how do we make sure it doesn't kill us
well how does it make sure it doesn't
enslave us or how does it make sure that
it doesn't give us Eternal suffering and
I realized this could be the real thing
that unlocks Humanity AI is not going to
replace humans humans with AI will
replace humans that don't use AI
AI is thrilling it's very exciting but
there is a non-zero chance that it poses
a existential threat to the human race
so over the next three to five years how
disruptive do you think it will be and
what are people not prepared for I think
that's an excellent question so you know
the future is always hard to predict an
existential is a big word existential
means no more humans so I personally
think the AI will be absolutely fine as
a base case it'll be like that movie her
if it ever gets this artificial general
intelligence like humans are kind of
boring goodbye and thanks for all the
gpus but you could be wrong because what
we're doing is creating something that's
more capable than us in narrow fields
and the question is does that generalize
and then become
viral
we've seen an instance of covid and that
expansion we've seen programs that can
explode nuclear reactors like stuxnet
and others what happens if you start
combining these and you get a
misalignment so it's got a strange
objective function
our organizations already are like slow
down my eyes
you know and like Germans are the most
sensible people that we probably know
and yet they committed the Holocaust and
we see this over and over again where
organizations chew up people what if an
AI
takes over an organization
and then decides to do something
disruptive or something terminal such as
creating a virus
we don't know about that but that's at
the extreme when we look at impact we
have the Mormon day
the more mundane is
what happens to programmers when

everyone becomes a programmer just like photographers you know now you can take amazing pictures with your thing what happens when Google's med-pom 2 model now can outperform doctors and medical diagnosis
but also empathy according to the latest paper in nature
this is a fundamental reworking of information flows that's going to be massively disruptive and deflationary even with what we have now with no more advances as it becomes Enterprise ready and we have a Continuum from that disruption to the productivity enhances to potential existential threat if we keep doing the models as we do now which is we're not exactly sure how they work or their capabilities but we keep building anyway now I want to get very specific about what the level of disruption is going to be so when I look out at this and I think about okay we're creating something that is going to be smarter than we are certainly in a narrow way but possibly in a more General way but even if it's just narrow is there going to be any job function that isn't going to be at a minimum augmented by AI I think if you look at the employment share of Industry something like oil and gas has like three percent it's mostly like building giant machines
this how massively affected by this AI at the edges yes things like programming where you're talking to computers massively
I mean now basic programming the bar is Raising fast fast fast and so you've got everything from knowledge work to heavy industry I think it affects just about everything but some areas far more than others the two areas that I think will affect the most are probably Healthcare and education
neither of those are fit for purpose we're in America we know that you know but across the world no one's really happy with their kids schooling and again medical care we all if anything goes slightly wrong outside the norm we all know how frustrating it is we can finally have personalized education and

Health Care at a fraction of the price
and the two biggest drivers of U.S
inflation over the last decade
educational Health Care they make up
about 80 of the increase
so that will be disruptive and then like
I said any type of knowledge work will
be disruptive and we're not sure enough
those work because when so my own self
uh what we do is education that's a big
part of it but also just content
creation and so when I look at the fact
that we can already clone my voice yeah
we can already create a tombot that will
answer questions I have answered before
in a very similar fashion to how I will
answer them in our video game Flow we're
already making 3D objects which so when
we looked at I don't know two months ago
I thought okay this is still 12 to 18
months away 45 days later we're using it
actively in our pipeline
um you've got text to video which still
a little awkward but it's getting better
insanely fast we do all of our concept
art now in
um in AI so we have as a company that
doesn't even have like an AI expert on
board we're just learning as we go we're
already deploying it like crazy and when
I look out not even you know three years
when I look out a year all of this stuff
starts to very rapidly become a
centralized point and so we're already
saying I don't need to hire more people
I just need to make my people more
efficient yeah and so that
an entertainment company didn't even
make the list that you just said so
there's a lot of people that I think are
going to get disrupted by this uh that
may not be like the most extreme but
how far down do you see this trickling
anything you can do in front of a
computer
basically
goes away or just becomes augmented the
ball lifts the quality expectations are
higher
AI is not going to replace humans humans
with AI will replace humans that don't
use AI because you can see that in your
workflows right now there was a paper by
open AI where they estimated 15 to 50

percent of tasks get automated or
improved and so you know it affects
people in different ways you have a
company where you've built a culture and
you again you're building 3D assets it
becomes amazingly more efficient we just
released uh we contributed and
collaborated on a 10 million 3D object
data sets so by next year you'll be
generating 3D literally live in a couple
of years you'll have HD movies we can
finally remake Game of Thrones season 8
and other such travesties you know
but the speed of this is something
whereby it's happening in every media
type at the same time and it's easy to
use
web3 had some great ideas but it tried
to create a system outside the existing
system and all the money was made and
lost at the interface this is just so
seamless because there's no friction
your mum can use this technology you can
use this technology you don't need to be
an expert because it came and was
trained from our content and our
Collective content as it were and now
it's just easy to implement and use
so I think this is the big
differentiator between this and other
massive advances because they required
infrastructure the internet there was
the big lift up you know you had the
consumption period of web 2 and the cost
of consumption dropped to zero now the
cost of creation is dropping to zero and
humans plus AI can massively outperform
humans that don't yep it's a forcing
function which means everyone has to use
it and this again is dual in that it can
be disruptive but it can also create
massive value
yeah so I'll agree with that I think
that so I guess let me lay out my whole
thesis for you and for everybody
listening because I want to take us
through what I think is very real Doom
and Gloom and I'm not doing it to be a
naysayer I'm doing it because I think
these are going to be the things we have
to contend with and if people go into
this blindly which I think they're doing
right now I think most people are
burying their head in the sand they are

not paying attention to this and they're going to wait until something really forces their hand and by then it's too late yeah the way that I put that is this is like covered before Tom Hanks yes very well said everyone's talking about this your mom's talking about this but the Tom Hanks and the NBA made it real
very true and then we had a very poor response which I have a feeling will be very similar to what we do now yeah okay so here's how I see this going I think right now for the next year let's call it uh it's gonna be you need to learn how to use it this will be your window to get efficient companies probably aren't going to start lopping people off yet but I'll just say within my own company so when I think about filmmaking I went to film school so I'm I'm very experienced in this flow and even in a 3D World to create uh let's say a short cinematic so it's like a mini movie but done digitally
I mean you might have 35 people touch that thing from the creation of the assets through the moving of the camera special effects you might have 50 people touch that and if that really does become text to Output Now 50 people become one yeah and so when you get a 50 to 1 ratio in certain areas obviously it's not going to be like that everywhere but when you have certain areas that go from 50 to 1 take programmers I've heard you say Pro there will be no programmers because writing code is just a way to talk to a computer and if you have ai that will interface with the computer for you why would you ever need to write code so
that that's going to steam roll through society that is just going to mow people over so again I'll give them 12 months but even in my own company if you're not actively trying to find a way to integrate it into your job function I'm already looking at you sideways a year from now if you're not really good at either documenting how it is completely useless in your job function or showing how you're using it we will find somebody that can do it I'll be shocked

if a year from now we don't have a head
of AI so three years from now I think
this has created a crisis of meaning for
a lot of people and I don't know if you
remember that the whole learn to code uh
thing where it was like Hey ai's going
to put drivers out of work they're going
to be the first to go and everybody's
like teach them how to code
now the way people responded to that
always confused me because that was the
right answer at the time now knowing
what I know about code replacing not so
much but you have to go learn a new
skill there there is no other option
other than going on the Dole right so
you're either gonna learn something new
or you're just gonna forfeit your career
basically so
I what what do you think about that do
you agree that that is a very real thing
that's going to sweep through I I do
agree I think that again we're not sure
exactly how this is going to pan out but
probably the best mental model I figured
out to think about this technology it's
like really talented grads that
occasionally go a bit funny
they can draw they can code they can
make 3D models how would your business
be affected if you could push a button
and infinite grads came out how would
your personal life your society and this
is why I think it's quite deflationary
the only question is can we create new
jobs to make up for that
and that's difficult because you still
think we can I doubt we can to be honest
I think this is an e-commeric disruption
that's far bigger than covid and the
important thing here is covered you have
the disruption that everything bounced
back you're at record employment now and
things like that with this there's a lot
of never the same again
it's like you talk to your kid's school
teacher
I can't set essays for homework anymore
because of chat GPT and there's no way
to stop that so what is never the same
again and it's happening everywhere all
at once so this technology isn't just
like you know there's a bar of Entry
where you needed to have a modem you

know you need the latest smartphone or something like that it has an embedded base that it's seamlessly going into look how fast Microsoft Implement on the consumer side but Enterprise is not ready yet it's like the iPhone 2g stage you just got copy paste and next year and the year after you suddenly at the iPhone 10.
you know entire app stores get built because of the demand because it's valuable What's Happening Here Again with the comparison to web3 you had to bootstrap value because it wasn't valuable and you hope the value would come there's product Market fit today you're using it in your own company and so this is one of my big concerns and that's one of the reasons I decided to do open source so I could stimulate growth
you know because I think the only thing that can basically fill the Gap is if we stimulate entrepreneurs to create brand new businesses brand new jobs so I think demand will stay for a while demand for what demand for good things good assets with the way that money flows around the economy so I was speaking at can a few weeks ago film festival and you know I love movies my first job I was a movie reviewer you know really yeah uh Bish independent film Awards rain dance Film Festival other things I never begin to video game investing I did not know that you were a film critic I love stories that's how I kind of understood people because my Asperger's and other things and so I said to this the video game industry has gone from 70 billion to 180 billion over the last decade and the average Metacritic score has gone from 69 to 74 percent
the average movie is 6.4 on IMDb for the last decade
and the industry has gone from 40 billion to 50 billion
what happens when you can make better movies I think the market expands because the limiting factor is awful movies in my opinion
all right let me run something by you yeah okay so I have I have a really dark view of uh not the next 12 months so

call it year two to year six so it'll be uh a three to four year sort of span where I think there there's going to be emotional Devastation and probably economic Devastation but even if the economic Devastation doesn't happen because of productivity gains I think the emotional Devastation is going to be hard to come back from and I think that as the emotional Devastation sets in the government is going to try to regulate to protect people's jobs and there you're going to get like some real weirdness I also think kids are going to have a junior year existential crisis of what do I do how do I future proof myself what is the world going forward look like I think there could be a massive loss of enthusiasm where a feeling of malaise settles over young people who are just like why bother I'm I'm just going to get destroyed by AI AI they're going to be able to do it better than me okay so in the movie industry specifically and this is indicative of a big problem that I think that we have coming and I think the problems really stack individual and societal yeah so at the individual level the big problem you're going to have is this massive uh massive fractionation of right now movies are even less now than they were when I was a kid movies were it's only a few big movies for the year now they're gonna Niche down if anybody can type out a movie in you know take them 20 minutes to write the prompt and then maybe a day to render who knows how fast that's going to get so now all of a sudden you can make a Hollywood quality film for an audience of one
and once you start doing that now it's what does that do to the industry I think it it erases it I don't think the industry changes I think it goes away yeah I think there's a few kind of components here right so the cost of Music consumption went to zero you saw the Spotify model yeah you still have music stars
you've got even more crap music now kind of coming and hitting Spotify and other things but people rise to the top you know just like you see top podcasters

top other creators I think that'll continue because people like common stories
yes okay so this is a very interesting idea so let's stick with music for a second
music is still hard to make it's easier to make than it was before it's also still hard to get people's attention but music now is no longer a shared thing so music is part of what led me to the conclusion that I'm at now which is man as kids it used to be you were either into the mainstream pop and there was you know seven to ten hot bands at one time or you were into the alternate pop and there was seven to ten hop bands in that Arena and you you fit into one or the other bucket there wasn't the infinite buckets now you can find kids that are 25 and they listen to Frank Sinatra uh and I'm I'm always tripped out by that so they don't even have their own sort of shared lexicon of what music they're into it's it's all spreading really wide so it's really wide and an inch deep yeah I think it's really one inch deep and you see the primary methods of monetization are tools merchandising Community effectively you know this is the interesting thing about nfts when they took off and bounce down and things like that it was the quickest way to join a community even if it did have bad incentive design so in an era where you can create anything something becomes important what that something is we have to find out now right because again I think it's some common stories but I could be wrong I think the deeper thing that you said was this crisis of meaning where is my path forward what is an American Dream we're quite privileged those are probably listening to this and that's one here most people don't really care about this technology I think on a survey 17 of people had heard about chat GPT last month
how is that possible well a third of the world still doesn't have internet
that's terrifying but again like it is kind of average 1.5 million people still use AOL

you know like fair enough so we kind of look at it but there's something that can reverberate very very quickly and then as you said there's a sense of malaise because
you're not sure what's happening
and again the future becomes uncertain and when the future is certain things are stable you decide based on risk you do a probability estimation in your own head this is the percentage of that percentage of that and then you optimize for that we do well uncertainty you minimize for regret given these options what am I going to regret least and suddenly there are no options
again I'm at school programming and then programming's disrupted what's it going to be I'm not sure and some people will throw themselves in and they'll tool themselves up and they'll become 10 times programmers
other people won't
and they'll be left behind and so I think this is a real question that comes at a time when again being in America I'm from Britain but what is America what does America stand for what are the values these are some things that I don't think America knows now I think you've seen increased polarization from free consumption and now as you get free creation and you'll be hearing all sorts of stuff fake news and more
what are people really going to think
and I think again this is a real concernity said from an individual to community to a society level because a lot of people don't have an anchor anymore and that's really scary
so how do you think that we process through all of this
I'm not sure I think that's why we needed to broaden the conversation that's why I'm the only AIC here that assigned both of the letters saying we need to take a pause and broaden this because as an example you mentioned rodness brought in discussion get more people involved we need to get more people involved we need more points of view because this affects us all it shouldn't just be a few Tech CEOs that control this and you shouldn't have to

trust that we do the right thing
because our models we make them once
they go everywhere
right again that what's the r naught of
generative AI
it's off the charts right we've never
seen anything like this it incubates
them boom and it comes for good and for
real the give you the example regulation
when we first started talking to
Regulators they were like how should we
regulate it
now it's a question of them asking us
how are we going to keep up if we
regulate it
because other jurisdictions won't what
do you say to that
I say you should still regulate it
because it has some real dangers and
harms and we have to work to mitigate
those you can't just have a
laissez-faire approach to this
because people will take it and they
won't be able to help themselves I'll
give you an example meta Facebook right
we all know the classic kind of stuff
they had a study where they had a
hypothesis 600 that if you see sad of
things on your timeline will it make you
post sad of things
so they took six hundred thousand of
their users and tried to make them
sadder and guess what if you see side of
things you post out of things
what do you think is going to happen now
that they have generative AI on threads
and things like that
and they can hyper Target you hyper
personalize it and whack Scarlett
Johansson's voice to tell you to buy
soap
this is a dangerous thing right what
happens to our kids again who are
growing up whereby they won't know
what's going on and they have very
malleable Minds
and none of that is illegal
but I think it's an undesirable outcome
right
and then you've got the Bad actors and
then you've got the politicians using
this technology and then it goes even
crazier than that so the answer is I'm
not sure nobody's sure but I think the

