Transcript

how would you describe this current moment in AI machine learning
whatever we want to call it
I think it's a pivotal moment chat GPT has shown that these big
language models
can do amazing things and the general public has suddenly caught on
yeah we
have because Microsoft released something and they're suddenly aware
of stuff that
people of the big companies have been aware of for the last five
years yeah what did you think the first time you
used chat GPT um it's well I've used lots of things that
came before chat gpg that were quite similar so chat GPD itself
didn't amaze me much gpt2 which was one of the earlier language
Brands amazed me
and a model at Google amazed me that could actually explain why a
joke was funny oh really yeah in just natural
language it'll tell you yeah you tell it a joke not for all jokes
but for quite a few of them it can tell you why it's
funny okay and that you it seems very hard to say it doesn't
understand when it can tell you why a joke's funny so if
chat GPT wasn't all that surprising or impressive were you surprised
by The public's reaction
to it because the reaction was big yes I think everybody was a bit
surprised by how big the reaction was
that it was the sort of fastest growing up ever yeah um
maybe we shouldn't have been surprised but people the researchers
had kind of got used to the fact that these things
actually worked yeah you were famously like half a century ahead of
the curve on
this AI stuff go ahead correct me go ahead not really not really
because there were two schools of thought in AI
um there was mainstream Ai and then there was years but it was all
about friends yeah that thought it was all
about reasoning and logic and then there was neural Nets which won't
call AI then
um which thought that you better study biology because those were
the only things that really worked
and so mainstream AI based its theories on reasoning and logic and
we based our
theories on the idea that connections between neurons change and
that's how you learn and it turned out in the long run
um we came up trumps um but in the short term it looked kind
of hopeless well looking back knowing what you know now do you think
there's anything you could have said then that
would have convinced people I could have said it then but it
wouldn't have convinced people and what I could have said then is
the only
reason that neural networks weren't working really well in the 1980s
was because the computers weren't fast
enough and the data sets weren't big enough but back in the 80s the

big issue was could you expect a big
neural network with lots of neurons in it compute nodes and connections between
them that learns by just changing the strengths of the connections could you expect that to just look at data and
with no kind of innate prior knowledge learn how to do things and people in
mainstream AI thought that was completely ridiculous it sounds a little ridiculous it is a little ridiculous but
it works and how did you know or why did you Intuit that it would work because the
brain works because that's you have to explain how come we can do things and how come we can do things like we didn't
evolve for like reading reading's much too recent for us to have had significant evolutionary input to it but
we can learn to do that and Mathematics we can learn that so there must be a way to learn in these neural networks
yesterday Nick Frost who used to work with you told us that you are not really
that interested in creating AI your core interest is just in understanding how the brain works yes I'd really like to
understand how the brain works obviously if your failed theories of how the brain works lead to good technology you cash
in on that and it gets get grants and things but um I really would like to know how the brain works and I think there's
currently a Divergence between the artificial neural networks that are the basis of all this new Ai and how the
brain actually works I think they're going different routes now so we're still not going about it the right way
that's what I believe this is my personal opinion but all of the big models now use a technique called back
propagation which you helped popularize popularize in the 80s very good
um and I don't think that's what the brain is doing explain why okay there's
a fundamental difference between two different there's two different paths to intelligence
so One path is a biological path where you have Hardware that's a bit flaky an
analog so what we have to do is communicate by using natural language also by showing people how to do these
imitation and things like that but instead of being able to communicate a hundred trillion numbers we can only
communicate what you could say in a sentence which is not that many bits per second yeah and so we're really bad at
communicating compared with these current computer models that run on digital computers it's almost infinite
they're able to that's a communication band with this huge yeah because they're exactly the same model they're clones of
the same model running on different computers and because of that they can see huge amounts of data because
different computers can see different datas and then they can combine what they learned more than any person could ever comprehend far more than any person

