Welcome to the rest of this politics leading with me, Rory Stewart and me, Alistair Campbell.
And our guest today is Mustafa Suleiman and Mustafa Suleiman is famous for being one of the people right at the heart of the AI, the artificial intelligence revolution.
He is co-founder of DeepMind, which was the big revolutionary company that did some of the very early work.
He's now co-founded another major AI company, he's probably one of half a dozen or a dozen people in the world who are at the center of all these changes.
But he's also actually, I think, very interesting for listeners in other ways.
He comes from a working class background in London, Muslim background, father Syrian.
And during the course of this interview, we will of course talk about AI and some of his anxieties about the threat that AI poses.
But we will also talk about a young man who left university after a year to set up a Muslim youth helpline who worked for Ken Livingston, the mayor of London, obviously had very left wing politics in his youth and the way in which he reconciles his position now as this very, very powerful wealthy figure.
He's talking to us from Palo Alto in California with his youth growing up on the Caledonian Road.
Alistair, anything you're interested to get into on this?
Well, I've just read his book, which is coming out shortly, called The Coming Wave, Technology Power and the 21st Century's Greatest Dilemma.
And it is one of those books that sort of, I don't know, keeps you awake at night.
A lot of worrying stuff in there, and it's about AI, it's about how this thing is working or isn't working.
And it's about, as you say, the threats that it poses and what, if anything, we can do about them.
A lot of it is about containment.
He has this idea that we have to contain it.
So, without further ado, let's talk to Mustafa Suleiman.
Can you just give us an overview of what on earth is technology?
I mean, we keep talking all the time about technology in the broadest sense.
What do you mean by technology?
That is a very good question.
First of all, Rory, Alistair, great to be with you both.
Thank you for having me.
Technology, very simply, is a tool for amplification.
I mean, anything that enhances, strengthens, accelerates, augments our capabilities as humans is a technology.
The glasses that Alistair and I are wearing are technologies.
They augment our eyesight.
And the quest to invent and create new advances, which make us as a species more efficient, more productive, more capable, is as old as humanity itself.
And so that's, I think, in the broadest sense, how I think about technology, rather than, I think, people might tend to say that it's digital technologies or it's something sort of more narrow than that.

I really think of it in the broadest possible quest to invent science and turn those scientific advancements into technologies.
Okay.
So you brought up our glasses, and that made me go to one of the lines in your book that leapt out at me, which is the fact that this exists now, Chinese police officers who wear sunglasses with inbuilt facial recognition technology, and presumably, can also have technology that then can lead them to recognize you, Mustafa Salaman, from just looking at you, but then actually know all sorts of things about you.
And that then leads me to worry about the whole sort of surveillance state.
This is something we talked about with Yuvon Noah Harari last week on the podcast.
And then if I may go to another bit of your book, you say, imagine robots equipped with facial recognition, DNA sequencing, and automatic weapons.
Imagine robots will not be scampering dogs that will be the size of the bird or a bee with a small firearm or a small vial of anthrax.
This will be accessible to anyone.
This is what bad actor empowerment looks like.
That feels quite scary to me.
So should we be scared or should we be confident that we can look after ourselves as humans surrounded by this thing that you guys have root upon us?
I think if you don't start from a position of fear, you're probably not paying attention.
And part of the reason why I sort of wrote this book is that during the pandemic, I kind of had a moment to sit back and reflect and really take in what is happening on a sort of multi-decade scale.
And what became very clear to me was that we are sort of at the mercy of these unfolding exponential trajectories.
And so what that means is that as technologies get more useful, they get easier to make, they get easier to use, and they get cheaper.
And so they spread far and wide.
And that's been the engine of progress for many, many centuries.
We invent things, other people demand them, they get more efficient, they get cheaper, they spread, they empower everybody.
And so if that is going to be the story of invention for the next 40 years or 100 years, we have to ask ourselves some quite fundamental questions about what that concentration and distribution of power means for the future of the nation state, for what it means to be human, and many other really quite fundamental questions.
And that in itself can immediately start to feel quite scary.
For the benefit of our listeners, but also very much for the benefit of me, a few terms and definitions, please tell us what is artificial intelligence?
What is a large language model?
What is machine learning?
Okay, so in short, artificial intelligence is the science of teaching machines to learn.
So rather than hand crafting a set of rules, if this then that, you want to get an AI system to learn and a representation of what is useful or productive in some environment itself.
So it's really about teaching the machines to learn.

A large language model is a type of artificial intelligence.
It is a very large deep learning network which has seen or been trained on many, many hundreds of billions of words from the open internet.
And what it does is really sentence completion.
You provide an input, and then it predicts, given all of the previous words it's seen, so everything on the open internet, what is the likely next word given some sentence or given some sequence?
So it's actually a very, very simplistic, but enormous matrix multiplication.
It's doing an all to all connection between all previous words and those words that are in current context.
And that's pretty incredible.
And the large language model is just one subset of artificial intelligence.
There are many other bits of artificial intelligence, right?
There are, yeah.
Is the dominant approach that has taken over in the last three or four years or so?
So the GPT series from OpenAI, Pi from Inflection, Google's models, they're all using large language models or various different flavors of the same infrastructure.
And can you just run through, and I understand that as a sort of spoiler alert, you have some practical proposals for how to contain the technology, and you're clearly not trying to spread fear and paranoia.
But could we start at the bad end and the worst case scenario, and just lay out some of the things which could go wrong with AI in relatively simple terms?
Well, I think, you know, Alistair pointed at some of the examples in the book.
I mean, essentially, these technologies make it possible to take actions in the real world.
So the last wave of AI was about classification, right?
So the last 10 or 15 years was about deep learning systems, learning to identify objects in images, learning to transcribe audio, and increasingly learning to understand what was in sentences, in language.
And that was really about classification, ordering, structuring, creating classes.
Having now understood, you know, or at least been able to classify effectively, these models can then take those classes and generate, they can predict.
And prediction, of course, is a necessary prerequisite to taking actions.
Before you do something in the world, you imagine, you simulate, you plan, like if I'm going to go to the train station, you lay out the root in your mind, and you have to predict something about, you know, what the weather is going to be like, and, you know, is this route going to be busy, and am I going to get there on time?
That's a generation.
It's a rolling out of a sequence of activities, which allows you to be creative, because you can roll out a whole range of different sequences about how the world around you might unfold.
That's essentially what we call imagination.
So being able to roll out these sequences and then pick one path is like taking an action, right?
It's actually now beginning to do things in the real world.
And so these models enable, I think, in the limit, you know, everybody to take actions