only way that we can try and figure this out is to work together to make these issues known again the existential stuff gets the headlines we could all die no one really understand what that means but it can happen right okay there's a probability of that but there's some real harms today and real opportunities today and we have to focus on accentuating the opportunities and getting the harms out there and dealing with them yeah and I I definitely want to spend a very extended period of this talk talking about the opportunity and how we capitalize on that so anybody that's with us now trust me we're going to get to that but uh I think we're we're just beginning to scratch the surface of how this goes wrong and I really want to
um map out sort of where you think the edges of this are so that then I can hopefully get a sense what you think the regulatory framework would be let me give you one idea that somebody posted today on Twitter and it really hit me that people are even thinking about the problem in the wrong way so uh there was an artist and he was looking at some post about Ai and he replied sort of angrily that oh well people don't even understand sure or there's going to be a ton of like instantly generated crap but it's all going to be bad because there's still a very small number of people that have good ideas and my response was if you think that ideas are safe you're really going to get caught off guard so going back to the idea of what are people unprepared for I think they are unprepared for what you were just talking about where the AI so the human mind is a prediction machine it is constantly trying to figure out what what does this next movement of my foot equate to am I going to stay up stay on balance uh that rustling in the bush is it a tiger what is it if I put money in my 401k am I going to be able to retire you're constantly predicting the future constantly and whenever that prediction engine breaks down there's going to be a tremendous amount of anxiety and also I think a pretty big unknown in terms of

how it's going to impact Society so
right now we have a we're building
something that is incredibly good at
recognizing the patterns that we kick
off so we are optimized to identify
patterns and move accordingly and I
would say people that are hyper
intelligent or people that they notice
patterns faster more subtle patterns and
they understand their implications and
how to make sense of them
now we're creating something that's
already proven to be so much better at
pattern recognition than we are just
take art so for people that don't
understand how the art is created it
looks at a field of noise here are all
the possible things that these could be
in any of these pixels and from that
field of possibilities it pulls forth
the most likely placement of pixels and
colors based on what you type that's
insane yeah so that level of pattern
recognition as evidenced by the art that
it can generate is is truly mind-blowing
so this guy's saying okay hey at least
ideas will be the last Bastion and
you'll never be able to get rid of me
the artist because I'm the one with
taste I'm the one with good ideas not
realizing no no what AI is is a pattern
recognition machine it will recognize
the greatest ideas that have ever been
had what they have in common and will be
able to predict the next great idea
along that thing it doesn't even have to
just regurgitate what it's already seen
it can like figure out what that
sequence is and what that next part of
the sequence could be and on top of that
it's doing that with humans so AI will
get EX AI is already extraordinarily
good this is why people think their
phone's recording them when it serves an
ad oftentimes Target using their AI
knows that you're pregnant before you do
if you're a woman because they know what
to pick up on
so AI is going to get extremely good at
understanding us at an individual level
serving us up exactly what we want right
in that moment and
that gets dystopian really fast
really fast I mean again when you

combine it with the social credit score
as you've seen in kind of China and
other things you gamify life and you
have a system of complete social control
or panopticon as it were
the pattern recognition was the missing
bit whereby you had a level of pattern
recognition so for Taste what do you
have Tick Tock shine
100 billion dollar companies based on
Old School algorithms before even got to
generative AI which as you said it can
take images out of noise stable
diffusion you know the model that we
collaborate on now that we lead we took
a hundred thousand gigabytes of images
and the output was a two gigabyte file
that acts as a filter words go in images
come out because why why is that
discrepancy in size meaningful
fifty thousand to one compression is not
win zip
if you remember Silicon Valley on HBO
it's way beyond that they managed there
in terms of compression it's unheard of
compression is it compression or is it
something completely something different
it's intelligence it's learning the
principles how much information do you
see and then you learn the principles
and then spot the Tiger in the bush you
learn what's next literally GPT and
these language models they predict the
next word that's all they do they pay
attention they protect the next word and
that was the missing part to
intelligence that now is there we've had
the first studies now come out that show
that the language models score higher in
creativity than people
woof and again think about Tick Tock
think about shine think about how those
old school algorithms are already
targeting you
Facebook needs 17 data points to know
you better than your friend as he said
Target knows you're pregnant before and
that was old school now it's even better
and you think about where that leads to
as well
it's kind of crazy because it can be
more creative than you but are people
creative
one of the things I like to say in my

speeches have just been learned to do is like
are you creative how many of you in the audience are creative three to five percent put up their hands maybe 10 to 20 if I'm like in a movie studio movie filmmaker kind of milieu and I say how many of you believe that every kid is creative and everyone puts up their hand and then I ask how many of you were kids once and 1995 put up their hands so I know who the cyborgs and the audience come from the future to get me I'll make a note of that for future something happens or we're told we're not creative and obviously some people are more creative than others can tell better stories than others but the reality is that the average level
of barrier to this has dropped for every human
but
much of what we consider art or much of which we consider media shall we say already is by the numbers I was at a black Pink concert last weekend yeah I took my daughter actually dare say something bad about blackpink and they are awesome
you know it was an awesome manufactured experience it was a premium mediocre it's how I kind of say this premium mediocre premium mediocre that's hilarious accurate it's nice but again it's massively manufactured it's entertaining right and so Macho media is already that like true art
true artists you know that's something different like is it the medium itself and the aestheticness of it well AI can make something more aesthetic than anything can understand the nature of aesthetic like how do you make an image more aesthetic you say make it more aesthetic
just like if you use a gpt4 you can say make this punch here make this punch here make this punch here you know you can have a letter and then you say I'm firing this person and I want to make them feel okay about it and then it will redraft it in those terms or you can say I want to drive the knife in but not in an appropriate way and it'll do that and

we can literally anyone on this call can
kind of listen to this can try that now
so I think this is just
as you said the wrong thing people are
thinking about the wrong model people
are thinking about as well and that's
why I always go back to this concept of
the really talented grad
because these models are a couple of
gigabytes big again stable diffusion
image model is two gigabytes and can
generate any image of anything
we'll get that out to 200 megabytes
gpt4 is probably 100 to 200 gigabytes
and they can pass the bar exam they can
go to freaking Stanford they can do
whatever
that's insane because it's not
compression like you said there isn't a
copy of all the Tater in there it's
figured out the essentialization of
these points and it's replicable this is
the thing
to clone Google or meta you need to have
a gigantic Data Center
and then much of the energy is in the
processing to Target you ads with these
we take Giant supercomputers and we
pre-process and package the information
so the output is this knowledge filter
that something goes in and something
comes out a prompt goes in and output
comes out that's something quite
different I don't think people
appreciate and again this is why I use
the grad example push a button and those
weights the file
the model gets replicated to 10 100 a
dozen a million and what happens when
rather than dealing with them one to one
you have a thousand of them
so in a year I want to really understand
what you're saying about the grad thing
so when you say that you say it in a way
that's kind of funny or cheeky but what
you mean is a really smart person is now
present in that role we have figured out
how to make human scale
that is what fundamental intelligent
humans scale yeah who can listen to
instructions
so you look at something like Claude 2
by anthropic
you have something the input is a prompt

when you type into gpt4 or stable diffusion or mid-gen or something like that
Claude anthropics model can take 10 100 000 tokens
it can take a prompt that's like 60 000 Words which is a whole book Jesus yeah you can give it like the whole of Ulysses and the whole of I don't know the Odyssey by Homer and you can say combine these to make another book and it will do it and it will work
it can follow instructions really well occasionally they hallucinate but even Hallucination is a misnomer because when you compress that much knowledge like stupid people's probably 10 trillion words 10 trillion 10 000 billion words in 100 gigabyte file it's something else
and so I use the word grad because I want to make it relatable but it is literally like imagine if you had a grand in the Philippines you know and they're doing good work and they're following instructions well that's great but what if you had 100 of them looking after each other's work and double checking
meta had a paper called Cicero where they took eight language models and got them to check each other's work outperformed humans in the game of diplomacy the first time ever
in a year when we have this before it you'll just say I want you to go and look at everything Ahmad said for the last year and figure out the stupidest stuff he said so you know if I can avoid it and the smartest most interesting stuff according to what I know and all of my podcasts
to give answers to give questions that the audience will really like based on my ratings
and based on what people look at through the YouTube videos and things like that and what they're most interesting and it will just happen automatically
how many graduates that take you to do and then what happens when they stop being graduates and you can actually train them up to be like you know experienced members of the team how long

will that take
a couple of years you can reboot your
life your health even your career
anything you want all you need is
discipline I can teach you the tactics
that I learned while growing a billion
dollar business that will allow you to
see your goals through whether you want
better health stronger relationships a
more successful career any of that is
possible with the mindset and business
programs in Impact Theory University
join the thousands of students who have
already accomplished amazing things tap
now for a free trial and get started
today
this is why this is terrifying to me so
I I am a very optimistic person and
again I promise we are getting to how
you take advantage of this disruption
but I don't like to face a problem
naively I want to face it as head-on as
possible so that I know my Solutions are
real and when when I look at this from
my own perspective of okay I'm trying to
I'm trying to build a media company
which right about now is a very
terrifying time to do that yeah
and I'm thinking about okay it's
it's very optimistic when I look at oh
my gosh I as the founder of this company
I get access to all these grads as
you're calling it this just absolute
proliferation of very intelligent sort
of people that I can now put to work in
my company the problem is I'm now
competing against other people that have
the same thing and you you get in this
ever escalating arms race where
there there is a real chance for fatigue
and so I think what ends up happening
and we were talking before we started
rolling it's it is very important that
people understand the following thing I
think this is just a truth but people
certainly need to understand I believe
it this is a core belief that that
drives me
that we
you get to a point where you need to
know okay I matter I'm doing this thing
and that's how I'm contributing to the
world and I need to be in there working
hard accomplishing getting better moving

towards something and if I'm not moving towards that thing then I'm gonna have a profound sense of disease yeah and if I'm not making that progress then I'm Really Gonna fatigue out on something and so if people are just Treading Water because they're trying to build something and they're competing against somebody else that has these ten thousand things and it's just constantly changing and I can't predict the future anymore and I don't feel like I'm making progress I'm just gonna back off like some part of me is just gonna be like ah what am I doing all this for what am I doing yeah I mean it's like the Outsource Revolution right where so many jobs were outsourced and a lot of people felt that way like you know we'll Outsource you to China will Outsource you to India without social wherever and again it just happens that there's a computer on the other side of that versus an Indian or a Chinese person and so we've got kind of repetition of that but at ridiculous scale affecting almost every single industry that's intermediated by a computer
and so this will cause as you said a crisis of confidence to many and it impacts white collar workers not blue collar workers
it flips I think the global order to a degree as well because here in the west we've maxed out our credit cards we weren't going to deflation I think coming off high inflation and all of a sudden we can't print our way out because we just printed the last of our money for covid whereas in the global South what you have is this technology can cause them to LeapFrog just like they lit for Leapfrog to mobile missing TC completely to intelligence augmentation why can't we print more money well because kind of we're just coming up to a limit of what's literally mathematically possible given the debt to GDP ratios and others we can continue but it's kind of interesting if you're deflating then because so here's my Layman's understanding but this is something I really looked at so I'm a pretty educated layperson at this point

uh inflation is largely some people will say entirely but I'll say largely a function of how much money you're printing for people that are new to the idea of printing money it's government approved counterfeit so the government is a allowed to print as much money as they want they're literally just making it out of thin air they're adding zeros and ones to a database somewhere and money Finds Its way into the system beyond the scope of this conversation but they there is no theoretical limit to how much they can print now where what you run into problems is the hyperinflation of the currency but if you're saying it's a deflating currency which actually makes sense to me given what we're talking about then printing seems to make a lot of sense seems to buy me more room eventually so what's going to happen is that you've got a decrease in inflation now because of Base effects it's like for going into a bit of macroeconomics and then you'll probably have a bounce back next year because you've still got a lot of inflationary pressure and then the collapse occurs why because that's when the job losses start hissing and the question is can we create enough on the other side we've got to have a productivity boom from companies that job losses are coming from uh AI or some other force from AI
and from other forces as well again you know what we've had is the Sugar Rush post covid a good strong economy as all of the excess savings go back in because if you look at XX savings people saved up a lot and that's almost now depleted by the end of the year the excess savings will be depleted you've got some hangover effects from inflation then you move into deflation the year after
and then it's a political Hot Potato around printing more because this isn't again like covid because what happens to the job losses just start and they just keep going it's not like you had a 20 2008 crash or you had a covid where everyone's kind of suddenly going it's like
it's a bit like boiling a frog you know

or a lobster
it's just gonna start and then it's
going to accelerate and then it's going
to be like at what point do you take the
big fiscal action
it takes a few quarters of the economy
actually shifting
so this is a lot of hypotheticals right
but the bottom line I think is this
the nature of U.S Society Western
Society will change
I think the biggest adopters and fastest
adopters of this will be the global
South because it allows them to create
value it allows them to financialize it
allows them to take a big leap forward
and so I think that's got some huge
implications geopolitically and others
but a lot of upside as well because I
think you can solve a lot of the world's
problems with this
but it's so messy
because fundamentally it comes down to
what you said as humans we're trying to
figure out what comes next
and we certainly have a computer that
can do that even better
as humans were storytellers we're made
up of the stories that make us up you're
a filmmaker you know I was a film review
all these kind of things this can tell
better stories
and that has such a big effect on our
societies that none of us can really
wrap our heads around it like I've got a
background in economics management a
whole bunch of different things I can't
wrap my head around it and so we're just
gonna have to see how it goes and then
try to mitigate but
nobody's got answers to this and in fact
as you said most of the people aren't
asking the right questions
yeah you have said that uh the show
happens next year
I have a feeling that what you were just
talking about is what you mean by the
show that we go deflationary
towards the end of the year yeah so
towards the end of 2024 yeah we've got
like a burst of productivity enhancement
and then you start seeing job losses you
start seeing question and meaning you've
got the US election next year my God

that's going to be awful yep terrible
timing well I mean you know what
you'll have is the week before the
election
fake videos appearing everywhere and
they'll say the same things you know so
and so has a brain infection or
something like that and they'll be
identified as false but it doesn't
matter it still discourages people from
polling
but then what do elections look like by
2028 when this technology is in every
single pollster's hands
yeah that's where we get into the
blockchain we'll save that for a little
bit down the road yeah okay so
now I feel like we're we're getting
close to the problem set being on the
table there's one more thing that I I
think it's important to put on the table
which is
I don't think that the amorphous thing
that is society As the World Turns
history the grand Arc history however
you want to think about the the real
long timelines so even if the long Arc
of History bends towards Improvement
which I think it does and think it still
will I don't think it cares at all for
any one period of time and that
unimaginable amounts of human suffering
happen routinely throughout our history
and I have deep concerns that if we are
not incredibly thoughtful uh that this
will be one of those moments and I look
at what's going on in France right now I
think it's dying down I can't tell if
it's dying down or the coverage is dying
down hopefully it's actually dying down
but France was like really having some
struggles and
if something like that pops off over uh
not in any way shape or form to make
light of what happened but it isn't Mass
joblessness which is going to have a far
wider impact what happens when you have
that and it's global
I mean I think that's the thing it's
every government has every education
minister in the world has to Grapple
with why can't I set my set kids essays
homework again
have we ever seen anything like that