could have become and yet somehow we're smarter than them still okay so they're like idiot savants right
chat GPT knows much more than any one person if you had a competition about you know how much you know it would just
wipe out any one person it was amazing at bar trivia yes it would do amazing it would do me and it can do all you can
write poems it can you know um they're not so good at reasoning we're better at reasoning we have to
extract our knowledge from much less data so we've got a hundred trillion
connections most of which we learn but we only live for a billion seconds which isn't very long
whereas things like chat GPT have run for much more time than that to absorb all this data but on many
different computers 1986 you publish a thing in nature that is the idea we're going to have a
sentence of words and it'll predict the last word yes that was the first language model that's basically what
we're doing now yes and no 1986 was a long time ago why still did people not
say oh okay I think he's on to something oh because back then if you asked how much data I trained that model on I had
a little um simple world of just family relationships there were 112 possible
sentences and I trained it on 104 of them and checked out whether it got the last eight right okay and how would it
do it got most of the last eight right okay it did better than symbolic AI so it's
just that the computers weren't powerful enough at the time the computers we have now are millions of times faster they're
parallel but they can do millions of times more competition so I did a little computation if I'd taken the
um computer I had back in 1986 and I started learning something on it
it would still be running now and not have got there huh
um and that's stuff that would now take a few seconds to learn did you know that's what was holding you back
I didn't know it I believe that might be what was holding us back but people sort of made fun of the idea
that the claim that well you know if I just had a much bigger computer and much more data everything would work and the
reason it doesn't work now is because we haven't got enough data on enough compute that's seen as a sort of lame excuse for the fact that your thing
doesn't work was it hard in the 90s doing this work in the 90s computers were improving but
um yes so there were other learning techniques that are on small data sets worked at least as well as neural
networks and were easier to explain and had much fancier mathematical theory behind them
and so people within computer science lost interest in neural networks within
psychology they didn't because within psychology they're interested in how people might actually learn and

these other techniques looked even less plausible than back propagation here which is an interesting part of your background you came to this not because
you were interested in computers necessarily but because you were interested in the brain yes I sort of decided I was interested in Psychology
originally then I decided we were never going to understand how people work without understanding the brain the idea
that you could do it without worrying about the brain that was a sort of fashionable idea back in the 70s but
um I decided that wasn't on you had to understand how the brain worked so we fast forward now to the 2000s is there a
key moment you think back to is a turning point when it's like okay our side is going to Prevail in this
around 2006 we started doing what we call Deep
learning um before then it had been hard to get to neural Nets with many layers of
representation to learn complicated things and we find better ways of doing it
better ways of initializing the networks called pre-training and the p in chat
gbt stands for pre-training okay um and the t is Transformer and G is
generative and it was actually generative models provided this better way of pre-training neural ads so the
seeds of it were there in 2006 in by 2009 we'd already produced something
that was better than the best speech recognizers and recognizing which phoneme you were saying using different
technology than all the other speech recognizers were then then the standard approach which you've been tuned for 20
for 30 years there were other people using neural Nets but they weren't using deep neural ads and then there's a big
thing happens in 2012. yes to actually two big things okay one is that the
research we'd done in 2009 done by two of my students over a summer that led to
better speech recognition that got disseminated to all the big speech recognition Labs that Microsoft and IBM and Google
and in 2012 Google was the first to get into a product and suddenly speech
recognition on the Android became as good as Siri if not better so that was a deployment of deep neural
Nets applied to speech recognition three years earlier at the same time as that happened within a few months of that
happening two other students of mine developed an object recognition system
that would look at images and tell you what the object was and it worked much better than previous systems how did
this system work okay there was someone called Faith a Lee in her collaborators that created a big database of images
like a million images of a thousand different categories you'd have to look at an image and give your best guess
about what the primary object was in the image so the images would

typically have one object in the middle yeah and I
didn't have to say things like bullet train or Husky or and
the other systems were getting like 25 errors and we were getting
like 15 errors okay within a few
years that 15 went down to three percent which was about human level
and can you explain in a way people would understand
the difference between the way they were doing it and the way your
team did it I can try that's all we can hope for
okay so suppose you wanted to recognize a bird in an image okay the
image itself
let's suppose it's a 200 by 200 image that's got 200 times 200
pixels
and each pixel has three values for the three colors RGB and so
you've got 200 by 200 by 3
numbers in the computer it's just numbers in the computer right and
the job is
to take those numbers in the computer and convert them to a string
that says
bird so how would you go about doing that and for 50 years people in
standard AI tried
to do that and couldn't got a bunch of numbers into a label that
says bird so here's a way you might go about it at the first level
of features you might
make feature detectors things that you take little combinations of
pixels okay so you might make a feature detector
that said look if all these pixels are dark and all these pixels are
bright I'm
going to turn on okay and so that feature detector would represent
an edge here okay a vertical Edge you might have
another one that said if all these pixels are bright and all these
picks as a dark I'll turn on that would be if each detector that
represent in a
horizontal Edge okay and you can have others for edges of different
organs we had a lot of work to do all we've done is made a box right
so we've got to have
a whole lot of feature settings like that and that's what you
actually have in your brain okay so if you look in a cattle monkey
cortex it's got feature
detectors like that um then at the next level you might say if you
were worried up by
hand you would create all these little feature detectors at the next
level you would say
um okay suppose I have two two Edge detectors that join at a fine
angle
that could just be a beak so the next level up will have a feature
detector that detects two of the lower level
detectors joining a fine angle okay we might also notice a bunch of
edges that sort of
form a circle we might have a detector for that okay then the next
level up we might have a detector that says hey I
found this beak like thing and I find a circular thing in roughly
the right spatial relationship to make the eye and
the beak of a bird and so at the next level up you'd have a bird
detector that says if I see those two there I think it might be a