and intervene in the world in very new and novel ways.
And I think that that's going to be the real shift compared to AI that we've had over the last decade.
I've heard you say in the past that there are a small group of technologists who believe that AI will get to a stage where it'll be indistinguishable from human self-awareness and capability, and that they'll begin to evolve faster than we can.
And if that happens, that will be a singularity that cannot be contained.
That will mark the beginning of the end of the species.
Can you expand on that?
What do you mean by a singularity which can't be contained?
And what would that feel like practically?
I mean, you're producing a beautiful philosophical analysis, but I'd love to get a sense of some practical implications in this.
Well, everything that we have produced in our world, everything that you look at in your line of sight today, is a product of intelligence.
It's a product of you looking at something, trying to understand it, and then trying to predict it well enough that you can intervene in that environment.
That is what intelligence is, and we're now trying to distill intelligence, the thing that has made us so productive as a species, into an algorithmic construct.
So if many, many people have access to that, then it will be the most productive period in the history of our species.
So when it comes to creativity and invention, we now have many millions of people who can make new things, and I think that that in itself is likely to be the most productive period in a very, very long time.
But of course, the flip side is, if you design those systems to have certain capabilities like autonomy or the ability to update their own goals or act independently of human beings, that's when they could potentially be pretty harmful or pretty dangerous.
But I get the feeling, Mustafa, that when you went away during COVID and you had a moment of reflection, you started to write this book, I get the feeling, reading the book, that you're much, much more pessimistic than you're sounding now.
Is that a misreading of the book, or is that maybe me imposing my own sense of pessimism upon what you're saying?
Because I guess I've got two problems with it.
One is I don't fully understand how it works, and that is always scary for something like me, I like to understand everything, and I don't understand when you say this thing has got the potential to essentially be more intelligent than human beings are.
And then when I read the statement, like the one that you signed with a bunch of other industry leaders mitigating the risk of extinction from AI, let's just read again, mitigating the risk of extinction from AI should be a global priority alongside other societal scale risks, such as pandemics and nuclear war.
I don't know how to read that other than with sort of pretty profound fear.
And then at the same time, you guys are there doing what you're doing, and you have people like you, people like Jeffrey Hinton starting to think, hold on a minute, guys, are we actually doing something potentially very, very bad here?

So I think the surprising thing has been that what we have thought of as unique and very special and impossible to replicate skills and capabilities that we have as humans, our intelligence has actually been relatively easy to replicate, right?
That's been the surprising thing that's happened over the last decade.
When we sort of set about this quest, when we found a deep mind in 2010, we had no idea how hard or how long it would take to make progress in some of these areas.
And slowly, but surely, we've been knocking down some of the big milestones.
Well, it's not been that slow, has it?
It's probably not been that slow.
It's true. It's been a decade.
It feels slow to me.
You know, we now have near perfect transcription.
We now have near perfect image understanding.
Increasingly, we have image generation, which is really very good.
We have text generation, which is very good.
The next big milestones over the next decade will be planning and imagination.
So, you know, it's true.
It's been easier than we previously thought.
But if I were to go back to the bird or the bee with the small firearm or the small vial of anthrax, would I be able to have access to that kind of thing quite easily within a few years, unless there is what you call containment?
Correct. So, I wouldn't say within a few years, but certainly within a few decades.
So, the trajectory, the historical evidence is very clear.
Where something is useful, it gets cheaper and it proliferates.
So far in the history of our species, everything, without exception, that is valuable and useful spreads far and wide.
So, that's the default that we have to contend with.
And that is extremely true in the context of software, because software isn't constrained by the traditional frictions of the physical world.
Software can be easily moved around on the web and so on.
It's an idea, right?
So, it spreads even faster than a physical invention like a motor car.
So, I think what I'm sort of trying to establish in the book is that if everything defaults to getting cheaper and easier and spreading far and wide, then we have to contend with power, the ability to take actions and do things in the world, being very, very widely available to many, many people.
To be cheesy, a couple of examples that I guess trouble people.
One of them is the next US presidential election and a real fear that AI can generate deep fakes or intelligently manipulate voters in such a way that it could affect the outcome of the election.
Or another fear might be that in a few years' time, some version of, I don't know, chat GBT-4 would allow me to start doing stuff that at the moment could only really be done by somebody with a doctorate in biochemistry.
And I could start making some pretty dangerous substances.