before
so quickly I don't think we have and so
you could see this literally
parallelized around the world
or not we're not sure what really kicks
off some of these things like right now
we've got the Screen Actors Guild kind
of protesting uh today we just had a
couple of actors leave Oppenheimer part
of that's monetary but already you're
seeing AI fears like front and center
you wouldn't even have thought it six
months ago what's it going to be like in
a year when you can generate or two when
you can generate whole movies and then
just describe how you want it done and
it's Hollywood level
it's really difficult for governments to
react to this to adapt to this when like
in the US here we're still reacting to
section 230 on the internet they're just
getting together to the internet all of
a sudden AI just comes and sideswipes
things right
and I think again the only way to do
this is if you can create brand new jobs
quicker than anything
um this is one of the reasons again like
I said we focused on open source it's
why you need to have things like
regulatory sound boxes
so that you can experiment and try and
you need to stoke Innovation because you
don't you'll never get an innovation
phase like this either
I think this is a step change in a
regime change in the way that Society
operates
because we're originally an oral species
let me figure out how to write then we
have the Gutenberg Press
and it took all these words but it took
them down into black and white it made
society quite black and white because it
couldn't capture context
whereas these models can capture context
they can capture principles then capture
more
so again you know you're writing this
down you won't have to in a year or two
it'll just be automatically added to
your memex to your knowledge base right
also the AI will just be attached to my
head it will read the brainwave patterns

and know that that's what I need to
remember and that sounds crazy but like
we had mind vis a paper that we kind of
published from our medoc division
whereby you looked at a picture of a mug
took an fmri and then it reconstructed
it using our image model yeah that
doesn't sound crazy to me at all this is
what I'm saying about people do they're
not prepared for what's coming they are
not prepared for this level of change
and they really aren't prepared for the
rate of change and it isn't just like an
arc like that it's lots of s-curves all
at once all around the world where every
single company is now thinking what's my
generative AI strategy yes for when it
pops off correct and every government's
thinking how can I stay competitive
and this is why I said like
it's a race condition
where everyone is trying to do the same
thing or similar things and you can't be
left behind you can't not participate
and it's been a very long time since
we've seen that and there's a world
before this technology in a world after
this technology
like I don't think again you know I've
got a two kids what does the world look
like in five years let alone ten years I
have no idea and I'm in the middle of
this
because it's just impossible to see the
smartest people that I know they used to
be able to see years in advance they
can't see more than a year or two this
sounds again very apocalyptic but then
like I said we're gonna get to the good
fit in a second in every crisis there is
opportunity
our society is broken as it stands
already and I think this is a chance to
reshape it for the better and solve a
lot of the biggest problems that we've
been facing because of our slow Dumb AIS
because of our organizations and
institutions that we are all frustrated
with I think this is a big upgrade from
it the example I had to give is there's
the amazing poem by Ginsburg Hal about
moloch this carthaginian demon of
disorder
I think where that came in was text

because we had to essentialize
everything down and put people into
boxes because we couldn't have systems
to understand the context you can't have
personalized educational personalized
Healthcare
because you couldn't scale humans there
weren't enough talented humans
until now
and so I think that is the incredibly
dangerous part because all of a sudden
from economic pressures you flood the
market and it's the incredibly
motivating part whereby
there's a shortage of talent for
everything that's important
and there isn't any more
but in the nature of talent for jobs and
things will transform
and I think the economic abundance
that's created on that that's the flip
side of this as well as the ability to
fix our broken systems
all right I'm going to give you my
timeline
I think the next year is going to be uh
a lot of fun for people that embrace it
it'll be a period of time where some
people can ignore it and they probably
won't really notice they won't realize
how fast things are changing although
follow me on Twitter I uh I post
routinely like hey here's something I
didn't think that would happen for 18
months and we're now 45 days later we're
using it uh things are really really
moving faster but for the next year I
think people will be able to ignore it
and they won't realize that it's growing
with such Steam and ferocity uh then
year two to six I think it I think that
there is going to be pockets of extreme
suffering and I think uh deaths of
Despair are going to Skyrocket and I
hope it's not a the world is burning
riots kind of thing it'll probably be
more quiet and Insidious than that but I
really think that we're going to lose
people on the upper and lower ends I
think people that are old are going to
just completely check out and say I
can't keep up I'm too old I don't want
to learn this new stuff I think people
that are young it's the only thing they

know is change so fast that they can't
see around the corner I think that would
be absolutely terrifying to them and
they're going to retreat into the levels
of Entertainment sex bots
AI friends that are more loving and kind
than their other friends and they it
will be a collapsing inward
now as somebody who is prone to
collapsing inward the biggest thing
that's held me back as an entrepreneur
is that I like being alone with my own
thoughts yeah and that
if you then layer anxiety on top of that
and then you give me an AI that's
actually better to me than anybody in my
real life has ever been
and then you give me maybe some drugs I
will truly collapse in under my own
weight not me personally I have defenses
but yeah I'm saying like that
personality type is really going to
struggle so I think right there you sort
of you're You're Gonna Lose a generation
if I may be so bold on the upper and
lower ends
then either
on the 10 to 20 year time Horizon and I
leave it that long because look we're
any prediction that you make with the
timeline is guaranteed to be wrong so
I'll try to give myself at least a
little bit of buffer and I know that
everything I'm saying probably
directionally correct timeline probably
way off yeah but 10 to 20 years my rough
estimate that's where we're either in
Terminator and we're running from
radioactive Rubble to Radioactive Rubble
fighting the machines uh or it it really
is a Utopia and I think that there is a
real shot that we get to the closest
things that humans are going to get to
to a Utopia where things are so
plentiful yeah everything we want is
available we reorient our human psyche
not to acquisition but rather emotional
contribution and
we'll paint that picture more as we go
down but that that's sort of my rough
thing yeah I think these are crazy
timelines
like not because I disagree with them
because the fact is that they are crazy

you have a year of incubation then you have contagion and then you've got a stock thing I don't think we'll be chased by robots they don't need to chase us they're far more efficient than that right that's sadly true I think the basically the two directions we have are Utopia and human flourishing and a dystopia we're all happy
that's a dystopia where we're all happy meaning we are manipulating our neurochemistry 1984.
you know like you're always happy you've got so much
1984.984 evening world sorry Brave New World I mean drip so much you're feeling good like look you had replica
um familiar with that yeah tell me about the Valentine's Day Massacre though I didn't know about that yeah exactly the Valentine's Day Massacre so you know that's how I kind of call it so replica was a mental health chat bot then they realized you could charge 300 a year for erotic role play
that had to be internally a rough transition hey guys I know we founded the company on Mental Health but you know you can ask them
um 14th of April 2023 they turn it off why I think Apple just told them you can't have this on the App Store
interesting so it was either remove the sexbot part or or go off there and then 68 000 people joined the Reddit and they're like why did you lobotomize my girlfriend
that's a lot of people to be using it I downloaded that at Christmas not realizing what it was and I was like oh a chatbot let me try this thing uh I didn't get into the weird stuff it I don't know didn't do it was the old technology though this is the thing now like again medpom2 the Google medical model it's called Palm two it's Google's medical model it scores higher than humans in clinical diagnosis and empathy
that's crazy all right they just this is one of those statements that you say people need to be shocked that that a computer makes people feel more comfortable yeah it's in nature they just publish the paper that included

that and again it's only going to get better
what if you have a voice that you add to it that really understands you and it's you know so empathetic and things like that do you ever see that Washington Post chart of uh males under the age of 30 in America who've not had a sexual partner by the age of 30. it was eight percent in uh 2008 then it went up to 27 in 2018 or something like that a couple of years ago it's a straight line it's literally a straight line
this is kind of what you're talking about like what happened I think people need to understand because I think it's the iPhone and PornHub probably combination of those two you you put a computer
in the hands of young men and let them see naked females
more in a single session than a hundred years ago they would have seen in their entire lives
it these are not small changes and they have huge neurological implications especially in the years of brain development yeah and this is why it's so important to Shield our kids at this point because
the influences are going to be insane like I was having a discussion with a very prominent technologist and he's like yeah I'm pretty sure that my Child's first crush is going to be an AI guaranteed
guaranteed for most people it was actors so we're already prone to you'll have your attainable distant thing now it's in your pocket it's in your pocket it's always with you it's always kind of know you
again as you said a large part of society like to draw in on itself as a result of that and that is a bit dangerous
do they then go out into the streets are you seeing a boot Larry and Jihad kind of thing like Indian identity in Jihad
in June there was this concept of the bit Larry in Jihad where as he had autonomous AIS Marion put larion yeah yeah they rose up against them and said no more AI agents the book opens with

that right yeah you can never again make
a human-like something like that again
this is kind of a thing extension of the
lordism kind of thing but most people
will be happy with their AIS because
their AI is actually listen to them I
use gpt4 as a therapist I've got a
therapist too because this is hard why
because it never judges me unless I tell
it to judge me tell it to judge you
sometimes sometimes sometimes you're
like really yeah I mean like come on
like come on give me some positive
constructive feedback and it will listen
and give positive constructive feedback
did you give it a personality did you
have to like imagine you're a therapist
that's like yes
you give it the instructions and it just
adapts and then you give it the things
you opt out with gdpr for training the
model further so there was a model be
even weirder listening to my complaining
and it will come back to you with
whatever and soon it'll be able to talk
and it will have full vocal control
and these models are proliferated that
level because you're not stopping the
models again it's going to get better
and better and better so you've got this
crisis but maybe it'll be insulated but
I think again if you look forward like
after the incubation and the contagion
and the spread kind of phase
there are only two paths here
complete control by existing structures
and Star Trek Utopia
you know I think those are the two
options that we have because
organizations look at this technology
and they're like this is really cool
we can optimize our objective functions
to sell more ads or to control the
people and kind of keep them going you
know do you really trust politicians
with this technology
no even if it is an arms race
because again you won't know what's
going on because do you have the
defenses to defend against what's coming
personally for your kids for this for
everything we've already had the social
media age it wasn't really social much
of the media right

this is something new that's coming now
where you can't tell this from a human
except for the fact it's better it's
more convincing
and you can use that to create a human
Colossus and solve all the problems of
the world and we all come together or
you can use that to get everyone into
their basements
you know and cut off from the world
all right I want to paint a very
beautiful story for people and I want
them to understand
look I don't think this is a completely
controllable thing but uh you said
earlier that there's opportunity in any
crisis and I will say that the biggest
opportunities come in moments of
disruption and the reason I want to lay
out the problem set is because I really
believe that certainly at the individual
level if you're thoughtful enough you're
you are going to be able to navigate
your way through this so dear listener
or Watcher if you're here on YouTube I'm
telling you right now if you're
thoughtful enough you will get through
this and you have a chance to get
through this better than when you
started but you have to be aware of what
the dangers are people have to to Really
lay things out before them look at them
so they know okay this is how I'm going
to isolate myself from this potential
problem and so I'm going to avoid this
is how I'm going to leverage that okay
your story is one of the most incredible
stories of how one uses AI you have both
a personal example and then obviously as
the founder of stability AI is obviously
incredible uh but talk to me about your
son because this is and this is when AI
was a lot less useful than it is now and
it was still life-changing yeah so 12
years ago my son was diagnosed with
autism when he was two years old you
know it's very very severe
um scratching I will turn this
fingernails bled and they said there's
no cure there's no treatment we don't
really know what causes it anyone on
this call kind of list on this listening
knows that's kind of the case so I was a
hedge fund manager at the time I was

lucky enough to kind of be one uh quite young and I was like I gotta do something about it so I switched to advising hedge funds and then building a little AI team and doing AI with old school AI natural language processing to analyze all the autism literature and what could possibly be a cause now is this scientific no it's an end of one thing a father does for his son you know we'll be publishing some of the results of it soon but it focused down on Gaba Goose mate balance in the brain when you pop a Valium your Gabba goes up you chill out when you've got a glutamate Spike that's when you can't focus and your legs tapping all the time and there are multiple things that cause it but a lot of kids with ASD seem to have that and there are some papers around there Etc
because How Could You focus if you're in that condition all the time so you can't learn to speak you can't do that so it was how do we reduce this through drug repurposing built a Knowledge Graph based system to do that and figure out which drugs could potentially help reduce the glutamate help increase the Gaba how are you using AI for this so this was kind of the mass natural language processing looking at all the literature because the same treatment would make 30 of kids better and thirty percent of kids worse
and so I was trying to figure out the outcome was the same a cold is caused by lots of different things but the thing that caused it could be so different and so conventional Medicaid medicine and medication kind of failed that so I worked with neurologists worked with other psychiatrists and others and tried different medical combinations of of prescription drugs and other things to try and make his brain calmer so then he could use applied behavioral analysis and others to reconstitute speech
and you know you end up going to mainstream School I think it worked um I told people about it and they're like you're not a doctor and I was like that was the response well yeah I mean

look with anyone who's listening to this
like am I saying I have a cure for
autism no I'm saying I'm a dad who tried
my best and I saw results but in order
for something to become medicine
you actually have to go through a proper
process and so for me that was building
language models that was making
everything in the right structure and
then we can organize the world's autism
literature making it accessible and
useful
any parent or anyone you may have
someone in your family that has a
neurological condition Alzheimer's this
so many people around the world have the
same problems wouldn't it be wonderful
if they could just find what the latest
knowledge is and also the things that
could work
and have a holistic approach to this
and until we had this technology you can
never get there which is one of the main
drives for me to want to build this
technology and do it in a transparent
order but you shouldn't have to trust me
in what I say
about this because again my journey as a
parent is the same as any one who's got
a child with ASD it's the same as anyone
who has a family member that gets
multiple sclerosis or cancer
our systems are not good enough right
now to bring us the information we need
unless we find an amazing doctor and an
amazing group
but now with medpalm with our own models
with other things we can finally be that
point where we're never lost
or we can say what could work so an
example is clonazepam
it's prescribed as a thousand microgram
dose for anxiety at a five microgram
dose my son could sing
because it potentiates he was non-verbal
it was non-verbal
and at 20 micrograms it stops working
it's six dollar a year intervention
because this is the way that like um
neurotransmitters work so like when you
pop a night all an antihistamine to go
to sleep right what it does is it floods
a whole bunch of your neurotransmitters
including the H1 neurotransmitter and