bird
okay and you could imagine wiring all that up by hand okay and so the idea of
back propagation is just put in random weights to begin with and now the featured textures would just
be rubbish whether it be garbage okay okay but look to see what it predicts and if it happened to predict bird it
wouldn't but if it happened to leave the weights alone um you got it right the connection strings but if it predicts cat
then what you do is you go backwards through the network and you ask the following question and you can ask this with a branch of
mathematics called calculus but you just need to think about the question and the question is how should I change this
connection strength so it's less likely to say cat I'm more likely to say bird that's called the ER the error the
discrepancy right okay and you figure out for every connection strength how I should change a little bit to make it
more likely to say bird and less likely to say cat and a person's figuring that out or the algorithm is set to work a
person has said this is a bird so a person looked at the image and said it's a bird it's not a cat it's a bird so
that's a label supplied by a person but then the algorithm back propagation
is just a way of figuring out how to change every connection strength to make
it more likely to say burden less likely to say cat it just keeps trying keep turning it just keeps doing that and now
if you showed enough birds and enough cats when you showed a bird it'll say burden when you showed a cat it'll say cat and it turns out that works much
much better than trying to wire everything by hand and that's what your students did on this image database that's why they did on the image check
device yes and they got it to work really well now they were very clever students in fact one of them Ilya
sutskova is also one of the main people buying chat gbt so that was a huge moment in Ai and chat
gbt was another huge moment and he was actually involved in both of them yeah yeah I don't know maybe it's cold in the room
you got to the end of the story I go shivers the idea that you do this little dial thing and it says bird it feels like
just an amazing breakthrough yeah I it was um mainly because
the other people in computer vision thought okay so these neural Nets they work for simple things like recognizing
a handwritten digit but that's not a real complicated image with sort of natural background with stuff it's never
going to work for these big complicated images and then suddenly it did and to their credit the people who've
been really staunch critics of neural Nets and said these things are never going to work when they worked they did
something that scientists don't normally do which she said oh it worked we'll do that people see it as a huge shift yes
it was quite impressive that they flipped very fast because they saw that it worked better than what they were doing yeah you make this

point
that when people are thinking both about their machines and about
ourselves in the way we think we think language in
language out must be language in the middle yes and this is an
important misunderstanding yeah can you just
explain that I think that's complete rubbish yeah um so if that were
true
and it were just language in the middle you'd have thought that
approach which is called symbolic AI yeah would have
been really good at doing things like machine translation which is
just taking English in and producing French art or something
your thought manipulating symbols was the right approach for that
but actually neural networks work much better than
Google translate when they switched from doing that kind of approach
to using your alerts really much better what I
think you've got in the middle is you've got millions of neurons and
some of them are active and some of them aren't and
that's what's in there the only place you'll find the symbols are at
the input and at the output
we're not exactly at the University of Toronto we're close to
University of Toronto at universities here and around
the world we're teaching a lot of people to code does this still
make sense to be teaching so many people to code
um I don't know the answer to that in about 2015 I famously said it
didn't make sense to
be teaching Radiologists to recognize things in images and because
within the next five years
um computers will be better at it yeah are we all about to be
Radiologists though well then Coopers are not better
I was wrong it's going to take 10 years not five I wasn't wrong in
spirit I just got I factor of two computers are now
comparable with Radiologists at a lot of medical images yeah they're
not way better at all of them yet but they will
get better yeah so I think there'll be a while when it's still worth
having coders and I don't know how long that'll
be but we'll need less of them yeah maybe or we'll need the same
number and they'll be able to achieve a whole lot
more um was talking about cohere we went over and visited them
yesterday you're an
investor uh in them maybe maybe the question is just like how they
convince you what was the pitch that convinced
you I want to invest in this so they're good people um and I've
worked with several of them
yeah and they were one of the first companies to realize that you
need to
take these big language models being developed to places like Google
and other places open Ai and
um make them available to companies so there's going to be it's
going to be
enormously valuable to companies to be able to use these big
language models um
and so that's what they've been doing and they've got a significant
lead in that so
that's why I think they're going to be successful another thing