These kinds of things seem to me to, well, initially, actually,
I remember talking to you about this in Japan.
And in Japan, we were on the table and I tried to say, is this like nuclear weapons?
And the optimists, there are a lot of optimists in this world said,
oh, no, no, no, no, it's much less dangerous than that.
You're making it sound much too scary by saying it's like nuclear weapons.
But what I remember you saying is that actually in some ways,
it has features which are more dangerous than nuclear weapons.
They're easier to proliferate these technologies than nuclear weapons.
Can you expand on that a bit?
Yeah, I mean, the cost of production of nuclear weapons
is one of the things that has halted its proliferation.
So we can clearly see where nuclear weapons facilities are in the world.
They're large, they're physical, they involve vast numbers of people,
they involve scooping up chunks of uranium-235, which is in itself very rare.
And that friction, the capital costs, as well as the kind of physical constraints,
make it sort of very different to the way that software evolves.
I mean, take an image generation AI today,
that's been trained on many billions of images that are available on the open web.
The model itself is now about two gigabytes, so it can sit on a thumb drive.
And in some sense, what's happened is it's compressed all the knowledge and insight
from all the different images that it has seen, if you like,
onto this tiny little transferable representation,
which can be shared and copied and moved around, et cetera.
And now it can be used to generate pretty good, not perfect, but pretty decent images.
If that trajectory continues, then I think we can all imagine what the world would be like
if we had perfect photorealistic audio, video, language generation that was entirely transferable
and indistinguishable from human-generated content.
In fact, it would likely be in the limit over a 10 to 20 year period, much, much better.
So your point about the election is a very good one.
I've said publicly many times that we should immediately call for a moratorium, a ban,
on the use of AI-generated content for any kind of electioneering.
And certainly, in the case of AI chatbots, those topics of conversation should be off limits,
because they're going to be very persuasive.
And we also can't fully verify that they're going to be accurate and reliable.
So people should not rely on them to do that.
And the manufacturers of them shouldn't put them to those purposes.
But if you think that one of the contenders in the next presidential election is almost
certainly going to be Donald Trump, it is impossible to imagine that Donald Trump
is not going to use something that might help him become president.
And I wanted to ask you about politics and this, because briefly it's true,
but you worked in politics and that you were an adviser to Ken Living, somebody who's a mayor of London. You were one of the AI people that had a meeting with President Biden not long ago to raise some of these issues with him.

First thing I want to ask you is, how is politics dealing with this in your view?
What are the tensions between the fact that your world operates at such pace
and the political world operates pretty slowly?
And I guess I want to ask very directly whether you felt Joe Biden got this.
And if so, great.
If not, do you know any of the current crop of world leaders who do get this?
I mean, so just going back to the Ken thing.
So I did briefly work for Ken when I was 21 as a policy researcher, not an adviser.
It was a very interesting experience.
It was only sort of nine months or so.
But I certainly learned enough to realize that I probably needed to get into technology
rather than politics.
So that was important.
Yeah, I mean, look, this is partly why I'm writing the book.
I mean, we need our institutions, those that defend and represent the public interest,
to evolve as quickly as the technology is evolving.
And everyone would agree that that's clearly not happening.
Of course, it's true that Trump will most likely, and many other politicians,
will want to use every possible advantage to win and get ahead.
And that's, again, the story of proliferation.
We have these huge geopolitical incentives, whether it's China or the UK or the US,
everybody wants to get that advantage.
That's what technology is.
It provides an edge.
It's an advantage.
It's a motivation to take risks and push the boundaries and advance your agenda, whatever that is.
So I think the White House administration in general, this one is on the case.
I mean, they have been very bold with their export controls, whatever you think of them.
That's a very significant intervention, holding China back from the cutting edge
of AI chips, which is a significant restriction.
They've been very proactive with the voluntary commitments.
So the top seven companies building the largest AI models in the world,
one of which is mine, Inflection AI.
And we all signed up to these voluntary commitments to expose our work to independent
audit to share best practices, et cetera, et cetera.
So I think that things are heading in the right direction,
and there's cause to be optimistic about that.
Likewise, I think that the EU AI Act is for sure heading in the right direction.
This threshold of evaluating risk of an application
rather than trying to define the capabilities makes a lot of sense.
Tell us a little bit about that, Mustafa.
Expand on what the EU is trying to do.
So to be fair to the EU, they have been ahead of the game for many years now.
They've been drafting this EU Artificial Intelligence Act for over three years,

maybe four years at this point.
And their approach has been to say, we will require developers of AI
to proactively assess the potential risks of their AI application.
And if there is a significant risk, if it exceeds some threshold,
then they have to disclose the data that that AI system has been trained on.
They have to disclose how the model operates, why is it making a recommendation?
Why is it not making a recommendation?
And there's a sort of reporting requirement
that basically imposes a kind of a precautionary constraint on activity.
But Mustafa, when I was talking to some of your colleagues about this
in the States over the last few weeks, many of them are very angry with the EU.
And some of them are saying, well, that's the end of the EU.
They're never going to make any technological advancements.
They're ridiculous.
They're Luddite.
They're in the Stone Age.
We're just going to go and do our business somewhere else.
And presumably that is the problem here.
But unless the entire world legislates, but like climate change, isn't it?
You just need a few bad actors prepared to host people doing crazy stuff
and the whole thing collapses.
Well, that's the story of technology, right?
The challenges that we're all collectively in this horrible tragedy of the Commons race.
But that raises the question of, is there going to be a race to the bottom?
Or are some people going to hold the line and defend their values?
I personally believe very much in the EU AI Act approach,
I think a precautionary principle is healthy and correct.
And I don't think the cost to innovation or anything like as big as what some people have speculated.
And I think that you might have been talking to the classics of the Silicon Valley types.
And I'm not one of them.
I guess the reason you're maybe not one of them is your background is quite interesting for,
I'm guessing you're meeting a lot of very smart,
very probably quite privileged and entitled people.
And yet your dad was, you're born in London.
Your dad was Syrian, a taxi driver, your mom, a nurse.
She was English here, British, yeah?
Yeah.
Just give us a little bit of how you got from that kind of upbringing,
pretty working class background in London,
to now being pretty successful, vastly wealthy by comparison with most of the kids you grew up with.
And as Roy said in the introduction, the cutting edge of one of the biggest issues
of change of our time.