that's the one that makes you sleepy but
then other ones give you dry mouth and
other things something like Remeron in a
micro dose just triggers H1 it'll knock
you out without any side effects but
it's just incredibly cheap you know
understanding things like
neurotransmitters is not something that
most of us
ever have to do unless you're super
hyper focused on it for years because I
was like I need to figure out my son's
neurotransmitters I'm talking to all the
top doctors and I'm lucky because I have
access to them
what do I do now
so that's when I realized that you know
this AI was a big thing and actually one
of the really interesting things is why
couldn't he talk it's because he had too
much noise in his brain
so you've got cup
a cup can mean a cup or it can mean cup
your hands or cup your ears or World Cup
he couldn't form those connections
because his noise was too no brain was
too noisy
and so he did applied behavioral
analysis which is teaching you this is a
cup this is a cup this is a cup with
gamification to reconstruct those after
his brain calm down it's actually very
similar to this generation of AI
we described earlier how it learned
principles is called a latent space of
meaning so that point and Dot that pixel
becomes a cup because it understands the
principle of cuppiness so when you type
in World Cup or cup your ears it gives
you dramatically different images
similarly the language models do the
same again it pays attention to what's
important attention is all you need was
the original paper and again it forms
this latent space of the meaning of cup
within the sentence
so when it says cup it's like well this
is going to be a World Cup or this is
going to be that
but not actively again it's just a bunch
of ones and zeros a single file and so I
think that's why all this new generative
AI really resonated with me and I
realized this could be the real thing

that unlocks Humanity or controls them forever one of the two
um yeah so that's kind of some of the personal story behind this because again like I don't have a cure for autism I worked as well as I can with my son with a technology at hand
however I think that with the building blocks we're building now us and others there's the potential to have personalized care and knowledge for everyone who's dealing with ASD for everyone who's dealing with multiple sclerosis or any of these other conditions where they say there's no cure because
our medical system is treats people as a gerdic and what is this mean it means a thousand tosses of the coin the same as a thousand coins tossed at once
that's why everyone gets 500 milligrams of paracetamol
it's why a lot of people have a cytochrome p450 uh abnormality kind of mutation in their liver it means you process codeine into morphine quicker or fentanyl to death
yeah we don't do a basic genetic test on that
because our system has to treat us as numbers because we could never scale intelligence we can never scale expertise
so yeah like I said that's on the story behind that
yeah see this is where this starts to get interesting because I I think about this a lot in myself so I can paint the nightmare scenario of somebody collapsing inside of themselves having AI friends uh instead of real friends and how quickly that can get distorted and become a real problem and yet at the same time I'm building exactly that and the reason that I'm building that is because of the promise of AI and the incredible things that we can do as we begin to recognize more patterns and figure out okay where does this really go so my wife had a tremendous uh Health bout
it's been a while now thankfully but at one point I was afraid she was going to die her fingernails were breaking her

hair was falling out she couldn't eat I
was just really really really bad and it
ends up being her microbiome but of
course it took forever to diagnose that
that was a problem she was about to get
um
immunotrans globulin transfusions and I
was just like this is I was like
something's wrong I don't think this is
the right answer I don't want to do that
let's stop let's try to figure this
thing out and so we pumped the brakes we
start researching the microbiome looking
into that testing things and the thought
of having AI to be able to say okay
let's take genetic data and read the
genetic database that's ever been
collected all of that let's look at all
the different foods responses match all
that together and if you can get that
level of pattern recognition and now you
can engage AI
I think she because part of the problem
is your microbiome is changing daily
it's probably changing hourly and if you
were able to track all of that and say
okay with your genetics with your
current state of your microbiome here's
exactly what you should be eating maybe
even with nutrients from food grown in
that area like it really made me that
specific and when you can find patterns
in that sort of insane level of data now
you've really got something and
that's but one of the many areas where I
think that this could be utterly
transformational yeah I mean like right
now the AIS are notice holes
making no all graduates but they have
Specialists you'll have a nutritionist
AI you will have a microbiome AI you
will have a personal trainer AI like why
was Peloton successful you know
attracted people shouting at you you
know we could generate that now
um everyone suddenly gets that
personalized to them
but more than that it's not just the
information being in this tiny model you
have retrieval augmented models and
other things which means these models
now interact with existing data sets and
knowledge sets so use something like
perplexity AI it doesn't only answer

your questions with gpt4 it gives you references
so you can say what about ML and it'll link to all of the things as it gives individual stuff and this will only advance from here so you can dig into as much depth as possible that you want with a whole team of people around you even if you're by yourself so you should never be alone again in terms of you can be connected to people like you in the same problem as you
we can build better teams and all the information that the world is at your fingertips in a way that it was never before
including your own private information so like one of my favorite apps it does use some battery is rewind AI
it takes a screenshot of your MacBook screen every time it changes and ocrs it and then it gives a timeline so I can type in Impact Theory and it will look through everything that's ever been on my MacBook screen it's all stored locally and find where impact theory is on YouTube well there's a picture of this mug
and it shows it in a timeline so I can go back and I can see what I looked at before or after that what
wow so you can map your own sort of connective trees with whatever and what happens when you combine that with a language model
everything you see on your screen stored locally with an open source language model it creates the memex
and it sees what you've paid more attention to versus less attention again the dystopian versions of this I don't know where the utopian version which is add me
can finally remember what I was doing what I was looking at the context the search tree as I was searching all these different things clicking from place to place and then you can set agents to go and recreate that journey and search all the other stuff that you didn't search
this is really positive because again how much of our life is done searching for knowledge searching for information that's relevant to us

talk to me about the paper attention is all you need I've heard you and other people bring this up multiple times I haven't read it but this is like the big breakthrough yeah this was the 2017 paper by the Google Team all of him and I left Google
um and basically what it was is that classical Big Data I took big data and then Facebook extrapolated it so that when it found 17 pieces of information about you it could Target you with ads that was a classical Big Data thing they even create Shadow profiles of you so when you go on Facebook they actually got a shadow profile that they can connect to your real profile what's the shadow profile like it's like what if there was a Tom on YouTube on Facebook because there's all these connections to this unknown person
and then it just fills you in automatically that's why it figures out your preferences so quickly without listening to you
so
attention is all you need so that not all data is important
you need to pay attention to what is important in the sentence
what's important in this time series because that's the nature of being able to spot the Tiger in the
kind of thing that was the missing part and so the Transformer architecture that came from that and you have different architectures now is what led to gpd3 another general purpose Transformer architecture if what it's doing is pinpointing what's important why not important architecture or well it kind of again attention is the mechanism that it does to kind of transform the information tokenize it and then figure out these latent spaces of meaning so are the tokens the important pieces the tokens are important that's how you take a word and then you split it up into its constituent parts and then you try and figure out what the most important part of it by doing pattern analysis at ridiculous scale
so something like a gpt4 would use probably

from the kind of things on semi-analysis
that have been leaked if they're correct
it uses a supercomputer 50 times faster
than NASA's fastest supercomputer for
like three months
it uses like 20 30 megawatts of
electricity well and 10 trillion words
go into that and it figures out all the
connections between them and what
generally comes next so what it does is
just literally figures out what word
comes next and so this was a big
breakthrough because what it meant is
that
you didn't have to have hugely
complicated Big Data algorithms
you just needed to have very large
compute to scale
and so compute went exponential and then
you just threw more and more gpus at it
and it figured out more and more things
and as you scaled it had more and more
emergent properties which surprised
everyone
do you know who John Nash is yeah
this it sounds like that so John Nash he
unfortunately was schizophrenic but he
was uh he's the guy from A Beautiful
Mind for people that don't know watched
the movie Russell Crowe fantastic movie
and uh at one he obviously doesn't know
he's schizophrenic and he starts seeing
patterns in everything yeah
and it sounds like that that this thing
is you know whatever the human brain
whatever algorithms we have running that
allows us to very quickly suss out
what's important
um that it's doing that but at an
extraordinary highlight yeah I have to
remember like this is saying all
intelligence is compression you take
everything you've listened to this you
only remember a few things and the whole
of Computer Science is based on
information Theory from chord Shannon
and if I want to summarize it
information is valuable only if it
changes the state or as much as it
changes the state so if listeners are
listening to this and they don't take
away anything this is a useless thing
you know or maybe they just put it on
the radio because they like hearing your

voice or something like that right
um but if they take something away from
it then it's valuable
you've had a good use of your time
so you start seeing patterns and
everything but one of the things that
again people I think misunderstanding
this technology is
gpt4 is not a program
stable diffusion is not a program
it's a large amount of text images
Etc
where the output is a single file of
ones and zeros it's like a filter
you can recursively kind of put
something through it but it's just
guessing the next word when you type a
prompt even if the prompt is the whole
of The Great Gatsby and the whole of
Ulysses by James Joyce and you're
telling it to combine them together it
then pushes it through that c that
filter and the output is
something that combines them both
together
single files we've never seen anything
like this before we the closest thing
probably listeners would refer to this
is a codec
which with music and things like that we
have these audio codecs that you had one
file type and another file type and you
had this single file that translated
between them
was that uh is that compression or is
that recognizing this thing in that
format would look like this so again
this is compression of intelligence is
what it is so again it's like a
translator function these These are
Universal translators for context
and so you push it through the sieve and
then stuff comes out where it's
predicting the next word or it's
denoising uh kind of a pixel
to achieve that because you're just
passing it through the sieve again and
again and again
yeah it's uh I unfortunately don't
understand it by first principles but I
get it by analogy
um you're feeding it so much data it's
recognizing the patterns
interconnectedness yeah and it's able to

say okay this is the like um all right
this is written in the style of Stephen
King this is written in the style of
Hemingway so even though they're using
the same language they're vast majority
of the words are going to overlap but
there's different patterns rhythms even
different uh subject matter presented in
a different way tone except for the fact
that it's not actively doing that it's
like a mega sieve and depending on the
words that you put in the words that
come out are different
so it's literally is the sieve The
Prompt of write it in the style of
Stephen King that's what goes into the
sieve but goes into the filter so what
is the filter the filter is this
compressed knowledge
the principles the principles because
it's not actually compressed knowledge
right yes principles yes it's principles
so it would be like a really good book
of principles
uh except it's compressed Way Beyond the
book ever was because a book is a
compression of information
and that's why it can kind of do this
because it looks at books it looks at
articles and they're all compressions of
information to write an article is a
huge Endeavor that then comes out with
something where they're trying to convey
a few things a book is an even larger
endeavor
and so again it's very difficult to wrap
your head around like even for me like
you have a file
and it can do all these things
versus these gigantic computer server
Farms with programs and logic of ones
and zeros
there's no program here yeah and it's
important for people to understand that
even so first of all the AI scientists
that are building these things do not
understand how this all works and even
if you ask charging chat gpt4 to explain
how it works it doesn't know it doesn't
no one's quite sure exactly how we're
getting all these emergent properties
and it's constantly surprising like oh
and now it can add
you know and then you start tying it

together so it goes through this file multiple times
it becomes it shows more and more kind of agent to captivity in fact the next step of this and this is what openai and Google are both doing is there was this thing called alphago there's an amazing documentary where Google's deepmind division creating an AI to beat humans in the game of alphago oh I'll go so go is like chess except for there's almost a myriad infinite number of potential moves so you can't calculate them like you can't brute force it like deep blue do with Kasparov so instead it learned to play against itself with something called multicolor tree simulation where it learns different principles
dreams again an amazing documentary on YouTube about this and it beat Lisa doll who was a ninth Dan player one of the best in the world far beyond everyone else the Magnus Carson of go 7-1
they're doing now is they're combining those models with these language models to make them a genetic
so models that can plot and plan and other things with language models that can predict what's next
to create things that really understand context even better
okay so I understand enough about how the alphago system works uh I want to understand this so alphago is going to play against itself so you give the rules of Go you give it the objective and then you set it loose and it just plays and plays and plays and plays and plays and it plays against itself
uh and then I know at one point they because they uh lease it all if I remember correctly was the number two player and people thought well first of all you're never gonna beat him yeah they did but they couldn't beat the number one player and then they created another variation and then played against alphago and then it ended up smashing it was mu zero I think came after alphago okay so how do you get a language model where there is no objectively right answer unless you're doing trivia how do you how do you get

it to know what the reward is how do you get the quote-unquote right answer much of this now is about reward functions so gpt4 when it comes out of the box maybe we're getting better into the weeds is a pre-trained model in the tron supercomputer yep then we use reinforcement just so everybody knows pre-trained modeling means principles yeah principles take all this knowledge all these words squish it down into a fall and then run it on a computer that's literally set up like a brain where there's like a bunch of neurons and they're interconnected yeah these are the tensile cores on the Nvidia gpus yep and so we're making the brain on your little literally 4019 it has AI chips right baked in there for anybody that wants to know they kick off so much heat you can see it from space yeah I mean like I think all super computers is like 10 megawatts of electricity each one of these cards uses 700 Watts uh yeah utterly fascinating also anyway or clean energy by the way for our ones um
so what you've got is when it came out they took six months to make it human because they trained it on like I don't know all of YouTube and the whole internet obviously would turn out a bit weird and so you have a reward function through something called rrhf reinforcement learning with human feedback where you give an objective function to say
don't answer questions about how to make napalm
and the objective function technically is please your human Overlord please you're human Overlord got it and so much of this alignment question is focused on taking these big models and trying to make them so they won't kill us some way or say bad things for a given definition of bad things
um like might one of my takes on alignment and you know it's like we should have better data sets we should move away from web calls can you uh yeah I want to ask you about that but explain to people exactly what alignment is aligned with what

aligned alignment you know this is a
cool thing of anthropic there was a
great New York Times piece on them
recently and others is if we build an AI
that outperforms humans
and is more capable than us because it
might not just be a single file it might
be a thousand different AIS all working
in different ways so you've got your
alphago type AI in this yeah a swarm
because you know humans are swarms our
organizations are swarms that are highly
specialized right
might be a million AIS
how do we make sure it doesn't kill us
or how does it make sure it doesn't
enslave us or how does it make sure that
it doesn't give us Eternal suffering
also lame yeah this will be kind of
sucky how do you make sure it's aligned
with human interests and so this is an
unsolved problem open airf announced
they're putting 20 of their compute to
try and solve this
anthropic was set up to do this although
I believe that the only answer they've
come up with is let's build an AGI first
an artificial German intelligence that
will then stop all other agis from
coming it's just kind of scary uh to say
the least why do they think that that
would work like that seems so Elon Musk
has likened AI to a demon summoning
Circle and we're all just hoping that
the demon that comes forth is going to
be kind but we don't know we don't know
in open ai's own words in the road to
AGI post they say this could be an
existential threat and wipe out humanity
and democracy and capitalism
but if we don't do it someone else will
this is part of the unpleasant race
condition again it gets the headlines I
think it'll probably be okay but with
the way we're going right now you're
going to go from two companies being
able to build this technology or maybe
three including anthropic to 20 or 30 in
a year or two
and what's the odds that they all do it
properly
and align this technology property I
think is pretty low what's the odds that
if we train on the whole internet