you've said that I just find fascinating so I want to get you to
talk about it is
the idea that there'll be kind of a new kind of computer that will
be sent to this problem what is that idea
so there's the biological route to intelligence where every brain is
different and we have to communicate Knowledge from one to another
by using language and there's the current AI version of
neural Nets where you have identical models running on different
computers and they can actually share the
connection strength so they can share billions of numbers this is
how we make a bird yeah so they can share all the
connection strengths for recognizing a bird and one can learn to
recognize cats and the other can learn to recognize birds and they
can share their
connection strengths and now each of them can do both things right
and that's what's happening in these big language models they're
sharing but that only
works in digital computers because they have to be able to do
identical things and you can't make different biological
brains behave identically so you can't share the connections yeah
but why wouldn't we stick with digital computers
because of the power consumption you need a lot of power it's
getting less as
chips get better but um you need a lot of power to do this to run a
digital computer you have to run
it at such high power that it pays exactly in the right way whereas
if you're willing to run at much lower
power like the brain is then you'll allow a bit of noise and so on
but that particular system will adapt to the kind of noise in that
particular system and the whole thing will work
even though you're not running it at such high power that it pays
exactly as you intended
and the difference is the brain runs on 30 Watts a big AI system
needs like a
megawatt so we're training on 30 watts and these big a systems are
using because they've got lots of copies of the same thing
they're using like a megawatt so you know you're talking factor of
the order
of a thousand in the power requirements and so I think there's going
to be a phase when we train on digital computers
but once something's trained we run it on very low Power Systems
so if you want your toaster to be able to have a conversation with
you and you want to chip in it that only costs a couple of dollars
but can do chat gbt that'd better be a low power animal chip what
are kind of
like the next things you think this technology will do that will
impact people's lives
it's hard to pick one thing I think this it's going to be everywhere
right it's already sort of sort of getting to me
everywhere chat GPT is just made a lot of people realize it it's
going to be everywhere but it's already you know
when Google does search it uses big neural Nets to help decide
what's the best thing to show you we're at a
transition point now where chat gbt is this kind of idiot savant and

it's also doesn't really understand
about truth is being trained on lots of inconsistent data it's trying to
predict what someone will say next on the web yeah and people have different opinions and it has to have a kind of
blend of all these opinions so that it can model what anybody might say it's very different from a person who tries to have a consistent world view
yeah particularly if you want to act in the world um it's good to have a consistent world
view and I think what's good one thing that's going to happen is we're going to move towards systems that
um can understand Different World Views and can understand that okay if you have this world view
then this is the answer and if you have this other world view then that's the answer we get our own truths well that's
the problem right because what you and I probably believe unless you're an extreme relativist is that actually is a
truth to the matter certainly on many topics on many topics or even most topics yeah like the Earth
is actually not flat it just looks flat right yeah so do we really want a model that
says well for some people like we don't know that's going to be a big issue and we don't know we don't know how to deal
with other present yeah and I don't think Microsoft knows how to deal with it either they don't and it seems to be
a huge governance challenge who makes these decisions it's very tricky things you don't want
some big for-profit company deciding what's true but they're controlling how
we turn the neurons Google is very careful not to do that at present um
what Google will do is refer you to relevant documents which will have all sorts of opinions in them well they
haven't released their chat product at least as we speak right um but we've seen at least the people that have
released chat products feel like there are certain things they don't want to be said by their voice right so they go in
there and meddle with it so it won't say offensive things yeah but there's a limit to what you can do that way there's always going to be things you
didn't think of right yeah so I think Google is going to be far more careful than Microsoft when it does release the
chatbot yeah and it'll probably um come with lots of warnings this is
just a chatbot and and don't necessarily believe what it says careful in the labeling or carefullying in the way they
meddle with it so it doesn't do lousy things all of those things careful in how they present it as a product and
careful in how they train it yeah um and do a lot of work to prevent it from
saying bad things and well who gets to decide what a bad thing is some bad things are fairly obvious but many of the most important ones are not yes so that is a big open issue at