Well, I mean, I grew up on Caledonian Road in London,
on Council State.
And then when I finished primary school, I was pure chance.
I was very lucky.
Someone at school said, oh, you should check out this really good grammar school
up in Barnett.
You have to sit an entrance exam.
And I was like, okay, cool, I'll try that.
And it was a nonverbal reasoning exam, two exams actually,
English and Maths for nonverbal reasoning.
And I did really well.
It was Queen Elizabeth Boys in Barnett.
It turned out to be the best state school in the country in terms of GCSEs and A-levels.
And it changed my life because I was surrounded by people who also kind of
wanted to work hard and enjoyed learning and they just pushed us really, really hard.
And it definitely changed my life going to that school.
It gave me a big opportunity.
What was in your mind that led you, via Ken Livingstone and other things,
but led you to what you do now?
What sort of mind do you need to be in the AI world?
Is it about words?
Is it about maths?
Is it about science?
What are the strengths that you need?
That's a great question.
I think it's about systems thinking.
So I did philosophy at Oxford.
And I think that what that training gave me was it honed my instinct for thinking
at varying levels of abstraction.
I can move from micro to macro and I think in terms of systems and causality.
And I'm also comfortable not knowing.
Like you sort of have to sit with a certain amount of ambiguity.
And I'm a translator and an interlocutor.
I get different groups of technical people to communicate well with one another,
to talk to product people, to talk to operations people.
And so I'm really playing this sort of constant translation role
in order to steward the big picture vision.
And I think that's quite an important thing to keep in mind
because much as people are scared about the potential outcomes,
they're also scared about the complexity of these ideas.
Everyone's like, oh, it's complicated AI.
I could never do it.
It's just nonsense.
It's like totally accessible.

There are scores of videos online.
You could watch two or three hours of videos.
And if you're patient enough,
you can get through roughly the big picture of like,
what is this system trying to do?
And I'm very lucky to have been doing that for 15 years now.
And I have a pretty good technical understanding of everything that's going on in the field.
And I think that's available to a lot of other people.
And we should be pursuing it because that's core to containment.
That you have to understand something in order to be able to steward it.
One of the interesting things,
and many interesting things in your life,
is you left Oxford after only a year or so
to set up something called the Muslim Youth Helpline
and worked for that, I guess, for three years.
And would love to know a little bit more about that.
So it's a fascinating thing to do.
What was it doing?
Why Muslim youth?
Why was it such an attractive thing
that you were prepared to throw away your undergraduate degree?
What does your parents think about this whole thing?
I mean, tell us a bit about that stage of your life.
Yeah, so I was raised as a very strict Muslim.
My mom was a convert well before she met my dad.
So, you know, the strongest type.
And yeah, when I got to Oxford,
I had managed to leave the religion.
And, you know, in my philosophy,
discovered human rights principles
and just found it tremendously inspiring.
I was like, here is a language and theory of justice
that applies to everybody, not just Muslims.
So we are not the chosen race.
And that always made me kind of uncomfortable and confused.
But this gave me an intellectual framework
for moving away from religion.
I became a very staunch atheist.
But during my kind of philosophy trading,
I also got quite frustrated
at how theoretical and abstract the work was.
And I wanted to kind of do good now.
I was truly like eager and motivated to have a positive impact.
And there were a couple of other people I met at Oxford

who were already working on a small helpline.
And I decided to quit, move to London and help them full time.
Was it a difficult time in your life
where people thought this is crazy?
He's left university.
Nothing's going to work out for him.
He's floating around.
Was that a difficult period?
And also related to that, Mustafa,
what are your mother in particular
think of you suddenly unfinding God, as we might say?
That was a very challenging one.
But, you know, I think going through sort of secondary school,
it was quite difficult for my parents to understand
what they had produced.
They were just like, why don't you leave?
My mom was like, why don't you leave school at 16,
get a trade?
Why are you messing around with all these ideas?
She was really eager for me to get a good trade.
And that's how you'll be safe, be a plumber.
And so I think they sort of didn't quite grasp
why I would want to go to Oxford.
Why was I doing philosophy and theology?
So by the time the Muslim Youth Helpline hit,
I mean, they just were, that was like, what are you doing?
So they were definitely confused by that.
But the Muslim Youth Helpline to me was,
it was a secular, non-directional, non-religious effort.
It's a very simple practical tool.
It took traditional counseling and peer-to-peer support
practices and just skinned it with a little bit of
faith and culturally sensitive language.
So it was sort of Samaritans for Muslims?
Yeah, exactly, kind of like that.
So the key thing was it was supposed to be non-directional,
which it was.
Everyone understood that this was about Sunnis and Shias
and people from Bangladesh and people from Sudan and Arabs
and Malaysians working together in the common interest
to provide support.
And it was non-religious.
That was really the main sort of goal here,
is to move beyond sort of sectarianism.

Mustafa, thank you very much.
Can we take a quick break?
Can I just, I mean, I'd like to bring in Alastair on this,
because Alastair's obviously got a very deep interest
in mental health.
And I'd be interested to hear the two of you just chat
for a second about that, because why was it particularly
mental health that you were drawn to?
And the helplines are very particular things and a very
particular way of doing good in the world.
Yeah, well, it was just after 9-11, so it was sort of 2002.
And all the talk was of where the next British-Muslim
threat might come from.
There was a lot of Islamophobia, although we didn't have
that word particularly for Islamophobia.
If you were British and Muslim, you were four times more
likely to be in prison.
There was this constant sort of fear and anxiety
that there was a threat from within.
And that kind of motivated me.
There were these basic issues that were getting
dressed up as mass political issues.
People were just fundamentally suffering, bullying.
They had difficult lives at home.
They were struggling with sexuality.
And on top of that came this huge sort of political narrative.
And I think it was causing a lot of confusion.
And that was really what I wanted to tackle.
Would you say you've got good mental health?
I would say that I have learned to manage my mental health.
I've learned coping mechanisms that help keep me
in moderate equilibrium.
But it's certainly something I've always wrestled with, for sure.
And wrestled upwards, moved upwards or moved downward or both?
You know, I think both.
I'm very driven and have a lot of energy.
But also, I'm constantly wrestling with demons and darkness
and difficult childhood and difficult memories.
And as I'm sure many people are.
And what was it about Islam that attracted
your mother?
And what was it about Islam that you decided it was not for you?
I mean, it was very clear to me that the kind of like righteousness
and the elite believers, I can remember the most basic question

