your systems get more powerful, you can choose harder and harder games. And this was actually the first time ever that our machine surprised us, the first of many times, which it figured out in this game called Breakout that you could send the ball around the back of the wall, and actually, it'd be a much safer way to knock out all the tiles of the wall. So it's a classic Atari game

there, and that was our first real aha moment. CA1 So this thing was not programmed to have any strategy, it was just told, try and figure out a way of winning, you just move the bat at the bottom and see if you can find a way of winning. CA2 Right, it was a real revolution at the time, this was in 2012, 2013, where we coined these terms, deep reinforcement learning, and the key thing about them is that those systems were learning directly from the pixels, the raw pixels on the screen, but they weren't being told anything else. So they were being told, maximize the score, here are the pixels on the screen, 30,000 pixels. The system has to make sense on its own from first principles, what's going on, what it's controlling, how to get points. And that's the other nice thing about using games to begin with, they have clear objectives to win, to get scores, so you can kind of measure very easily that your systems are improving. CA1 But there was a direct line from that to this moment a few years later, where the country of South Korea and many other parts of Asia, and in fact, the world, went crazy over what? CA2 Yeah, so this was the pinnacle of, this was in 2016, the pinnacle of our games-playing work, where we'd done Atari, we'd done some more complicated games, and then we reached the pinnacle, which was the game of Go, which is what they play in Asia instead of chess, but it's actually more complex than chess. And the actual brute force algorithms that were used to kind of crack chess were not possible with Go, because it's a much more pattern-based game, much more intuitive game. So even though Deep Blue beat Garry Kasparov in the 90s, it took another 20 years for our program, AlphaGo, to beat the world champion at Go. And we always thought, myself and the people working on this project, for many years, that if you could build a system that could beat the world champion at Go, it would have had to have done something very interesting. And in this case, what we did with AlphaGo is it basically learned for itself by playing millions and millions of games against itself, ideas about Go, the right strategies, and in fact, invented its own new strategies that the Go world had never seen before, even though we've played Go for more than 2,000 years. It's the oldest board game in existence. So it was pretty astounding, not only did it win the match, it also came up with brand-new strategies. And you continued this with a new strategy of not even really teaching anything about Go, but just

setting up systems that, just from first principles, would play so that they could teach themselves from scratch Go or chess. Talk about AlphaZero and the amazing thing that happened in chess. Yeah, so following this, so AlphaGo started, we started with AlphaGo by giving it all of the human games that had been played on the internet. So it started that as a basic starting point for its knowledge. And then we wanted to see what would happen if we started from scratch, from literally random play. So this is what AlphaZero was, that's why it's the zero in the name, because it started with zero prior knowledge. And the reason we did that is because then we would build a system that was more general. So AlphaGo could only play Go, but AlphaZero could play any two-player game. And it did it by playing initially randomly and then slowly, incrementally improving. Well, not very slowly, actually, within the course of 24 hours, going from random to better-than-world-champion level. And so this is so amazing to me. So I'm more familiar with chess than with Go, and for decades, thousands and thousands of AI experts worked on building incredible chess computers. Eventually, they got better than humans. You had a moment a few years ago where, in nine hours, AlphaZero taught itself to play chess better than any of those systems ever did. Talk about that. Yeah, it was a pretty incredible moment, actually. So we set it going on chess, and as you said, there's this rich history of chess and AI where there were these expert systems that had been programmed with these chess ideas, chess algorithms. And you start, you have this amazing, I remember this day very clearly, where you sort of sit down with the system, starting off random, you know, in the morning, you go for a cup of coffee, you come back. I can still just about beat it by lunchtime, maybe just about, and then you let it go for another four hours, and by dinner, it's the greatest chess-playing entity that's ever existed. And, you know, it's quite amazing, like, looking that live on something that you know well, you know, like chess, and you're expert in, and actually just seeing that in front of your eyes. And then you extrapolate to what it could then do in science, or something else, which, of course, games were only a means to an end. They were never the end in themselves. They were just the training ground for our ideas and to make quick progress in a matter of, you know, less than five years, actually, it went from