present I think Microsoft was extremely Brave to release chat GPT yeah do you
see this as like a larger some people see this as a larger societal thing we need either regulation or big public
debates about how we handle these issues well when it comes to the issue of what's true
I mean do you want the government to decide what's true speak problem right yeah you don't want
the government doing it either I'm sure you've thought deeply on this question for a long time how do we navigate the
line between you just send it out into the world and we find ways to curate it
like I say I don't know the answer and I don't believe anybody really knows how
to handle these issues we're going to have to learn quite fast how to handle these issues because it's a big problem with president but
yeah how how it's going to be done I don't know but I suspect as a first step
at least these big language models are going to have to understand that there are different points of view and that
completions it makes a relative to a point of view some people are worried that this could take off very quickly
and we just might not be ready for that does that concern you it does a bit until quite recently I
thought it was going to be like 20 to 50 years probably have general purpose AI yeah and now I think it may be 20 years
or less so okay some people think it could be like five is that silly I wouldn't completely rule
that possibility out now and where's pre a few years ago I would have said no way okay and then some people say AGI
could be massively dangerous to humanity because we just don't know what a system that's so much smarter than us will do
do you share that concern I do a bit um
I mean obviously what we need to do is make this synergistic
have it so it helps people and I think the main issue here well one of the main
issues is the political systems we have so I'm not confident that President
Putin is going to use AI in ways to help people like even if say the US and Canada and a
bunch of countries say okay we're going to put these guard rails up then how do you yeah it's it's
pretty for things like autonomous lethal weapons okay we'd like to have something like
Geneva conventions like chemical weapons people decided they were so nasty they weren't going to use them except just
occasionally but I mean basically they don't use them people would love to get a similar treaty for autonomous lethal
weapons but I don't think there's any way they're going to get that I think if Putin had an autonomous lethal weapons
he would use them right away this is like the most pointed version of the question and you can just laugh it off or not answer it if you want but what do
you think the chances are of AI just wiping out Humanity can we put

a number
on that it's somewhere between 100 percent
okay I think I think it's not inconceivable okay
that's all I'll say I think if we're sensible we'll try and develop it so that it
doesn't but what worries me is the political Citizens We're in yeah where it needs everybody to be sensible
there's a massive political challenge it seems to me and there's a massive economic challenge in that you can have
a whole lot of individuals who pursue the right course and yet the profit
motive of Corporations may not be as cautious as the individuals who work for them
Maybe I mean I only really know about Google that's the only Corporation I've worked
in and they've been among the most cautious they're extremely cautious about AI because they've got this
wonderful search engine that gives you the answers you want to see and
they can't afford to risk that yeah whereas Microsoft has
Bingham well if being if being disappeared in Microsoft would hardly notice yeah
but it was easy for Google to take it slow when there wasn't someone nipping at their heels and this seems to be
exactly yeah so Google has actually been in the lead I mean Transformers were invented at Google right the big
language models early ones were at Google but and they kind of kept it in your lab they're being much more
conservative and I think it might be so yes but now they feel this pressure yeah and so they're trying to they're
developing a system called bad that they're going to put out there and they're doing lots and lots of testing of it
um but they're going to be I think a lot more cautious than Microsoft you mentioned autonomous weapons let me
give you a chance just tell the story what's the connection between that and how you ended up in Canada okay there
were several reasons I came to Canada but one of them was certainly not wanting to take money from the U.S
defense department this was at the time of Reagan when they were mining the
harbors in Nicaragua and it was interesting I was at a big university in Pittsburgh and I was one of the few people there who thought that mining the harbors in
Nicaragua was really wrong [Music] um so I felt like a fish out of water and
you saw that this was where the money was coming from for this kind of work so that department almost all lemon came
from the Transformer you started to talk about the concerns that bringing this technology to Warfare could present what
what are your concerns oh that um the Americans would like to replace
their soldiers by autonomous by AI soldiers and they're trying to work towards that
and what evidence do you see of them I'm on a mailing list