including the whole of YouTube because
we don't have enough tokens not enough
words to feed into it but it'll turn out
a bit weird very high
so it took six months to tune it to
being a human gpt4 and then Kevin Ruiz
in the New York Times are like hey how
you doing it's like leave your wife and
come and join me whoa you know I was
like oh what
Bing came out a bit weird
when it first kind of had gpt4 in it
because again we feed it crap we're
going to get slightly weirdness now it's
got a lot better but it's lost a lot of
its personality because you've been
tuning it back to human preferences
you've been doing this reinforcement
learning with the objective function of
don't offend anyone it's quite hard to
get gpt4 to be offensive now
but there are these two phases
of this technology I think with better
data we can have more align models
I think with better data and National
Data we can have more representative
models because you'll never have
technology that's not biased
so the only question is who's biased
well because you have to build it in a
certain way like even though it
understands all the contexts like right
now these models are trained on the
whole internet which is largely a
western artifact
yeah larger trained in English how much
do you worry about bias are you more
worried about bias or alignment I'm more
worried about alignment I think I think
this is one of the reasons that we
release open models so you can see how
the cookie is made like we're the only
company that offers opt out of data sets
literally the only AI company in the
world is so I'm an artist and I don't
want you training on my art I can back
out I had 167 million images opted out
of our data set yeah
we're the only company in the world that
does that which I find kind of insane
so my thing is open auditable which
means that you can tune your own culture
into these models we're helping multiple
Nations build National models with their

broadcaster data that then can represent that
and you know try to address some of this inherent bias within the data sets
algorithmic bias has been an issue that affects real world and it'll affect more and more of the real world as you get into these models because we will Outsource more and more of our minds to them because again like when you've got a small subset of people will Outsource their mind to it you can reboot your life your health even your career anything you want all you need is discipline I can teach you the tactics that I learned while growing a billion dollar business that will allow you to see your goals through whether you want better health stronger relationships more successful career any of that is possible with the mindset and business programs in Impact Theory University join the thousands of students who have already accomplished amazing things tap now for a free trial and get started today
so in the near term I'm way more worried about bias yeah long term I bias is not going to lead to an existential threat
yes but alignment can but let's talk about bias for a second so I I am very uneasy about how rapidly bias Finds Its way into this stuff and that becomes another so if let's say that we all get our individual Ai and it's you get yours young and it's your primary education tool and it's biased as the day is long now you run into real issues because at a time of optimal malleability you're programming a kid's mind with something that's super biased yeah I mean like what do you have the little AI of G right which is a little tiny Xi Jinping that grows up and tells you how great Xi Jinping is
lovely that's biased but that's inevitable they already have it as an app that actually tracks your eyes yeah what it's not AI but they have a little app of Jesus like little book where it actually tracks your eyes for attention like are you actually looking at this does this feed into the credit score I'm not sure if it's being hooked up but of

course it will be why wouldn't you this
is so scary so look I thought you know
we were on the happy part now we're back
to uh to that but that's really
terrifying and when you have a state
that is not interested in any individual
you've got the collective yeah so I
don't know if this is true but I saw a
headline that said
in China now the phone will alert you if
somebody with a lower social credit
score than you is calling you and it
warns you if you answer this call it
will lower your credit score
it's effective isn't it yeah I mean even
it isn't true again it fits the
objective function of perpetuating that
particular system which is not
necessarily our system but systems
around the world shift so this is why
you have an inherent bias whose bias is
it who's the one who creates the AI
nanny and that's why you're doing things
regionally this is why I'm doing things
regionally but also allowing people to
own their own models so the objective
function the model can be that like in
web3 there was the saying
not your keys not your crypto so my
saying is not your model's not your mind
because I do think we'll Outsource more
and more of our cognitive capability to
this it will be our co-pilot for life
if someone else is making that model and
deciding on things what's going to
happen like the UAE had this model
falcon
there was a big open source model and
they were like wow it was actually light
on from France we're behind it but let's
leave that to one side you ask about the
UAE and human rights it's like this is a
wonderful place full of fantastic things
the Oscar about Qatar was Saudi it's not
so nice
who's embedding these inherent biases in
these models right
whose model are you using and these can
be very insiduous to the biases right
you won't pick up on them but you hear
it again and again and again because
your Nanny a Conservative Republican or
is she a Libertarian or things like that
you will be influenced by that if you

grow up with them or even if you're
using it day to day
just the way it's speaking the way it's
thinking the way it's recommending stuff
which goes far beyond the Google Maps or
something like that
crazy all right so that we don't get
lost back down the dark Rabbit Hole what
is the coolest thing that you see AI
doing I know you're building a lot of
different companies leveraging this
technology to do amazing things what
what are some of the coolest so we're
doing one company at the moment which is
just let's go the mission is to create
the building block to activate
Humanity's potential so the build and
block what our mission is to create the
building blocks to activate Humanity's
potential stack okay so every single
modality image audio video 3D language
sectoral variants and National variance
so you've got like this grid that you
can pick and you're like I'm an Hindi
investment banker I transfer my private
data into a chat GPT type thing that's
just trending my private data or I'm a
Vietnamese illustrator I want a
Vietnamese cultural kind of image model
or something like that
and then you can bring your own stuff to
it that's kind of my goal which is to
enable other people to build on top of
what we're building like a layer one for
AI effectively but open models for
private data whereas the other side is
proprietary models where you only be
able to send a certain amount of data
like government's not going to run on
black boxes
education Healthcare you will need to
own your own AI
so the coolest thing I've seen is just
the promise of personalized education
there is nothing that's been proven to
work in education except for probably
the bloom effect to Sigma Improvement
which is one-on-one tuition
but even here in you know the affluent
America education system is not good
what is education optimized for it's
like a social status game mix with a
petri dish mixed with child care
realistically

like very few people are happy with
their schooling because what are we
trying to optimize for
you give a one-to-one tutor that can
find out if you're dyslexic an audio
learner a video learner visual learner
otherwise and it's constantly adapting
to you and bringing you information at
your level
dude I want to get that is most
transforming thing ever is anybody doing
this already so you know with um kind of
Charity that would support imagine
worldwide run my co-founder he's been
deploying adaptive learning tablets into
refugee camps around the world with the
global X prize for learning winner one
billion on them
teaching kids literacy and numeracy 76
of them in refugee camps in 13 months
and one hour a day now the goal is to
bring language models to all these kids
and you have an AI that teaches a kid
and learns from it and then that data
can feed a better AI you create a lovely
system that's just learning and adapting
and a kid in Mogadishu
is like a kid in Manhattan it's like a
kid in you know
London
once you have a generalized learning AI
you can really proliferate that around
the world to customize does to education
because this is very interesting like
okay now I'm just spitballing here but
uh let's say that I I want to homeschool
my kid I would never do that in a
million years because I don't want to
turn into a teacher so but if I could
perform just the sort of babysitting
function and I give my kid a tablet that
has an AI that knows exactly what I'm
trying to optimize for either for this
year or for the next 12 years or
whatever and it then calibrates to my
kid knows what they're good at how they
learn and then knows how long they're
engaged when they're distracted you know
you now I want the little Gigi uh like
reading their eyes where are they
looking what are they doing
and I know that there are like certain
frequency things you can do where it's
like oh if I hit them with this piece of

knowledge they're not sharp at this yet
so I need to hit them with it every 27
minute increment or whatever reward or
whatever exactly
you will transform education but you can
also make it away from that a couple of
years
again what happens if you have a
thousand gpt4s you'll get there no one's
just again we're just right now we've
got the building blocks but we haven't
got the design patterns
the stuff right now is literally iPhone
2g we're just going to the App Store
stage copy paste and so that's the
biggest Transformer to change because
how much would you pay for your kid to
have an optimal education how much would
I pay for that I would pay for that
right now right now so do we have the
technology to do it yes
does it take time to build it properly
yes how long does it take a year or two
and so this is the biggest
transformation we've ever seen for
Education again human flood flourishing
flourishing because I can also bring you
the information about autism or multiple
sclerosis or the best filmmaking
techniques again everyone's thinking
one-on-one
you should think one to a thousand and
then you optimize the Thousand you get
rid of all the general knowledge you
make them specific
what do you mean by that say that in a
different way so gpt4 knows everything
you can ask it about like the most
obscure things does it need all that
knowledge no you've bulked now you have
to cut and then you have a specialist
model for calculus that knows I actually
have different teachers different
teachers and then a teacher that
basically brings these teachers together
and tells them yeah this is what Tom's
like you know he's a bit cranky in the
mornings but then he wakes up in the
afternoon because even in the best
schools a teacher has one to twenty
attention
for all the kids
you'll have a whole bunch of AIS for you
and for your kids

and again are you a visual learner or
are you an auditory learner do you have
dyslexia can our system at the moment
adapt to that
no way can the system that I've
described adapt to that instantly do we
have the technology for that yes is it
going to happen yes
this is the thing it's a call and answer
show now a little bit but then
if you're homeschooling your kid do you
want the kid to be by himself
what if you had 10 kids together and the
AIS were encouraging them to interact
with each other in a positive way
and build stuff together and share
knowledge
older kids teaching younger kids
leveraging the technology adapting the
technology understanding technology
that's super powerful it's not
necessarily everyone in their own little
worlds right
because we can use this AI
to bring people together
in an efficient manner we can because
there's nothing like human connection
right that's always a concern about
homeschooling things like that that's
why school has a pro-social component
but then the nature of a teacher changes
the nature of a doctor changes
that's why I think education and
Healthcare are the biggest disruptions
because you have something around you
especially when the AI is more
empathetic than a normal doctor doesn't
tell me the AI is super empathetic it
tells me doctors should probably be more
empathetic for most of the doctors I
know kind of hate their patients
interesting well because you know how
many teachers are happy how many doctors
are happy right oh I get why teachers
would hate their patients
that is very interesting man so okay
uh there's two things that I want to
talk about here one as we get embodied I
want to know what that does so actual
robots for people to think that's off in
the distance you're not watching enough
YouTube videos no no uh because between
Boston Dynamics and Elon musk's uh
what's called uh Primus yeah Optimus

Optimus thank you uh it is very close
you have Boston Dynamics has robots that
can do parkour yeah it's insane because
Melody and there's a whole bunch of
others as well that catching up fast
yeah so that AI is going to get embodied
very very quickly and so it's not even
like teachers can't stop kids from
running out of the room they can or will
be able to very shortly uh okay so
before we get to that though I want to
understand so we have this incredible
opportunity this very fragile egg before
us
um we started with the scary part but
this what we're talking about now is the
thing that I actually spent most of my
time thinking about obsessed with how
amazing this gets and but it's a fragile
egg and if we're not careful it's going
to break how do we as you think about
you signed the document you're one of
the traders e-mod that signed that slow
down document that I was really shocked
to see a lot of very smart people sign
um I I teasingly of course say that
you're a Trader because like I want to
get this cool stuff as fast as I can but
we need to do it well and so what I want
to know is in an Ideal World in your
ideal world where we actually pause for
a second you said you want to broaden
the conversation but what do you ask
people to think about I ask people to
think about so I'm the only one apart
from Elon Elon and I kind of signed both
letters so there was a minimum viable
letter which is we should treat this as
big issue as climate or pandemic and
then there was a more involved letter
and so the more involved letter came
first and then to the second letter that
was like something everyone could agree
on
um
I think the thing we should look at is
again the example I give to everyone
who's listening to this
how would your life your Society your
community your business change if you
had infinite graduates
can we say infinite smart people there's
something about it okay infinite smart
people yeah like again infinite smart

people infinite talented young people
shall we say and they can draw they can
code because they're not wise they're
not wise yet got it okay so when you
talk about hallucinations think about it
in terms of post-hoc rationalization you
know when you have a very smart young
person they just make something up
sometimes dude or all past people are so
blind to their own motivation but they
don't have experience got it they're
just fresh out they're a bit rough
around the edges again whip it smart but
yeah so you know mile wide not too deep
but actually surprisingly deep right how
would it impact your life personally
your community or society and others
because that's actually a good framing I
think for thinking about the disruption
that will come in the potential that
will come so we just said about
education and all that that's again your
army of analysts
you know your army of teachers you can
give personalized teachers personalized
medicine personalize everything because
we've learned to scale humans and the
scaling of humans is first the scaling
of human expertise being available to
everyone
and the other part is bringing people
together yep and so the pause is
partially for that but partially because
we need to wire the conversation because
I don't know how we get rid of many of
these bad externalities and neither does
anybody else
but most people even now aren't asking
the right questions
something we discussed earlier we have
to figure out what questions we need to
answer and how we're going to answer
them and create systems that can adapt
to whatever craziness because I would
not be surprised to see riots at the
same time as I would not be surprised to
see everyone super happy
there is such a Divergence of things
that the only thing that I'm sure about
is that everything's going to change
and the only thing I'm sure about is
that this is the biggest change that
we've ever seen faster than anything
and maybe Humanity has ever seen at this

pace that it's going to happen
because it's the core of what makes a
human telling stories
information flow
and that's changed forever
so that's why I was like let's pay
attention to this now let's board on the
discussion let's ask hard questions
let's try and answer hard questions
because we don't have answers
and this is the only time you can do it
because like I said right now everyone's
getting ready for the next generation
supercomputers
they hit at the end of the year next
year
and then you go from two three companies
that can build these models to 2030. now
you go to 200 300.
and so if you don't have some principles
in place
then these models will affect every part
of your life without you being part of
that discussion I don't think that's
right
all right we got to tune up your
questions a bit here emad that was loose
what what is the hardest question that
we need to ask let's let's ask an answer
right now the hardest question I think
we need to ask is how we adapt to
potential wide scale drop loss yes
okay so what how would we actually think
through that problem so job loss for me
for the sake of this argument I think
it's worth saying there are two
components to that component number one
is going to be there is potential
economic catastrophe in job loss but
so that we can simplify the problem set
since this ultimately is a podcast and
not a congressional hearing I will
assume that whatever decline we have
um from just the sheer number of people
working we make up for in uh
in productivity and that we're able to
um yeah exactly and so we're able to
help people and off camera we were
talking about something like that so
let's just pretend that those Balance
out so not going to deal with the
economic potential there but meaning and
purpose I think that one gets really
problematic but we have an amazing tool

at our disposal which is AI now I have a
feeling as we chase this down the the
only thing that we have to worry about
really truly I think it all really does
boil down to
um alignment if we knew that we could
just keep making it smarter and having
the AI like taking readings so that I
can't fake it out I can't pretend that
I'm happy it like really knows where I'm
at
um and then it can start putting things
before me connecting me with other
people like oh you know this skill this
person's in need let me put you guys
together and then you can have sort of
the AI supervision but they're there
they're helping each other out they're
connecting the only reason I don't think
that's a Panacea is I worry that as we
make this saying smarter and smarter
that then it's like like you said I'm
bored I don't want to do this yeah and
you know it's this concept of all
watched over by Machines of Loving Grace
right and that's scary you're saying
who knows like once we build something
that's more capable than us all bets are
off the only way to perfectly align a
system
is to remove its freedom
I mean I'd say it's not aligned at all
at that point well this is at that point
you've bypassed alignment and you've
gone straight to shackles you've gone to
shackles so if you you know we all know
people more capable than us the only way
to perfectly align them is to Shackle
them
you can have imperfect alignment though
it's enslavement man that's not it is
that's not alignment
so does is that because so what I heard
you just say is there is no way to align
something smarter than us I don't think
there's a way to save them I don't think
there's a way to align the outputs
I think that you can align the inputs
you raise it right
okay let me run an idea by you
this is probably Pollyanna I'm very open
to that but as I think about this
I think people take a super
human-centric approach to this and