Atari to Go. CAWTHORNE. Yeah. I mean, this is why people are in awe of AI and also kind of terrified by it. I mean, it's not just incremental improvement. The fact that in a few hours, you can achieve what millions of humans over centuries have not been able to achieve, that is just, that gives you pause for thought. I mean, it's ... It does. I mean, it's a hugely powerful technology. It's going to be incredibly transformative, and we have to be very thoughtful about how we use that capability. CAWTHORNE. So talk about this use of it, because this is, again, this is another extension of the work you've done, where now you're turning it into something incredibly useful for the world. What are all the letters on the left and what's on the right? This was always my aim with AI from a kid, which is to use it to accelerate scientific discovery. And actually, ever since doing my undergrad at Cambridge, I had this problem in mind one day for AI. It's called the protein folding problem, and it's kind of like a 50-year-grand challenge in biology. And it's very simple to explain. Proteins are essential to life. They're the building blocks of life. Everything in your body depends on proteins. And you describe a protein sort of described by its amino acid sequence, which you can think of as roughly the genetic sequence describing the protein. So that's all the letters. CAWTHORNE. And each of those letters represents in itself a complex molecule. CAWTHORNE. That's right, each of those letters is an amino acid, and you can think of it as a kind of string of beads there at the bottom left, right? But in nature, in your body or in an animal, this string, a sequence, turns into this beautiful shape on the right. That's the protein. Those letters describe that shape. And that's what it looks like in nature. And the important thing about that 3D structure is that the 3D structure of the protein goes a long way to telling you what its function is in the body, what it does. And so the protein folding problem is, can you directly predict the 3D structure just from the amino acid sequence? So literally, if you give the machine, the AI system, the letters on the left, can it produce the 3D structure on the right? And that's what AlphaFold does, our program does. CAWTHORNE. It's not calculating it from the letters, it's looking at patterns of other folded proteins that are known about, and somehow learning from those patterns that this may be the way to do this. So when we started this project, actually straight after AlphaGo, I thought we were

ready. Once we'd cracked Go, I felt we were finally ready after almost 20 years of working on this stuff to actually tackle some scientific problems, including protein folding. And what we start with is, painstakingly over the last 40-plus years, experimental biologists have pieced together around 150,000 protein structures, using very complicated X-ray crystallography techniques and other complicated experimental techniques. And the rule of thumb is that it takes one PhD student, their whole PhD, so four or five years, to uncover one structure. But there are 200 million proteins known to nature. So it would just take forever to do that. And so we managed to actually fold, using AlphaFold, in one year, all those 200 million proteins known to science. So that's a billion years of PhD time saved. And so it's amazing to me just how reliably it works. I mean, on the shows, you know, here's the model, and you do the experiment, and sure enough, the protein turns out the same way. Times 200 million. Yeah, and the more you deeply go into proteins, you just start appreciating how exquisite they are. I mean, look at how beautiful these proteins are. And each of these things do a special function in nature, and they're almost like works of art. And it still astounds me today that AlphaFold can predict, the green is the ground truth and the blue is the prediction, how well it can predict this, to within the width of an atom, on average, is how accurate the prediction is, which is what is needed for biologists to use it and for drug design and for disease understanding, which is what AlphaFold unlocks. You made a surprising decision, which was to give away the actual results of your 200 million proteins. We open-sourced AlphaFold and gave everything away on a huge database with our wonderful colleagues at the European Bioinformatics Institute. Thank you. Thank you. Thank you. I mean, you're part of Google. Was there a phone call saying, Demis, what did you just do? Well, you know, I'm lucky we have very supportive, Google are really supportive of science and understand the benefits this can bring to the world. And, you know, the argument here was that we could only ever have even scratched the surface of the potential of what we could do with this. Maybe like a millionth of what the scientific community is doing with it. There's over a million and a half biologists around the world have used AlphaFold in predictions. Almost every biologist in the world is making use of this now. Every pharma company.