from the U.S defense department I'm not sure they know I'm on the meeting list it's a big list they didn't notice
you're there you might be off tomorrow I might be off tomorrow what's on the list oh they just describe various things
they're going to do there's some disgusting things on there okay we'll discussed you the thing that disgusted
me most was a proposal for a self-healing minefield
so the idea is look at it from the point of view of the minefield when some silly civilian trespasses into
the Minefield they get blown up and that makes a hole in the poor my field so it's got a gap in now so it's not fit
for purpose yeah so the ideas may be nearby Minds could communicate or maybe
they could move over a bit and they call that healing and it was just the idea of
talking about healing for these things that blow the legs off children I mean
and the healing being about the Minefield healing yeah that disgusted me there is this argument that though the
autonomous systems might play a role in helping the Warfighter it's ultimately a
human making the decision here's what worries me if you wanted to make an effective autonomous Soldier
you'd need to give it you'd need to give it the ability to create sub goals in other words it has to realize things
like okay I want to kill that person over there but to get over there how am I going to get over there and then it
has to realize well if I could get to that road I could get there more quickly so it has a sub goal of getting to the
road so as soon as you give it the ability to create his own sub goals it's going to become more effective
and so people like Putin are going to want robots like that and but as soon as it's got an ability to create sub goals
you have What's called the alignment problem which is how do you how are you sure it's not going to create sub goals
that are going to be um not good for people not good for you
who knows who's on that road who knows on that road and if these systems are being developed by the military the idea
of wiring in some rule that says never hurt a person well that's they're being designed to
eat at people yeah do you see any way out of this is it a treaty is it what is it
I think the best batch is something like a Geneva Convention but it's going to be very difficult I think if there was a
lot of public outcry that might persuade I can imagine the Biden Administration going for something like that with
enough public outcry but then you have to deal with Putin yeah
um okay we've covered so much I think I have like two more things there's one more thing I want to say yeah yeah go
for it you can ask me the question some people say that these big models are just autocomplete well on some level the
models are autocomplete we're told that the large language models they're just predicting the next word is that not so
simple no that's true they are just predicting the next word and so

they're just auto-complete but
ask yourself the question of what do you need to understand about
what's being said so far in order to predict the next
word accurately and basically you have to understand what's being
said to bring language so you're just already complete
too um in the sentences they are you can predict the next word maybe
not as well
as chat gbt yeah but to do that you have to understand the sentence
so let me give you a little example from
translation it's a very Canadian example okay suppose I take the
sentence the trophy
would not fit in the suitcase because it was too big and I want to
translate that into French
well when I say the trophy would not fit in the suitcase because it
was too big you assume the it refers to Trophy I do
and in French trophy has a particular gender so you know what
pronoun to use yeah but
suppose I say the trophy would not fit in the suitcase because it
was too small now you think that it refers to suitcase
right and that has a different gender in French so in order to
translate that sentence
to French you have to know when it wouldn't fit in because it was
too big it's the trophy that's too big and when
it wouldn't fit in because it was too small it's a suitcase that's
too small and that means you have to understand about spatial
relations and containment
and so on yeah so you have to understand just to do machine
translation or to
predict that pronoun if you want to predict that pronoun you've got
to understand what's being said it's not enough just to treat it as
a string of
words yeah yeah I mean this gets me to another thing you've pointed
out which is kind of a
either exciting or troubling idea that you working intimately in
this field for
as long as anyone describe the progress as well we had this idea and
we tried it
and it worked and so we get a couple decades of back propagation we
have this idea for a Transformer now we'll do some
trick but it could there's hundreds of other ideas that haven't been
tried out yes so I think even if we didn't have
any new ideas just making computers go faster and getting more data
will make all this stuff work better we've seen
that as they scale up chat gbt it's not radically new ideas there I
think it's just more connections and more data to
train it with yeah but in addition to that there's going to be new
ideas like Transformers and they're going to make it work much
better are we close to the
computers coming up with their own ideas for improving themselves um
yes we might be and then it could
just go fast that's an issue right we have to think hard about how
to control that
yeah can we we don't know we haven't been there yet but we can try
okay that seems kind of