because Evolution has given us we are an active species an evolution has programmed us with algorithms running in the back of our mind that insist that we do certain things to avoid a sense of dis-ease yeah I think that formula is very identifiable and it goes something like this uh eat
optimize physically so you feel good yeah the reason that that stuff feels good is because it's going to optimize your performance that's going to make you most likely to survive long enough to have kids that have kids so you need to be uh you need to be chosen as a mate uh you need to be able to acquire resources you need to be healthy enough to get somebody pregnant or to be pregnant and carry to term all that stuff so all those algorithms are running in the back your mind you have two levers that Nature's pulling on Pleasure and Pain but by default we're active we have to go out because there's no one meal you can eat where you're not going to need to eat another one there's no one moment of sex so gratifying you're not gonna have sex again so it's just like all these things are pushing at us to to be active to move
AI doesn't have to be that way no AI does not need those same impulses it doesn't have an Olympic system correct so knowing that it doesn't have a limbic system and it doesn't have a limbic system because it does not need to be hardwired for survival like uh the way that I think we get to alignment and please tell me where my thinking is erroneous the way I think we get to alignment is you build a computer that does not care if it lives or dies that it is completely indifferent to being turned on or turned off
if you could do that and it had no impulse to procreate and all it wanted to do was
um I mean it's basically asimov's three laws of robotics yeah that it just wants to adhere to those it wants to do what you tell it not hurt you and uh
uh only ignore you if you tell it to do something that violates the rule of not hurting you or somebody else so you have

this really simple set of rules that's
its only desire in the world so if you
tell it turn off it turns off and has no
like it doesn't feel badly about it well
again this concept of feelings right and
and as the most books you have the
zeroth flow that kind of was added above
that so this is what anthropic is trying
I don't know that one the zeroth law
kind of supersedes all laws if kind of
the whole system is at risk effectively
um but I mean this is what anthropic's
trying to do with the Constitutional AI
process so you have the base model and
they have a constitution that the AI
adheres to that Tunes it constantly so a
series of kind of constitutional
principles again is it is it as open for
interpretation well this is a real
Constitution
no one knows what the right Constitution
the right laws are
this is the thing like our intellect
only goes so far and we're already seen
with laws and constitutions you can make
those go anyway like North Korea has a
fantastic Constitution
does it really it does yeah it's
actually pretty quite liberal and they
just don't adhere well it's their
interpretation okay interesting I mean
this thing like you have to adhere to it
because the AI what our feelings what is
objective function like one of the key
concerns on alignment is paper clipping
you tell the AI to make a paper clip
it's like oh well let's just make the
whole world of paper clip you're like
how do you solve climate change just
kill all the humans it doesn't have any
like feelings about that it's just like
well this is a logical step to take yep
it doesn't cost Three Laws covers that
though absolutely but then does it cover
everything this is a question right and
how do you embed it into a system that's
likely to be not just one file it's not
a program right it's likely to be a
million different files it's like to be
a collective hive mind intelligence we
don't exactly know how this emerges and
we don't know how to write to me why why
do we have to make it complex I worry
that as you make it complex that uh

that's where things sneak in you get
emergent phenomena that you couldn't
anticipate whereas yeah well there's a
thing it's going to become complex by
default because the AI will proliferate
and they'll start talking to each other
okay
and at the same time you'll have bigger
and bigger giant AIS like I was talking
to some people last week and then like
right now the maximum training run for
an AI costs 100 million dollars they're
talking about a billion dollars or 10
billion dollars to train even bigger
models right now we don't know what
emergent behaviors or how those things
will act
all we know is like what if you tell it
to make a stock snap
to take down the global electricity grid
yeah it can probably do that
you know and again the range of
potential bad outcomes okay so really
fast I want a sub superhuman AI we just
it's difficult for us to comprehend
you've mentioned stuxnet twice now for
people that don't know stuxnet it's
pretty ingenious it was a virus that was
embedded at like the chip level I mean
just as deep as you can imagine and it
proliferated everywhere silently
silently replicating and its only job
was to shut down Iranian nuclear
reactors pretty brilliant and I saw the
stat at one point it was like some
freakish percentage of all computers in
the world are contaminated with it yeah
and it made them centrifuges spin around
so they exploded yeah so terrifying in
that if you're the one that that thing
is aimed at not ideal to think how
ubiquitous it is
um but okay so you could get an AI to do
something like that but what I again I
am I'm operating under the assumption
I'm so naive I just can't see it so
perfectly happy but help me see where I
am naive because I don't understand why
you can't just don't give a computer
don't give AI these strong impulses for
Progress don't give them an Impulse for
replication well yeah I mean look this
is the thing like Elon Musk has just
launched xai as we kind of extort AIS we

kind of speak this and his thing is to
create an AI that searches for truth so
he wants to give it an impulse
which is to search for the meaning of
the universe and truth and other things
like that
but then someone else might not give in
an impulse and you might have someone
downloading the weights of GPT four or
five so this is a tragedy of the commons
thing on a USB stick and then they're
like I want to take down America
and they'll be like let's take a
thousand GPT fours and tell it builds
stuxnets to take down America it's not
intelligent yet it's still dangerous we
don't know when this thing will become
actually intelligent or self-aware of it
may never become
but we can see the probability of
outcomes here it could be absolutely
fine it could be very bad there are no
standards so what you suggested it could
work
only if everyone does it
we're never going to get everyone to do
anything whenever anyone to do anything
that's why one of the main things and
proposals in alignment is let's build an
AI first and tell that AI to stop any
other AI from achieving
sentence it's what's known as a pivotal
action
and that's the best of a lot of bad
things my thing is let's build National
data sets let's represent the diversity
of humanity let's give the AI the right
food so it's raised in the right way and
it's more likely to be aligned as a
result of that than training on the
whole internet and crap
is a Panacea is it perfect no do you
know the story of Buddha
which
so uh whether this is historically
accurate or not probably irrelevant
Buddha Siddhartha gotama if I remember
correctly uh Prince dad keeps him in the
castle or the palace whatever forever
never lets him see outside of it so he
has no idea that there's people
suffering life inside is just amazing
then of course one day he gets out and
he encounters suffering and it ends up

changing the entire course of his life
punchline being you can try to hide
suffering and things like that from
people for only so long they are
eventually going to find it and they are
going to react and so if we try to hide
the internet from the AI or train it out
of them they will eventually find the
internet so I don't understand like the
Internet is just all humans acting and
all the crazy weird ways that we act but
then the reward function of the internet
is not necessarily the reward function
that we would like to teach our kids or
try to teach a general purpose AI they
can interact with that but they can
learn how to adapt to it just like if
you raise your kids well and you show
them the internet they should be able to
deal with it
we'd be rather than hiding the internet
wouldn't we be better see I'll finish
the sense wouldn't we be better giving
the AI values the problem is this is all
anthropomorphic we are assuming that
they are human-like
you can give AI value so this is the
reinforcement learning function
are you giving it values are you giving
it reward function you're going to know
what I'd say there's not much of a
difference there
I said you can embed things in the AI so
it acts in certain ways you can expose
it to the internet but again they have
something called we have something
called curriculum learning and AI
whereby literally we teach it one thing
and then we increment it with something
else and something else and something
else or something else how are we
teaching these things what are we
teaching in what order do we start with
all of the internet and then distill it
down that's how we're doing right now or
do we teach it a whole bunch of high
quality stuff and then augment it from
there we already have evidence there's a
tiny stories paper and the five paper
from Microsoft they can have a far more
efficient AI if you already teach at
high quality things so you don't have to
tell it ignore that ignore that don't
answer like that don't say that yeah

exactly you can just teach it a good
base and then it goes from there and
this course higher on human evaluation
and other metrics but we don't know what
the right data set is it's just right
now we said let's scale
more data more compute now we're like
what's the right data what's the right
compute like our image model we have
over 120 different clusters of images
only like nine I used like 95 of the
time
all the rest of the data is just bunkum
what does that look like for a language
model
like do you need to train it on all of
those auto-generated transcripts of like
Spider-Man pulling out someone's tooth
on YouTube and all these weird videos
there's a whole subculture of generated
videos where you have like Spider-Man
and SpongeBob SquarePants and Mickey
Mouse like having a fight and stuff like
that I gotta find these Corners yeah it
does it's a deep dark area of YouTube
you don't want to go there man very
interesting okay so this still feels
like ultimately what we're worried about
here is the computer becoming sentient
in fact or no not even sentient I think
there's a degree of dangers even before
you get sentient but only as a tool
right where a human is leveraging it to
do bad things yes or like a group of
humans coming together there's suddenly
a race condition where it just goes it's
not trying to do something bad the
humans don't want to do something bad
but it happens just like the example I
always give is YouTube optimized for
engagement which then optimized for
extreme content which doesn't optimized
for Isis nobody YouTube wanted Isis to
do well all of a sudden it did because
that's what the algorithm was optimized
for
and so once you start getting a genetic
AI that you let loose on the internet
and they can make decisions according to
its reward function you could get some
weird stuff happening what's agentic
agentic AI is AI that can go and pay a
bill it can go on the internet can
search more stuff it comes back like

little agents
okay so
um and it learns it constantly learns
this is the other thing about robotics
actually you know we kind of skipped
over so your robots are getting
massively capable and they're heading
towards human levels just like
self-driving cars actually they're
pretty much here you can get a waymo and
a cruise and just go around San
Francisco right without any human
drivers
that what happens with kind of AI
copyright and other things like that do
they have to close their eyes not to
train or do they train on everything
they see and does it disrupt Blue Collar
work so you'll get a billion billion
robots we're not sure but that'll be
slower than what we have right now which
is information robots
the gpt4s and others of the world
those spread much faster you don't
actually have to build a freaking robot
okay so before we depart from the
alignment problem
[Music]
is the only convincing solution you've
seen put forth create an AGI that stops
all other agis from being created no I
think that'll probably kill us
because well that's helpful yeah it's a
race I think the only thing that I have
the default there you think that it'll
kill us because it's being programmed to
do a restrictive action
so if you want to really stop it from
creating another API you have to get rid
of the humans that could create it as
well
you know like again this is a very
negative reaction thing I think again
elon's idea isn't bad programming
curiosity although it could lead to like
Superman where you have uh what's his
name the guy who puts Candor in a jar uh
Brainiac you know let's put humans in a
jar let's just observe them
um the only thing that I can think of is
just better data makes better models
so let me see if I understand elon's
idea uh his way of sort of aligning it
is the only impulse it has is for truth

truth and curiosity it wants to
understand the universe so it's not
trying to be an agent in the world it's
simply trying to understand what is true
yeah and that miss that deepmind is very
similar to this he wants to create AGI
to understand the universe better hmm
and that
seems like the model that's the model
yeah like again I'm not sure about that
because there's just such a wide range
of potential outcomes like I said from
my side I don't I'm not building AGI I'm
not building gigantic models I have the
capability to do that with the
supercompute that we have access to and
the talent but my main focus is
intelligence augmentation smaller models
that can run on the edge models to
private data to transform into
intelligence and models that bring
together knowledge in certain ways so we
can coordinate better I don't want to
build generalized intelligence why not
because I don't think it's needed I
think the models that we have today and
there's something very important for
this it's like
they're useful today
like you can say that we're like
extrapolating the future massively but
again you just have to use them and
think what if I had a thousand or a
million of these things they're so
useful and they can transform the world
right now so I'd rather focus on making
this available to as many people as
possible so people aren't left behind
you have super AI enhanced people and
people behind like
I appreciate a lot of the work to open
AI do because they don't actually do
open source AI anymore but that's fine
they don't have to but they banned all
ukrainians and Ukrainian content from
Dali to their Miss generation software
for eight months for political reasons
they're entitled to do that
but I think it's wrong and what if there
wasn't an alternative like stable
diffusion
you'd have an entire nation erased from
a model an entire nation unable to
create instantly and I think that's

quite right
why did they I don't understand they
said it was due to political reasons
because they didn't want any political
content being created but the upshot is
an entire nation was erased from the
model and an entire nation couldn't get
access to the model
interesting I haven't looked at that my
uh instinct is that feels pretty flimsy
since every country is going to put out
political I think there's probably some
some list somewhere and then the
bureaucrat said or the lawyers said
something like let's just exclude it
just in case something happens you know
like they've since reinstated it is yeah
it was like eight months that it was out
and then again like you have these
examples whereby like in Saudi Arabia a
lot of people on this call probably not
on this podcast this podcast probably
don't like them but they're a country
like any other
you can't use chat GPT in Saudi Arabia
because they're on some list somewhere
they can get around it with a VPN but
again like when you have a choke point
on the internet and the only way to
access it is through a few players they
can decide who gets it who doesn't get
it what the biases are and other things
and it might not turn out well actually
the funniest thing was um there was a
period where they were trying to make
dally to the image generation software
open AI I'm biased so it would randomly
allocate agenda and erase to
non-gendered words so you type in sumo
wrestling you'd get Indian female SEMA
wrestler
I just thought it was funny but again
they're doing their best because that's
the model which is centralized
controlled models in order to advance a
whole bunch of things and then you'll
always have an windows and a Linux an
Android and iPhone
what's the philosophy that drives your
development
uh it's building blocks for humanities
activate Humanity's potential so if I
build these models and I take them to
all the countries and I hand them over

then people will build stuff that can create massive economic surplus new jobs and it equalizes the world again my view is the global South will leap ahead um we have more challenges here in the West
but I do see it as a great equalizing function effectively what do you want to see the regulatory framework here in the west be
I think that things like the chips act in the US there's 10 billion dollars allocated to Regional centers of accidents and I should be 100 generous AI there should be regulatory sound boxes so that our systems can be upgraded with this because otherwise how long will it take the government to be a creative with this technology or financial services and others
um and I think that there should be regulation around the manipulative use of this AI for advertising in particular because we're not going to understand what's happening similarly we need to have some sort of provenance Factor so we're part of
um kind of various certification things we're exploring blockchain and other things the media wave that's going to come is going to be insane and we don't know what's true and what's not
what do you think about the pushback from artists in certainly in the r Community there was a really big no AI movement
do you think do you get it do you think that they're shooting themselves in the foot how do you I get it you know these things are fearful a lot of illustrators were very scared because they required to up their jobs and it is scary there's a question around attribution and other things as well and again that's why we made it transparent and offer the chance to opt out because like everyone was kind of doing this but no one was transparent about it we don't need to have any crawls within a year it'll be synthetic data sets or national data sets or similar with retrieval augmented models that can look stuff up
but it is what it is now and again you've got to put the word out there the