asking my dad when I was really young was like,
well, what about the good people who aren't Muslim
but are going to go to hell?
That doesn't seem right.
They're not that bad.
I was thinking about my friends at school, of course.
And there's just no good answer even to a child
for any of these like really fundamental questions.
And so I was always left like frustrated and it felt incomplete.
You know, for my mom, she was friends with Kat Stevens.
And when he converted to Islam, I think in like 1980 or something,
she sort of followed that path and took his name and so on.
So, you know, that was her sort of journey.
This is not in the book.
This is not in the book.
But this is a freebie for you since you asked.
But it fascinates me because I think that this whole thing
about what we believe and how we believe it,
how deeply we believe it,
but I think it's one thing to be raised in a faith
but it's quite another one to choose a faith you're not raised in.
It's definitely very different.
Yeah.
I mean, and many people choose, you know, sort of savior
and choose the kind of support network.
I mean, there is something very attractive about the idea
that you can begin again and that you're going to be, you know,
sort of supported and embraced by a group of people who are moving
beyond race.
I think it's very important in a lot of the Islamic narrative
is like it's post race, you know, provided you believe this,
we can all be equal other than those, of course, who don't believe it.
So, you know, I think that's very attractive to some people.
But do they believe you're going to hell?
Because you don't believe now?
Well, I think over time, my mum has sort of left the faith as well
when I was a bit older in my 20s.
So probably not.
But I think my dad probably does, which is a funny thought.
And how do they feel about your kind of mega success
and the sort of millions in the bank and all that stuff?
You know, I'm fairly nonplussed.
I mean, I think they're busy doing, you know,
you know, how is this sort of doing their day to day life,

just like any other, you know, older parents these days,
busy with the small things like fixing the washing machine and so on.
You wouldn't just say, listen, let me, if he's broke, I'll buy you a new one.
Oh, of course.
Yeah, no, no, I've done all of that.
Yeah, yeah, of course.
No, I bought them houses and all the rest of it.
Yeah, definitely.
But there's clearly, oh, not clearly,
but there seems to be a through line from the Muslim youth helpline
to the latest product which you've developed with AI, which is Pi.
And Pi, for people who don't use it, is a large language model.
But unlike chat GBT, it's less about spewing information,
and it's more about the AI developing an empathetic human relationship.
In fact, it feels a little bit like a helpline.
I mean, is there a sort of connection between that very first thing that you did at 20
and what you're doing now that you're applying a new technology
to a similar issue that you care deeply about?
Yes, in a way there is.
I mean, I think that in the future,
everybody is going to have a personal intelligence,
an AI that essentially functions as a chief of staff.
It will help to schedule and coordinate and plan and book and buy.
It will teach you.
It will provide you with advice, but also support and be a companion.
And so we've started with an empathetic and kind and supportive AI.
You know, certainly that's inspired by some of the stuff I did earlier in my career.
But really what I wanted to do was demonstrate that AI can actually be good.
I mean, it really can be controlled.
The behaviors can be engineered with very nuanced and precise tones.
So if you use Pi, you can really see that it's quite boundary.
It's super respectful.
It's very empathetic.
It's very hard to push it to be offensive or abusive in any way.
It's virtually impossible.
I mean, we don't actually suffer any of the jailbreak,
the red teaming attempts that many of the other LLM large language models face.
And so that was really my quest is to sort of demonstrate
that we can actually reproduce a very controlled set of behaviors.
And so I do worry sometimes that maybe like many of us,
you are very, very powerful at seeing the challenges and the problems.
And then you feel, because we live in an optimistic age and publishers and others
expect it, that you have to produce a very clear 10-point conclusion on how to sort it out.
But sometimes it feels to me a little bit like, you know, what I remember in Afghanistan,

which is people would provide an incredible analysis of everything that was going wrong
in Afghanistan and then have a 10-point thing at the end saying, not astounding the fact that,
you know, Pakistan's destabilizing intelligence or my own life, right?
I write a whole book about politics, about how miserable it is and how grim it is.
And then I feel forced to produce a sort of concluding thing with a positive story
about what we can do to sort it out.
Do you ever fear that you've sort of taught yourself into optimistic solutions,
but that actually if someone were to read it in 100 years time,
the thing that would be truest and most powerful,
are your analysis the problems rather than your solutions?
I think that's a fair point.
I think if you read the book, I don't think you can accuse me of being an optimist
or a pessimist, depending on which page you read.
And I think that that's an honest look at both sides of the coin, right?
This is the reality.
And I open the book with this idea of pessimism aversion.
We have to stop this false narrativizing, good or bad,
this catastrophizing or this techno, you know, optimistic utopia,
and it's all going to be wonderful.
And we have to actually engage with the very nitty-gritty practical reality
of how these huge incentives produce technologies which default to proliferation
and we have to look at those consequences.
And then we have to decide collectively what that should mean
for how we manage governments and the future of the nation-state.
So it's too simplistic a frame to say,
are we optimistic or pessimistic?
It's much more important to rip apart the detail of my proposals around how,
you know, I'm actually suggesting we do containment, right?
Yeah, well, just on that then, I'll hand back to Alistair.
I mean, be lovely to hear you talk a little bit more about the 10
practical policy suggestions you have about how to contain AI.
But I think my fundamental problem with most of them
is that you're assuming goodwill and cooperation,
and you're assuming that it doesn't proliferate pretty quickly to some bad actors
who don't give a monkey's and are going to push ahead for advantage.
And a page after page, when I find myself noting,
I just think, whoa, that's pretty hopeful there.
That sentence to saying, you know, if we can do this, if we can do that,
I'm thinking all the time, wait a second,
am I sure that North Korea or China or Dubai is going to play ball with this idea?
Well, I mean, I'm not sure that we're sure of anything in this world, right?
So I think that one has to make proposals which, you know, are both practical,
but they clearly lean towards a hopeful outcome, right?
I mean, we're trying very hard to lay out a roadmap for what containment looks like.