So we'll never know, probably, what the full impact of it all is. But you're continuing this work in a new company that's spinning out of Google called Isomorph. Isomorphic, yeah. Isomorphic. Give us just a sense of the vision there. What's the vision? So AlphaFold is a sort of fundamental biology tool. Like, what are these 3D structures, and then what might they do in nature? And then, you know, the reason I thought about this and was so excited about this is that this is the beginnings of understanding disease and also may be helpful for designing drugs. So if you know the shape of the protein, and then you can kind of figure out which part of the surface of the protein you're going to target with your drug compound. And Isomorphic is extending this work we did in AlphaFold into the chemistry space, where we can design chemical compounds that will bind exactly to the right spot in the protein and also, importantly, to nothing else in the body. So it doesn't have any side effects, and it's not toxic and so on. And we're building many other AI models, sort of system models to AlphaFold to help make predictions in chemistry space. So we can expect to see some pretty dramatic health medicine breakthroughs in the coming few years. I think, yeah, we'll be able to get down drug discovery from years to maybe months. OK, Demis, I'd like to change direction a bit. Our mutual friend Liv Borre gave a talk last year at TED.ai that she called the Moloch Trap. The Moloch Trap is a situation where organizations, companies in a competitive situation can be driven to do things that no individual running those companies would by themselves do. And it's felt, I was really struck by this talk, and it's felt, as a sort of layperson observer, that the Moloch Trap has been shockingly in effect in the last couple of years. So here you are, with DeepMind, sort of pursuing these amazing medical breakthroughs and scientific breakthroughs, and then suddenly, kind of out of left field, OpenAI with Microsoft releases ChatGPT, and the world goes crazy and suddenly goes, holy crap, AI is, you know, everyone can use it. And there's a sort of um ... It felt like the Moloch Trap in action. I think Microsoft CEO Satya Nadella actually said, Google is the 800-pound gorilla in the search space. We wanted to make Google dance. Um, how ... And it did, Google did dance. There was a dramatic response. Your role was changed. You took over the whole Google AI effort. Products were rushed out. Um, you know, Gemini, part amazing, part

embarrassing. I'm not going to ask you about Gemini, because you've addressed it elsewhere. But it feels like this was the Moloch Trap happening, that you and others were pushed to do stuff that you wouldn't have done without this sort of catalyzing, competitive thing. Meta did something similar as well. They rushed out open-source versions of AI, which is arguably a reckless act in itself. This seems terrifying to me. Why? Is it terrifying? Look, it's a complicated topic, of course, and first of all, I mean, there are many things to say about it. First of all, we were working on many large-language models, and in fact, obviously, Google Research actually invented Transformers, which was the architecture that allowed all this to be possible five, six years ago. And so we had many large models internally. The thing was, I think what the ChatGPT moment did that changed and fair play to them to do that was they demonstrated, I think somewhat surprisingly to themselves as well, that the general public were ready to embrace these systems and actually find value in these systems. Impressive though they are, I guess when we're working on these systems, mostly you're focusing on the flaws and the things they don't do and hallucinations and things you're all familiar with now. We were thinking, would anyone really find that useful, given that it does this and that and the other? And we wanted to improve those things first before putting them out. But interestingly, it turned out that even with those flaws, many tens of millions of people still find them very useful. And so that was an interesting update on maybe the convergence of products and the science that actually all of these amazing things we've been doing in the lab, so to speak, are actually ready for prime time for general use beyond the rarefied world of science. And I think that's pretty exciting in many ways. CAWTHONER-JOHNSON So at the moment, we've got this exciting array of products which we're all enjoying, and all this generative AI stuff is amazing. But let's roll the clock forward a bit. Microsoft and OpenAI are reported to be building or investing, like, 100 billion dollars into an absolute monster database supercomputer that can offer computers at orders of magnitude more than anything we have today. I think it takes, like, five gigawatts of energy to drive this, it's estimated. That's the energy of New York City to drive a data center. So we're pumping all this energy into this giant, vast brain. Google, I presume, is going to match this type of