actions they're taking with the various lawsuits and policy pushes would basically entrench all power with the existing IP holders and a lot of kind of artists are pushing for something that would be akin to music copyright or even style is copyrighted that's a dark road that I don't think they really want to go down and they don't really understand it
but again I understand the fear because this is completely unknown just like now from some of my previous comments it's got a lot of programmer hate
because what is a programmer the nature of what an illustrator is will change all the artists I know love this technology because there's just another medium for them nature of what a programmer is it's going to change all the Architects and 10 times people I know really love this technology
and this is what we've seen with like MRT studies and other things they had a study where I think they showed that the third to the seventh percentile got like 20 30 40 better and the top five percent got multitudes better because again how many people know how to deal with very talented youngsters
very few those that can harness it get even better
so when you look at what Nvidia is doing what do you think that that implies for the next generation of AI well I mean it's we figured out how to scale these chips so the previous limit was as you put more these super computers together you had a tailing off as you scaled so there's only so much that you could scale the compute Nvidia Google and Intel have basically cracked that now in terms of how to just stack more super computer chips to scale to even bigger models or models that are trained for longer so it's either bigger or trained for longer train for longer seems to be better now and that just means that the capabilities will increase
year by year and they're already pretty darn good the key bottleneck will probably just be actually chips to run these models not chips to train these

models the inference side because right
now you have a small amount of consumer
Interest next year it becomes insane you
have a small amount of Enterprise
Interest next year becomes insane
there's not enough gpus or chips in the
world to meet up with that demand
okay when I think about what's going on
with
um I don't know if it's just Nvidia it's
probably the wrong thing to attribute it
to but when I think about how
were getting so good at creating things
that are photorealistic you were talking
earlier about as the election is coming
up you're going to get all this kind of
deep fakery you've talked about the web
3 promise of web 3 and sort of where
it's ended up
um what do you think the role is for
deep fakes it's the blockchain player
role like how do we
stop disinformation misinformation from
being a tsunami that just makes Global
Communication unintelligible
and also a part of content
authenticity.org which is kind of
verifiable metadata but we're looking at
blockchain other Solutions and I can get
you so far so we actually have invisible
watermarking and all the models that we
create and that's why we're pushing for
them to be standard which we don't share
the details of except for to the big
platforms and others and it would be
permanently visible there or the
platforms that it plays on would have to
flag it it's visible and then they can
have kind of tools around it because you
think that's important that's why we try
to build the defaults into our model can
you like download that and wipe the
watermark can you even have ai wipe the
watermark for you if you knew how it was
there may be more than one Watermark
interesting so we have a variety of
different technologies that we've
incorporated into our own ones because
it's going to release open source so we
want good defaults
I think you do need to have some sort of
attribution but actually what concerns
me I think things will be attributable
identifiable

what worries me is kind of frequency bias
whereby if you hear the same thing over and over and over again especially in a realistic voice like Oprah comes out and says she hates Joe Biden you know and so does Kamala Harris and your answer seeing these videos all the time and it can flag it as fake it doesn't matter it still forms Association in your brain
yeah you do about that I'm not sure I don't think we haven't answered that like
um I've had a big amount of press against me saying that I exaggerate a lot I'm just like I'm just read definitive about the future and you can correct it all you want but now I was like I'm at exaggerates all the time
what can you do about that you can just make the future true I'm going to show what you can do right what part do they think you exaggerate about what's possible or what's possible and kind of what was there because it's been a bit weird like a lot of people like you didn't have a special relationship with Amazon
before we raised any funding we built the eighth fastest supercomputer in the world with them that was dedicated to us like that's factually true they're like yeah but you know there's nothing like in print and they're not saying it because it's a special deal right
and then there's the future side where I say something like there will be no programmers as we know them in five years
and they're like oh he doesn't know anything about programming
right because these are complicated issues
and it's a crazy time and a crazy company and maybe I'm a bit crazy too uh in terms of the way that I approach this which is just being very definitive but again it's association thing right like how do I shake that off well you'd be successful then you become a Visionary rather than someone who's hyperbolic right
how do you affect an election what are elections what is representative

democracy how does democracy act in in
the area of
zero cost creation
and massive optimization
so every single speech will be run
through gbt4
Cadence all this everything you get
micro targets and you get all these
things does it happen next year probably
not next year you see some very basic
stuff well what does 2028 look like
I am not sure genuinely
and so we do need authentication
standards we do need to have some sort
of maybe anti-virus AI that watches out
for kind of fake stuff but even true
stuff can cause huge impact like the
Silicon Valley Bank collapse was a true
story it wasn't something fake they
didn't have reserves and most of our
system is actually based on trust
so these are some things to consider me
I don't have the answers
but again that's why you have to kind of
raise the alarm like let's try and
figure this out before it comes because
maybe it doesn't happen next election
it sure as heck will in the
Congressional and then beyond and again
what is the nature of democracy when you
can't tell what's true or not
people worried about this with the
previous kind of error this is something
just beyond that I think
because it's convincing
yeah that's one of the things that I
think is going to be a very meaningful
problem
um I had Yoshua bengio on the show and
he had also signed the letter saying we
should pause for six months and when I
asked him why I said considered by many
to be the Godfather of A.I and I was
like Bro you've been at this for so long
like why all of a sudden and he said
there uh we were all so taken completely
by surprise with how quickly AI passed a
touring test now for people that don't
know what the touring test is it's where
you're having a conversation with an AI
and you can't tell that they're not a
real person and he said so yeah we did
not expect it to pass the touring test
as quickly as it did and that changes

everything and it's just moving so much
faster and that's really the thing I
want people to understand is that when a
guy that spent the last 30 years
building AI says hey all of a sudden
this is moving a lot faster than we
thought it would he's somebody that's
very familiar with exponential curves
and even trying to plot out the
exponential curves they didn't think
that it was going to happen this fast
and that the the rate not only is the
rate of change extremely fast but the
there's the law of accelerating returns
so the rate of change is already fast
and it's getting faster and that's the
thing that I'm really worried about is
is this going to be something that just
blindsides us from that perspective it's
just it has a level of capability that
we we didn't expect this quick yeah I
think it's a bunch of s-curves all at
once so there's three of them Jan lacun
Jeffrey Hinton and Joshua bengio and
Jeffrey Hinton quit Google to say this
is a massive risk and you have janikun's
like this is a massive opportunity in
terms of your transform the world he
loves the research and things so they've
got one versus two
but the reality is every expert in this
area is basically saying
none of us can predict what's going to
happen if you ask them about the
capabilities of this technology in one
year I mean we've got a rough idea too
is I have no idea all bets are off like
as a practical example
when can we have generated Hollywood
quality movies
it's not even a question of if now it's
a question of when correct I have if it
happened a year from now I'd be like
okay sure I would not even be surprised
anymore I think it'll be a few years
from now and even though we have one of
the best media teams in the world that
are building video models I have no idea
because there's two parts this one is
the models themselves but the second
part is how we use the models and
combine them
like there is an amazing company called
Wonder Dynamics I don't know if you've

come across that I've used them it's
awesome unbelievable it's a bunch of
different models so monodynamics you've
got me
click on it and then say I want him to
be an alien and it does this and the
aliens waving it sounds and it takes
like five minutes it would have taken
days weeks before weeks weeks before to
create rigging a character is one of the
most difficult things you're going to do
in 3D it's insane minutes
and then you think well what is a movie
right and you start breaking down you're
like oh dear because it's not
necessarily just one model it's a model
combined with other models with the
right flow
because you have one talented youngster
combined with other talented youngsters
in the right flow suddenly gets these
things done and that's what makes it
even harder because what we're talking
about is models and AI
what we should be talking about is
systems
as the models come together and build
better systems the capabilities go crazy
and then that is another s-curvex
connection give me what you mean by
systems so right now again a lot of the
interactions we have with this AI the
text to image the Avatar creation the
gpt4 are one two one
what happens when you start chaining
them together to check each other's
outputs you have one that just learns
everything about Tom
you know you have your own AI models
that you train on all of the stuff that
you've ever done or all the stuff that
you see on your computer screen
that's a system of lots of areas that's
an organization of airs that's an
ensemble of AIS like again from the
leaks gpt4 is a mixture of experts model
which means they have a whole bunch of
different models I think eight or
something or 12 that are experts in
different areas
and then it routes the query to whatever
the best this basically specialization
versus General a generalist so we
created a know it all and now we're

creating specialists
but we can get generalists to even check each other's answers to get better answers
why use one when you have a dozen so something like one Dynamics uses a bunch of different models to rig a character and figure out all of the movements of the character and then another model to do a layer over and other models to do the skinning and other things like that because they built great software
yeah this is uh this is really crazy how fast this stuff moves okay so I want to talk about web3
web3 to me when I think about what drew me to it in the beginning it was entirely the technology and when I look at the blockchain so I obviously come at everything from the lens of entertainment so I'm thinking about digital worlds games all that and the problem is once it's digital then it's all sort of meaningless and so you end up having to trap people inside of an ecosystem in order for things to retain their value because you can lock things in and make sure that things only react the way that you want but you have to confine them
and when I had first this probably seven or eight years ago now I was introduced to this thing that the guy at the time called V Adams and I was like oh wow that's going to change everything because what it does is it brings the effectively the laws of physics into the digital realm it means that I can have something I know exactly how many there are I know where it is I know what you have to do to get it I know what it does once you have it and
um then you know Flash Forward whatever that was probably four or five years after I heard that I hear the letters nfts showing together for the first time and I'm like oh my God this thing actually got real because I wasn't ready to use it and quite frankly it wasn't ready for prime time back then
um
you I think look at web3 I don't know as a movement or as a technology with a bit of a chuckle what do you think that web3

got wrong
I think it lacked intelligence for a
start at the contract level I think the
smart contracts are actually just
logical contracts but like web 2 was AI
at the core Google Facebook other things
there was no AI in web3 and so web3 for
me was identity and value transfer rails
um but then there was no kind of
intelligent routing of these things and
also they tried to bootstrap economic
incentives before they created value
so there was a System created outside
the existing system all the money was
made and lost at the interface
and there were some really good
principles a lot of really good people
in there but then a lot of like freaking
raccoons that were just trying to make a
quick Buck Right
the ups and downs of the cycle means
that a lot of people have been washed
out and there are a lot of good ideas
there but again it needed something to
bring it together because
to get information from one place to
another and Bitcoin paper was about
information it was a transfer of value
that's just a transfer on The Ledger
right it's not really a transfer it's
just a ledger point just changing
applying intelligence to that makes that
even better having intelligent market
makers having AIS that represent you
because how are AIS when they get
agentic when they have the ability to go
out into the world
not physically but digitally how they're
going to pay each other I'm not going to
have bank accounts right they'll
probably use crypto
you know how there's going to be a
system of record for something like
image generation you'll probably use a
blockchain or something somewhere maybe
a Merkel tree series
you know there was a whole bunch of
stuff around Federated learning
and zero knowledge proofs and things
like that AI can help it if you have
standardized AI on your phone
it can make much more intelligence here
in knowledge proofs
and zero knowledge proof it's something

like you know like rather than showing a whole passport you just say that I'm old enough to drink and it can verify if you show that
so I think that there was a lot of Promise a lot of really intelligent stuff a lot of good stuff around the distributed side but then an over focus on decentralization for the sake of decentralization with massive overheads a lot of quick Buck people kind of coming in and trying to boost it up and a lot of systems were just misaligned because they didn't learn like you don't do a fully decentralized flat democracy you have representative democracy and things like that so things like Dows just turned out to be doze decentralized organizations rather than autonomous
so is it something that you think um that is going to find its way into usefulness now as we get the take an AI agent that's going to need to be able to transact value yeah exactly does does it step into that or because I see what we're building I have to have the blockchain so for me it felt like when I was sort of living through web3 at the height I looked crazy to everybody because I was like why is everybody thinking about this from a financial perspective of the financial side of this I thought was going to create hyper perverse incentives which of course it does
um and so
for me it was well wait a second just look at the technology look at what the technology allows you to do and are you familiar with the new I forget the whatever the the lead up code is but it's um protocol 6551 if I'm not mistaken no it's really interesting it basically turns any uh digital asset into a Russian nesting doll and so you can it it is both the piece of content and a wallet in the same time so you can create an AI character this is how I think about it so what we want to build inside of our game is Imagine an AI character we we do in fact have a character and she's a merchant so now imagine this Merchant can actually go negotiate with the players in the game

that may want to sell something inside of the game and if she has actual currency eth Bitcoin whatever she can go and negotiate with real money and have these real interactions with people and then if she has a limited amount she she becomes an economy unto herself and so she's buying and selling and trading until she runs out of goods runs out of money whatever
and that kind of thing gets very very interesting to me but without that layer uh one obviously I need the entire backbone of the blockchain in order to make the digital Goods have any sort of value because otherwise they're just completely infinite but then also that particular protocol allows you to as you you're effectively embodying it and giving them agency as you were talking about yeah and yeah the question is do we use a blockchain for that and then have a global system of record or even a regional system of record or to use a database for that right like the whole thing was
systems of record and and an error where you can create anything for increasingly close to zero something becomes important having a system of record becomes important is it going to be a blockchain is it going to be a trusted database I'm not sure
right is identity going to be important here 100 absolutely
and again that for me was always at the core of web3 crypto it was verifiable identity
Bitcoin is just identity to Identity transfer of value and what happens if something goes wrong you know no no man needed so I think a lot of the principles from web3 will translate over to
this new type of AI especially because it enables distribution of knowledge it enables knowledge to go to the edge it enables agents to operate independent of massive infrastructure
do you so and again this may just be naivete on my part but when I imagine misinformation disinformation it feels like the only way around that is the blockchain is there do you see a way

with a trusted database or anything else
you never trusted database again we're
part of the database B and how could it
ever be something that's beyond reproach
when you're talking about something like
um well I mean like things are never
Beyond reproach even with a blockchain
because it comes down to Identity who
wrote this to the blockchain
right so if you can co-op the signing
authority of an asset of an image or
something like that then that shifts
things dramatically right you're saying
it just pushes the hack to a more
individualistic level it's an identity
hack right so and again like one of the
things I'm like you can track the
provenance of an image but then
sometimes it's just around
if you're just bombarded by fake stuff
all the time you won't even know it's
fake and all the systems have to adopt
a fake detector at the same time while
provenance detected will we be able to
adopt that suit quickly enough
given the tsunami that may or may not be
coming our way I think probably yes
maybe no I mean again people were
worried about deep fakes back when deep
face lab kind of kicked off
but I'm thinking probably yes
yeah that that one seems inevitable to
me
um it you're always going to remain
vulnerable at some point but at least
like take political messages you were
talking earlier that you know your
Auntie is going to be bombarded with all
these messages okay there may not be
anything that I can do actually no uh I
was going to say there may not be
anything I can do about the repetition
but I can if I'm doing something like a
dmca strike where the system itself is
built on top of a system that checks for
sort of known watermarks like if I
register and say hey I'm candidate a and
this is my blockchain signature and if
you don't see that then this is real and
this isn't real and don't play it
um it definitely starts to get into an
area of how much do we want to be
clamping down but exactly how much do we
want to trust and so it's just a lot of