And that is going to require cooperation because if your assessment is correct,
which is that nation-states are fundamentally zero-sum,
will always be competitive and will always be opposed to one another,
then it really is all better off because we're about to create extremely powerful tools
that are inevitable over the next century.
I mean, it might not happen this decade or two decades,
but over, you know, a century scale, it really does turbocharge many, many, many actors,
big and small, and we have to contend with that.
I mean, I think I've read the book in much the same way I think as Roy did,
feeling very, very alarmed and pessimistic as I sort of traveled with you through a lot of the kind
of both the history, but more so what's happening now.
And then along comes this 10-point plan, which I did, I agree with Roy, I think is so reliant
on the people that you would need to implement such a plan in every country in the world,
because they're going to be worried about this threat to the nation-state.
I think that's something that a lot of people are thinking about.
And then you have this lovely line, which I, you know, faced with the abyss,
geopolitics can change fast.
In the teeth of World War II, peace must have felt like a dream.
Now, that is true. It must have felt like that.
But at the same time, there was back then the sense of a global structure which emerged from it.
And there was time. What I feel with what you're describing the first three quarters of the book
is that we're slightly running out of time.
Now, am I just being too pessimistic here?
Am I not falling enough into the pessimism aversion trap?
Look, I think you're right that time is against us.
And I think that the pace of change is phenomenal.
And everybody can now see that, particularly in the last 18 months.
But, you know, the same trajectory is also true on synthetic biology.
And so, which I think is a decade or so behind of where large language models are in terms of being
able to generate arbitrary compounds, which are of enormous value just in natural language
instruction.
So, yes, in that sense, it is a warning and it is a provocation.
The goal is to basically say, if I'm wrong, then propose an alternative path.
What does the non-competitive or let's say the cooperative path to resolution look like?
And I think that the comparison to, you know, the Second World War is appropriate because
it was, you know, sort of inconceivable that there would be peace.
And yet peace was found and out of that over a decade during the 50s,
many of the world's most significant institutions were born, which have driven
untold prosperity and progress.
The problem is we're mere mortals, right?
So, if we resign ourselves to inaction and fear and heartbreak,
then we won't get anything done.
So, you have to be hopeful and positive.
Yeah, but you also have to be realistic about the world in which we're operating,

particularly at the political level.
And, you know, you make the point that for something like climate change,
we kind of have the metrics.
We know what bad looks like and we know then we can work out the things to avoid bad.
With this, because it has happened so quickly, we don't really have the same metrics.
I don't know what a bad AI outcome or a good AI outcome looks like.
And I don't know who's meant to be explaining that to me.
Yeah, I think that the world is catching up rapidly.
I mean, people have gone from basic...
The average person has gone from basically zero understanding
to increasingly getting a grasp of what these things can do.
So, putting them out there in the world is clearly a useful first step
because people get the measure of them.
And I think in two or three years, we'll start to see the kind of social consequences
and what they actually feel and what they're like for our politics and culture.
And I think that will give people the resources and the kind of energy
to start thinking about other strategies for containment.
How do I make sure that an AI system is always accountable to me,
is on my side, is, you know, working in the public interest?
Those are the questions that people will start to wrestle with in the next three or four years.
Who are your sort of peers and colleagues at the sort of level at which you're operating?
I mean, we get very little glimpse of them and you're interacting with them daily.
You know all these famous people, Bill Gates, Mark Zuckerberg.
I mean, it's extraordinary, for example, the endorsements in your book.
I mean, you're being praised by Al Gore and you've all known Harari and all these people.
So, what is this strange culture that's emerged?
You have a group of people who are unbelievably wealthy
and who are also in a way that probably wasn't true of the kind of robber barons
the late 19th century. They're also quite intellectual.
I mean, I guess, you know, Henry Ford and Rockefeller were clever people,
but they were quite practical clever people.
This is now a group of sort of billionaire nerds.
They're people who, you know, incredibly good at maths or did unbelievably well in exams.
I mean, if Alistair and I were to sort of hang out with them,
you're talking to us from Palo Alto where presumably you're working 14 hours a day,
but when you're not working 14 hours a day, you're going meeting these people at conferences.
What is this culture? What are they like and what does that mean for the development of
civilization to have a civilization run by billionaire nerds?
I'm not sure that the civilization is run by billionaire nerds, but I think that it is a
culture of curiosity and a willingness to speculate on the far out future.
That's what politics is lacking today. No one in your world is proposing a plausible path
to a positive vision over the next 10 to 20 years. You know, politics feels broken to people.
It feels broken to the people who are in it. And where are the great visions of the future?
You know, what are the stories that we're telling ourselves about how we're going to create peace