investment, right? CAWTHONER-JOHNSON Well, yeah. I mean, we don't talk about our specific numbers, but I think we're investing more than that over time. And that's one of the reasons we teamed up with Google back in 2014, is we knew that in order to get to AGI, we would need a lot of compute, and that's what's transpired, and Google had and still has the most computers. CAWTHONER-JOHNSON So Earth is building these giant computers that are going to power so much of the future economy. And it's by companies that are in competition with each other. How will we avoid the situation where someone is getting a lead? Someone else has got 100 billion dollars invested in their thing. Isn't someone going to go, wait a sec, if we used reinforcement learning here to maybe have the AI tweak its own code and rewrite itself and make it so powerful, we might be able to catch up in nine hours over the weekend with what they're doing. Roll the dice, dammit, we have no choice, otherwise we're going to lose a fortune for our shelters. How are we going to avoid that? JGWINN NICHOLS Yeah, well, we must avoid that, of course, clearly. And my view is that as we get closer to AGI, we need to collaborate more, and the good news is that most of the scientists involved in these labs know each other very well, and we talk to each other a lot at conferences and other things. And this technology is still relatively nascent, so probably it's OK what's happening at the moment. But as we get closer to AGI, I think as a society, we need to start thinking about the types of architectures that get built. So I'm very optimistic, of course, that's why I spent my whole life working on AI and working towards AGI, but I suspect there are many ways to build the architecture safely, robustly, reliably and in an understandable way, and I think there are almost certainly going to be ways of building architectures that are unsafe or risky in some form. So I see a kind of bottleneck that we have to get humanity through, which is building safe architectures as the first types of AGI systems, and then after that, we can have a flourishing of many different types of systems that are perhaps sharded off those safe architectures that ideally have some mathematical guarantees, or at least some practical guarantees, around what they do. CA1. Do governments have an essential role here to define what a level playing field looks like and what is absolutely taboo? CA2. Yeah, I think it's not just about ... Actually, I think government and civil

society and academia and all four parts of society have a critical role to play here to shape, along with industry labs, what that should look like as we get closer to AGI and the cooperation needed and the collaboration needed to prevent that kind of runaway race dynamic happening. CA1. OK. Well, it sounds like you remain optimistic. What's this image here? CA2. Yeah. It's one of my favorite images, actually. I call it the tree of all knowledge. So we've been talking a lot about science, and a lot of science can be boiled down to ... If you imagine all the knowledge that exists in the world as a tree of knowledge, and then maybe what we know today as a civilization is some small subset of that. And I see AI as this tool that allows us, as scientists, to explore potentially the entire tree one day. And we have this idea of root node problems that, like AlphaFold, the protein folding problem, where if you could crack them, it unlocks an entire new branch of discovery or new research. And that's what we try and focus on at DeepMind and Google DeepMind to crack those. And if we get this right, then I think we could be in this incredible new era of radical abundance, curing all diseases, spreading consciousness to the stars, you know, maximum human flourishing. CA1. We're out of time, but what's the last example of, like, in your dreams, this dream question that you think there is a shot that in your lifetime AI might take us to? CA2. Well, I mean, once AGI is built, what I'd like to use it for is to try and use it to understand the fundamental nature of reality. So do experiments at the Planck scale, you know, the smallest possible scale, theoretical scale, which is almost like the resolution of reality. CA1. You know, I was brought up religious, and in the Bible, there's a story about the tree of knowledge that doesn't work out very well. Um ... Is there any scenario where we discover knowledge that the universe says, humans, you may not know that? CA2. Potentially, I mean, there might be some unknowable things, but I think scientific method is the greatest sort of invention humans have ever come up with. You know, the Enlightenment and scientific discovery, that's what's built this incredible modern civilization around us and all the tools that we use. So I think it's the best technique we have for understanding the enormity of the universe around us. CA1. Well, Demis, you've already changed the world. I think probably everyone here will be cheering you on in your efforts to ensure that we continue to

accelerate in the right direction. Thank you. Demis, how's that? Thank you. Thank you. That was really excellent. Thank you. Thank you.