concerning um yes do you have any you're seen as like a
Godfather of this industry do you have any concern about what you've
wrought
I do a bit on the other hand I think
whatever's going to happen is pretty much inevitable that is one
person stopping doing research wouldn't stop
this happening if my impact is to make it happen a month earlier
that's about the limit of what one
person can do there's this idea of the and I'm going to get it wrong
the short Runway and the long takeoff maybe we
need time to prepare or maybe it's better if it happens quickly
because then people will have urgency around the issue rather than
like creep creep creep
do you have any like thoughts on this I think time for repair would
be good and so I think it's very reasonable for
people to be worrying about those issues now even though it's not
going to happen in the next year or two yeah people should be
thinking about those issues we
haven't even touched on job displacement um which is just my mistake
for not bringing it up
is this just going to eat up just job after job after job after job
I think it's going to make jobs
different people are going to be doing the more creative end and
less of the routine end but what's the creative if
it can write the poem and make the movie and all of that well if you
go back in history and look
at ATMs these cash machines came along and people said that's the
end of bank tellers it wasn't actually the end of
Bank tell us um the bank tellers now deal with more complicated
things and take coders so people say you know
these things can do simple coding and usually get it right you just
need to get it to write the program and then
just check it so you'll be able to work 10 times as fast well either
you could have
10 of the programmers well you could have the same number the
program as producing 10 times as much stuff yeah
and I think there's going to be a lot of trade-offs like that you'll
the once these things start being creative
they'll be hugely more stuff created this is the biggest
technological
advancement sense is this another Industrial Revolution what is this
how should people think of it
I think it's comparable in scale with the Industrial Revolution or
electricity electricity maybe the wheel or maybe the
wheel yeah that was earlier yeah okay so buckle up
yeah one of the reasons I got a Toronto got a big lead in AI
is because of the policies of the granting agencies in
Canada which don't have much money but they use some of that money
to support curiosity-driven basic research okay and
so in the states the funding comes you have to say what what
products you're going to produce
with it and so on yeah yeah some of the government money quite a lot
of it is given to
professors to employ graduate students and other researchers to

explore things
they're curious about and if they seem to be good at that then they get more money three years later
and that's what supported both Joshua Benja and me it was money for curing
curiosity driven basic research and we've seen that before even through Decades of not being able to show much
yes even through decades not being able too much so that's one thing that happened in Canada another thing that happened was there's a Canadian
organization called the Canadian Institute for advanced research that provides extra money to professors
in areas where Canada is good and provides money for professors to interact with each other when they're
far apart like in Vancouver and Toronto but also to interact with researchers in other parts of the world
um like America and Britain and Israel and so on and see far setup of programming AI is set
at one originally in the 1980s which is the one that brought me to Canada which was in symbolic AI
yet you came I was an oddball okay I was kind of weird because I did this stuff everybody else thought was nonsense they
recognized that I was good at this kind of nonsense and so if anyone's going to do the nonsense it might as well be him
one of my letters of recommendation said that it said you know I don't believe in
this stuff but if you want somebody to do it Jeff engines to go okay um and then after that program finished I went
back to Britain for a few years and then when I came back to Canada they decided to fund a program in deep learning
essentially sentience I think you have complaints with the even just how you define that right yeah I
when it comes to sentience I'm amazed that um people can confidently pronounce
these things are not sentient and when you ask them what they mean by sentient they say well they don't really know
so how can you be confident they're not sentient if you don't know what sentient means so maybe they are already
who knows I think whether they're sentient or not depends on what you mean by sentient so you better Define what
you mean by sentient before you try and answer the question are they sentient does it matter what we think or does it
only matter whether it effectively acts as if it is sentient
it's a very good question Matt and what's your answer I don't have one sure okay because if it's not sentient
but it decides for whatever reason that it believes it is and it needs to achieve some goal that is contrary to
our interests but it believes in its interests does it really matter if in any human I think a good a good context
to think of this in is an autonomous Lethal Weapon yeah okay so it's all very
well saying it's not sentient but when it's hunting you down to shoot you um
yeah you're going to start thinking it's sentient we're not really

caring not an important standard anymore the kind of
intelligence we're developing is very different from our
intelligence so it's this idiot savant kind of intelligence
yes so it's quite possible as if it is a tool center is essentially
in a somewhat different way from us but
your goal is to make it more like us and you think we'll get there
and my goal is to understand us oh okay no but yeah and
I think I think the way you understand us is by building things like
us okay so that's I mean the physics is called
Richard Feynman said you can't you don't can't understand things
unless you can build them that's the real test of do
you understand it and so you've been building so I've been building
yeah