and prosperity and more wealth? And I think the challenge is that we technologists are telling ourselves the story that we have the chance to produce, you know, radical abundance.
And I truly do believe that that's my primary motivation for working on building intelligence. And I think we have every chance of doing it. And like I said, I really do think this is going to be the most productive few decades in history. So I think that's the challenge for conventional politics is to try to respond in a way that like basically takes control and takes all of the this kind of strengths and benefits of the technological ecosystem to try and harness them in the public interest. I mean, that's really the goal of my book. It's a it's a love letter to the nation state. It's a warning. It's saying now is the time to act. Now is the time to refresh and regenerate our politics to make sure that it remains in control and relevant and capable of intervention, given what's coming technically. Your book is obviously written for more than a British audience. But can I imagine that it might have been your homeland that you had in mind when
you said, when a government has devolved to the point of simply lurching from crisis to crisis, it has little breathing room for tackling tectonic forces requiring deep domain expertise and careful judgment. I mean, it's true, right? I mean, if if if that's an example of an area where I'm being optimistic, because I'm saying that we need cooperation, then I'm sorry, what we need is nuance and patience and respect and forgiveness for one another and, you know, enough of the kind of bashing of each other and so on. I mean, without that, we're not going to move forward. And yet, obviously, particularly, I think in the UK, we're mired in this pretty awful depressing situation where where I think politicians have lost confidence in themselves and the political process has been so damaged by the last decade that people don't have a clear sense that this is the sort of infrastructure that can help navigate the next 20 years.
And paradoxically, though, technology has been one of the drivers of polarization populism post truth, whether through social media or through the massive changes in our economic structure. And so it's a very, very odd world in which to some extent, some of your billionaire nerd friends have contributed to the forces that erect our politics, as well as offering the great future visions for how to fix it. Yeah, this is true. I mean, this this this is sort of what happens when we when we sort of unleash technology in a in a sort of unbridled way. I mean, you know,
I think one of the big failures of the story of the internet over the last 20 years is that technology companies sort of claimed that platforms were really just neutral reproducers of information that they that they they sort of hid the fact that re ranking and ordering was in fact moderation, right? And if the signal for driving that re ranking was in fact singularly engagement, then things that are, you know, less true or maybe completely untrue or things that are outrage inducing would clearly, you know, find their way to the top of the the ranking quicker than others. And that in itself would trigger our dopamine releases. And I think everybody is sort of more familiar with that story now. And that that's the sort of unintended impact of technology platforms that that go out there without thought. And I think that's one of the things that is slightly changing with this new wave of AI. I do think that we're learning those lessons and more proactive and more conscious and more deliberate about the way that we design these systems. I think you can see that in Pi. But I think you can also see it with the responsible AI principles of the other AI companies starting to think proactively and adopting the precautionary principle.
Now, you've talked about EU Act and other interventions by some of our political leaders that

you clearly take seriously and have some respect for. But is there a sense that a lot of the people in your world really do feel now that they're kind of masters of the universe that they can operate without worrying too much about the political process? I think what a lot of people forget is that the nation state is still the backstop of power in our world, right? Managing the military, managing taxation, and essentially operating the judiciary is always going to be a thousand times more powerful than corporations. So whilst government institutions move slower, they do fundamentally wield way, way, way more power, right? And I think that's a good thing. And so for those who are sort of techno boosters and maybe do believe that they're kings of the universe, and they are kings rather than queens at this point, sadly, I think that they're going to get a tough shock when political systems kick into action. Because if the trajectories continue as they are over five to 10 years time, then people aren't going to sort of accept these kinds of extreme concentrations of power or the risks that might unfold from this kind of technology.
We're coming towards the end because you've been very patient, but let me hit you then sort of really my final area, which is what is the consequence of AI on employment when we're really thinking about the politics the next five to 10 years? What's AI going to mean for jobs? And what's that going to mean for our societies? I think that the way to understand this is that for a period of time, it is going to make everybody much more efficient and more productive. And the question is how long that period of time is because everybody who wants to invent or create or learn is now going to have a personal AI in their pocket, which is going to provide them with the most patient, the most personalized and the most kind of useful knowledge synthesized, perfectly given their context to that question in their style. And that's going to make us all way, way smarter and way more productive. I think the challenge is going to be how quickly those sorts of tools end up getting integrated into the workplace and ultimately replace a lot of what I would call intellectual cognitive manual labor or cognitive manual labor. So sending simplistic emails, processing orders, managing call center inquiries, the kind of basic back office administration
that powers most large bureaucracies, you know, I think over the next five years, maybe seven years is likely to be automated. And so there's going to be a very disruptive transitionary period where we need people to retrain and reskill, and we'll have to fund that and support that. And that basically means taxation.
And we could be looking at 20% unemployment over the next five to seven years?
Not five to seven years. No, I don't think that's, I don't think that's possible.
I think that over a 20 year period, you could see double digit structural disemployment. So, you know, people who want to get a job and they can't compete in the labor market, maybe I think that that that is a serious scenario whose probability has increased in a non trivial way. And that's why I say taxation and redistribution is going to become even more important. I mean, at the moment, you know, in the US, labor is taxed on average 25%, whereas software and equipment is only taxed at around 5%. And I think we're going to have to make
this shift towards taxing capital significantly. There are large stocks of wealth in land and in property and in, you know, stocks and shares, which, you know, we're going to have to call on to fund this transitionary period over the next 20 years.
To finish for me, last question. What do you think irritates you about these sort of interviews? What do you think people like Alistair and I often miss? If you had a sort of final

couple of minutes just to reflect on the end of the interview, think,
I wish we'd got onto this, or I'm so frustrated that people ask these kind of questions and
don't ask those, what would you say in the last couple of minutes?
I would say I'm the expert in technology and in my area AI. You guys are the 30 year veterans
and experts of politics. I think it's on you to help answer the question that you put to me
about how we evolve politics and governance so that it's fit for the modern technological age.
So that's a kind of the discussion that I think it would be good to have because I'd be looking
to both of you for, you know, answers to the kinds of 10 point plans that I laid out and those kinds
of questions. And I think that's the goal of my book is the provocation has been laid out.
I've done my best to lay out, you know, the ways in which I think containment might be possible,
the practical things we can do. If they seem insufficient, then please take up the mantle
and help us all to figure out how we evolve governance and politics for the 21st century.
You know, the problem is with stuff. Rory and I basically share your analysis
of the state of our politics at the moment and don't quite feel we're empowered or have the
strength and energy, just the two of us to kind of, you know, get it back in the right place.
But I thoroughly enjoyed the book. I mean, it enjoys the wrong word. Actually,
it did alarm me woke me up. I think Rory and I have talked a lot about artificial intelligence,
but I think what your book did was was really kind of bring it up close to me
how this doesn't need far greater contribution from the political space. Because without it,
the same thing is going to happen is happening with social media is that by the time
all the implications are fully understood, the politicians are already too late.
And I worry that's what's going to happen with this as well. So you're absolutely right that
the politicians need to wake up and we can definitely be part of that dialogue.
Glad to hear it. Thank you.
Thank you. Thank you so much.
Thanks for your time.
Thanks, guys.
So, what do you make of that?
I sense the tension there between his book and himself. Or maybe I just because I don't
really get this world is not my world at all. By reading the book, I thought,
God, here's a guy who's coming along and he's really, really trying to warn us that this thing is
bad. But then he has, as you say, his 10 points at the end says he's going to be fine provided
we do this. But I don't know. I don't feel that assured.
I think it's very tough. I mean, I think he's, I mean, as you would have picked up,
unbelievably bright. And I think actually in many ways, and we didn't really get on to his
politics. But I think if we had, you would have found a lot in common. I mean, he's somebody
who very much came from the left, probably a bit to the left of you, I guess, when he's working
for Ken Livingston. And he's quite unusual in the Silicon Valley setup in really emphasizing
social justice and encouraging people to push up taxes. And he thinks that the way to deal with the
challenge from mass unemployment that might come from AI is not just income taxes, but wealth
taxes. And I think he's happy to put his money where his mouth is. I think he does believe that
there is a huge obligation for the very wealthy to make sure that this technology isn't having an
impact on the poor. Yeah. And the other thing he does, which came through both in the interview
and in the book, I find a lot of these kind of tech guys are pretty anarcho capitalist and they

have a pretty rich contempt for politics and politicians. Whereas I think part of his thing is that politics has to be part of the solution. It can't be done just by the tech people. Which I think is right, but also I think is good coming from him.
But he's also, I mean, one of the things that I think you said to me just immediately afterwards when we were chatting is that you thought that he was a bit sort of defensive. And I think that is partly because he's so central to this whole thing, to this company, to the AI movement, that you sense that he has to be quite careful what he says. He can't quite kick back and relax in the way that some of our guests can because every word that he says about the threats AI is affecting international legal policy, share prices, companies. I mean, in a way, it's sort of tribute to how much he's in that game that he has to be quite careful measured in the way that he deals with this stuff. It's very, very sensitive. And we're right at the cusp of this.
Yeah. One of the other points in the book is that of all the themes and trends and the technology and the change that he's writing about in the book, which is vast, he says in terms of the jobs, the people working in all of that, it's about 150,000 people around the world. It's tiny.
And yet, as you say, these people have real impact upon stuff that we do understand and stuff that we don't understand.
And they're quite remarkable people, too. I mean, whatever one thinks of them, and I like Mustafa, I think he's an unusually intelligent, thoughtful person, but they are all in their own ways, very remarkable. I mean, I obviously am pretty horrified by what I see at a distance of some of them, but they are all incredibly bright. They all read a lot. They think a lot. They can be very narrow. I mean, if you think about, you know, Marshal Mark Zuckerberg, there's a sense that they're very, very much engineers with a strong kind of technical bent. But others of this group, which includes Bill Gates, are making real differences in the world. And Bill Gates has driven huge changes on malaria, on all forms of vaccination and global health. It's a very interesting period that the richest people in the world, so much more interested in kind of ideas and thinking than maybe billionaires were 40 years ago, 50 years ago.
Or they just wanted to make money.
Well, I guess it's that to make it to the top in these tech companies, people need to be very, very adept at maths, computing analysis, philosophy, as well as working very hard. I mean, that's the other thing about them. I mean, Mustafa, we had an interview with him there for an hour, but you know, what I know of him is that he will have then gone on to work a 14 hour day. So he's part time here. Exactly. He's a bit like you, Alastair, is what I'm trying to say. I think in fact, actually, strangely, he is a bit like you.
Talking of the other guys, apparently the Musk Zuckerberg fight is off.
Oh, no. No, no, no.
You were so excited about that, weren't you?
I was very, very excited about that.
Yeah, no problem.
I think strangely, you'd find them, you'd find more incomparable than you'd think. They're very, very driven. Many of them have had challenges around mental health, but are therefore unbelievably productive when they're up.
Yeah, he sort of touched on that, didn't he, for himself, but I could sense he didn't really want to go there. No, he hasn't. I mean, I think that was probably the most open he's been about

his personal life that I've ever heard. I mean, he was much more open about his parents and his father and his faith and his mental health than I've heard before.
And I think, is that the first mention of Kat Stevens on our podcast? I think it might be.
Very good. Well, I think as we come to, I mean, I think he is a very, very remarkable British person. I think it's an extraordinary story. Working class father, Syrian immigrant taxi driver who's become one of the most powerful central voices in this revolution that's changed the world. So thank you for doing it with me. No, I enjoyed it, Roy. Thank you.
Thank you very much. All right, bye-bye.
Hello, Restors Politics listeners. It's Anita Arnand from the Empire podcast, which I host with... me, William Dalrymple. And we are here to tell you about our brand new series of the Russian Empire and the Great Game. With Russia dominating the news at the moment, we wanted to look into
its history and see if there are any answers as to why Putin is doing the things he's doing, thinking the way he's thinking. Yes, after the Russian invasion of Ukraine, Sergei Lavrov, the foreign minister, joked that Putin had only three advisors that he listened to. They were Ivan the Terrible, Peter the Great and Catherine the Great. And we've got episodes on all three.
That's right. And this week, we're telling the story of Putin's last advisor, Catherine the Great, the most powerful woman in history. It was under her reign that Russia took control of huge areas of modern Ukraine, annex Crimea and built the frontline towns like Kherson and Sevastopol that are all too familiar to us from news bulletins at the moment. So if you want to see how Russia became the world power it is today and look at how Putin is influenced by said Russian Empire, you could do worse than listen to Empire wherever you get your podcasts.