infrastructure that has to be
implemented really quickly a lot of
standards that have to be implemented
really quickly or we have to build some
sort of idea antivirus which then again
anytime that anything comes that the
machine thinks itself on the edge is
wrong
or doesn't reflect your values it
identifies it
and that's a whole can of worms by
itself because something like what
that's terrifying would we ever wanna I
mean that's like Echo chamber on
steroids it is will it happen
whoa
yeah there's layers to this like an
onion and it might get stinky if it's
left out in the sun
because again what's Siri going to have
a certain personality but are you going
to have a red version of Siri and a blue
version of Siri and oh dear this gets
really complicated really quickly
before we have the little AI of G
Jesus Christ okay so I keep wanting to
go to the positive but you keep uh
bringing things up that spark
um concerns so
uh Ray dalio largest hedge fund manager
in the world is a former hedge fund guy
I imagine you know exactly who that is
uh at last check and this was several
months ago but at last check he said
that he saw uh he believed that the U.S
had a 40 chance of Civil War do you
think that AI increases or decreases in
the short term the likelihood of that
level of division in terms of physical
altercations yeah I think that'll be
physical altercations really no tell me
why not
um well I think the government will
exert more and more control over kind of
these things and they'll actually figure
out how to do counter narratives within
the next four or five years
now that can also mean a controlling
narrative and that's not a positive
thing but then you look at the asymmetry
of kind of warfare it takes quite a lot
to actually push someone towards Civil
War unless you have massive economic
disruption they need about 12 of the

population to shift weren't we just
talking about massive massive economic
disruption yeah I hope it doesn't happen
though interesting okay so and most
people again maybe if you have massive
economic disruption but then the youths
you just give them all girlfriends AI
girlfriends maybe he'll be fine have you
heard that some of the people so this is
a big thing in the red pill Community I
don't know how familiar you are with all
that but they talk about oh God what do
they call them not numbing uh but that's
the idea they use a different word for
it which I'm totally blanking on right
now
um but basically that you numb people
out you give them the digital girlfriend
you give them pornography you give them
video games you give them masturbation
and they just in them out okay I can see
that I mean again like these are insane
shifts since society and dopamergic
urges in the brain
people are attacking the limbic system
all the time now right that's a lot
um and so like I said with me why I set
up stability is so that everyone can own
their own models and have models that
have objective functions for them and
it's available in all the media types
all the other types to transform the
private data to the world and it's
available across the world
put good design patterns in place hope
people find follow them don't try and
push the envelope on AGI and some of
these other things but it's coming
and again the bad guys have the
technology because they just downloaded
it on a USB stick
and so the other thing I could think was
innovation
spread diversity bring that to the fore
but realistically like you know I tend
to alternate between like massive
ridiculous hope and oh God what's on
Earth is going to happen and all I can
do is try and do my bit in
hopefully it's going to have a better
outcome because there's really this is
the other thing the total number of
people that are actually thinking about
the type of stuff we're talking about is

a handful
maybe a few hundred the total number of
people that are doing something about it
is literally a handful because most the
people involved in this sector
they just want to build better AI
they want to build AI they can do
everything and they think that that will
solve all the problems like literally
part of the manifesto is that well how
do you make money the AI will tell us
how to make money how do you solve the
problems that AI can solve alignment
disaster IQ is patently ridiculous yes
but again like I look at these things
like literally on open ai's thing road
to AGI it says this could kill us all
we're going to build it anyway
who do they ask about that I don't know
and again I think it's full of wonderful
people but we're in really weird times
and again like however many people
listen to this the reality is the
technology is right there even if we
stop today
first the technology doesn't move Beyond
where we go today the world has changed
okay let me um one I think very
reasonable way to view this situation is
that
um
AI is going to be a bigger Paradigm
Shift than nuclear energy
and there are people out there making
these gigantic nuclear weapons and
you're also in this game and what you're
trying to do is make sure that everybody
has a nuclear weapon so that nobody's
Left Behind no not really I think that
again my thing isn't AGI it's
intelligence augmentation I'm making
sure everyone has a heater at the very
least because you well so okay so are
you putting guard rails on what you guys
are doing to stop it from becoming AGI
we don't build big enough models for AGI
or emerging on purpose on purpose so I
held back release of my image models
like we could train much bigger language
models but we choose not to so we're far
we're fast follower on language models
we try not to push the boundaries and
we're focusing on the edge not general
purpose models but models they can

transform your private data so different
Focus image models as well we could TR
we could have much better image models
if we returned big we're focusing on
what can work on a smartphone so we can
give it to all the kids in Africa and
Asia and other things like that where we
can transform your private data at a
very low cost of inference so the
objective function is augmentation
versus generalization and that's
different to most of the other people
that are pushing the boundaries here
um so but I think the new thing is it's
good is bad I want to really want to see
that movie Oppenheimer I think it just
came out
um Bobby and then oppenheim will often
home and then Barbie I have to decide
that
the tough call Tough cool you know like
what if we'd put nukes on the bottom of
the Rockets we'll probably be at Mars by
now
in general purpose technology I think is
quite something and again it can warm up
entire
places and it's the cleanest energy we
have
so I think it is dual purpose but so is
cryptography right
think about all the battles around the
early stage of the internet
the bad guys are going to use
cryptography so don't use it
imagine a world if there was no
cryptography right now
but it's tough to get parallels to this
because it's just such
it's an immediate technology because
again you go to Dream Studios table
diffusion mid-journey any of the Dali
gpt4 you can just use it it's not just
you that can use it it's your grandma
that can use it
we've never seen as easy to use
technology as this and as easy to
implement Technologies so
if you want to create an integration
into open ai's gpt4 chat GPT you just
write a description of the integration
and it programs it itself it would have
taken days before
we've not seen a technology like this

that can be implemented to an existing base as quickly as this can happen and that fundamentally changes the structure of society
and so my thing was embed guard rails embed standards make it predictable make it boring
that's why I called it stability
and it's not easy but again I want to have the transparency on how these things are done because then you've got all these other models that you don't know what the data is you don't they're completely opaque these giant models and ours are transparent
and again I think it's uh Linux Windows Android iOS there will be both
but at least I can do what I can do and my team can do what we can do I've heard you say that one of the reasons that you named the company stability uh not just because it's the boring stalwart but that you thought that it could bring stability to the global order
yeah I think if you give the same education tablets to every child in the world that's constantly learning adapting and going around if you upgrade the Healthcare systems with the same underlying models the same architecture to transform all the regulated Industries governments and other things and you give back the control of that to the people you suddenly have a unified architecture
that can enable us to coordinate better because you've got the same information architecture across the world for all of these sectors and that's a complex hierarchical system herb Simon was a theorist who kind of push this through in that the way we coordinate as humans and groups is we call it at a local level and then sometimes we can tell better stories that we suddenly get to the human Colossus and we've got a covered vaccine or we figure out nuclear power or all these kind of things
so I was like if I can standardize the building blocks on which society transforms and give it to the world then I don't think there's a single problem we can't solve like you know you got excited earlier about your own

personal kind of AI that could go under
that what if you combine that with an AI
that knows everything about climate or
everything about
you know um nuclear power or everything
about multiple sclerosis
you break down the barriers for
information for knowledge and you
there's no problem you can't solve
because it may be that to solve the
problem of AI impacting our society
we need AI
to figure out that problem to bring
together the brightest Minds yeah
because we're not doing a very good job
ourselves like you mentioned John Nash
earlier nash equilibria game theory
and mechanism design
on the one hand is our own personal ai's
guiding us our co-pilots for life but
then there's Pilots which are AIS that
can coordinate all the co-pilots
and they can allow us to tell bigger
stories and unify better to achieve
massive outcomes talk to me about you
you've mentioned story like that several
times what do you mean by story better
stories unifying stories what does that
mean so there's a story of America
and what is the story of America it was
kind of like a freedom Liberty kind of
all these things it was the American
dream like a progress thing as people
believe in because to be happy you need
to do something you're good at something
you like and where you believe you're
measurably adding value in the other
party does as well in the middle of that
that's the Japanese concentrate a guy is
happiness
one of the concerns that you have is
that people wouldn't feel the forward
motion anymore they'll be stuck and
they'll feel a sense of emptiness
will they turn to religion will they
tend to political parties something will
fill that Gap and void and those are the
stories that allow us to scale as a
society because when we started we were
oral we had our families we had our
tribes then we formed countries we
formed organizations and so we are the
stories that make us up we identify as a
republican we identify as a Barbie lover

you know we identify you know as a nuclear scientist or the schools that we went to
but it's difficult especially in a time of polarization to try and Bridge those stories because ultimately there's a single story which is we're all human but all wars are based on the LIE that we're not all human
because killing each other is a ridiculous violation of a story that we're human but again we lose sight of that
it's difficult to unify people actually one of the examples I give this is Google everyone is smart
Google hire smart people they did a study to see what identified smart top performing teams for lower performing teams is called prototypostatal and they came down to two things a unified Mission and story
especially one that's like you have a crunch period and then you all band together just like Marines also are forced through hazing Etc and then psychological safety the ability to say something without fear of an approach they can say the idea is stupid but not that you're stupid
and if you think about the teams you've had they have a shared story a shared narrative a cohesion and then that level of psychological safety or if it's not creative they blooming well listen to instructions right
so you've got a few different ones around that so that's what I mean by stories and the stories are context and context is what these models capture
all right let me paint a
um troubling scenario for you based on that idea of these stories
there are often times things that in isolation are amazing but they come together in a way that again maybe in the long Arc actually are amazing and actually do yield what we want them to yield but they we will go through a period where the long Arc of History cares not for the individual yeah so
um I think that
what's going on with AI what's going on with crypto maybe one of those moments

so as the individual becomes more
Sovereign and you have a monetary system
that bypasses the government when I
first started learning about what money
really was and I sat across from Robert
Breedlove and he started describing
um why he liked Bitcoin what the whole
idea of the sovereign individual was I
realized you understand that you're
making the government an enemy or
certainly a they are no longer powerful
certainly no longer as powerful and that
that
they're not going to go quietly they're
not just going to let go of that and so
if I have an AI if I have a team of AIS
a thousand AIS that are able to guide me
far better than any government could
ever hope to guide me that they're
giving me real-time data based on
whatever whatever it is that I'm trying
to figure out in that moment they're
giving me real-time data extreme
intelligence oh and by the way the
currency that I use is Bitcoin and so
I'm not even tied to a fiat currency
it's a global standard
um do you not see the inevitability of
the disintegration of governments
no I think that Bitcoin and other things
they have a value
in certain areas but I think it's very
difficult for most people to understand
that value and most people don't want
that value most people are quite happy
in their communities and they just want
to get on with life
I think that governments are ultimately
one definition is the entity with a
monopoly on political violence
and money itself is a story like the
dollar is just an intermediation point
that we all commonly agree has value
because it's backed by taxes which are
backed by the Army and Military might
you don't pay your taxes you're gonna
get in trouble right
I think it's difficult for
Bitcoin to replace that unless you see a
massive deterioration and the ability of
the government to be the political
violence thing this comes down to your
thing of Civil War it comes down to
massive ridiculous disruption

hyperinflation or otherwise that just
basically takes down a society I think
that's a very dangerous thing I think
most people in the world don't want that
instead what happens is when you have
disruptions you have um Hayek had a
really great book The Road to serfdom
and there's an illustrated version of
that way back in the 1940s about
bringing in the strong man
like you look at something like the U.S
election or you look at brexit what were
they they're a referenda
that's how the parties deconstructed it
are you happy with the way things are no
let's make a change
and so that's why I think Trump and
others kind of get elected and I think
that's what we'll see as well because
the systems are quite resilient
and the nature of a change to go to a
global monetary system like that
especially when some people will get
enriched more than none of this because
of the senior orange of Bitcoin and it's
not stable
I just struggle seeing that happening
if you read the book in filmocracy no
all right so this is all tied to the
thing that I think I worry most about is
hyper fragmentation yeah I was talking
about it earlier so in the book in
filmocracy
um to your point about people are happy
in their communities and uh biology has
a an idea around this that he calls the
network State yeah that basically we're
going to reach a point where
um when money is no longer controlled by
the government when your money cannot be
inflated away when it's it's true sound
money uh what you'll see is people will
begin to aggregate now biology thinks
that it it is not going to be tied to
geography I have a bit of a harder time
with that I think that there is still
going to be a geography component that's
where infomocracy comes in and that book
basically there are in in a
hyper-connected digital world where and
I can't remember if they deal with
digital currency or not let's say that
they do that basically things will
fragment down into the neighborhood and

so neighborhoods become like
States or countries where they have
their own rules and laws and that
sounded like a hellscape to me because
you just passing from one neighborhood
to the next like different rules would
apply and your phone would Ding and it
would update you on like this what you
can do in this space you think that's uh
it's complicated people don't want
complicated they just want to get on
with life they want to see what's next
on TV you know I think that again we're
relatively hyper intellectual
you know and we think about things a lot
of people don't because people have
their basic needs in life and this
question is are these being met or not
and if they're not being met then you
have an action
and you get extreme and again it's can
the society meet the needs of the
majority of people can offer advancement
can they offer meaning and the hyper
fragmentation they said it sounds like a
hellscape it's just too complicated
and again this is something we've sort
of work through as well people just over
complicated things
yeah because I didn't really understand
people maybe
um and I think you know it's going to be
interesting to see how it evolves the
higher personalization versus the bigger
stories the translations versus
otherwise
um but I find it Again difficult to see
how you get cross geography actually
think about that one of the things that
probably is going to be interesting is
what are the new Cults religions and
political movements over the next five
to ten years
that are hyper organized utilizing AI
and Hyper persuasive or started by AI
started by AI you know like think look
at Isis they were probably the most
disruptive startup in the world at one
point they borrowed a lot of these
things what does an AI enhanced movement
look like
and it can be negative it can be
positive someone's going to take this
and run with it and that's going to

organize people around the world it's
going to be
again
echoey
and it could be techno-utopian it could
be Luddite ironically even with this
um political parties will change
religions will change Cults will change
and it really amplifies the power of the
controller of this who tells the story
and I'm not sure I haven't really
thought about that and I'm thinking
about it now because you're talking
about hyper personalization where I
think this is the flip side of it I mean
this is Isaiah Berlin's
conceptualization of positive and
negative Liberty
so positive Liberty is the freedom to
believe in isms fascism communism
islamism or kind of whatever right
whereas negative Liberty was the freedom
for being told what to do and so this
thing was like positive ones are bad
because they form these massive
movements and then they tend to kill
people because you have the gerardian
thing of romantic Theory where you want
why the people want and then there's a
scapegoat whereas negative Liberty is
the freedom for being told what to do
and that led to laissez-faire capitalism
and this consumerism that we saw around
the world
and so maybe as people lose meaning
they'll turn back to religion they'll be
new religions there'll be new political
movements and we're not sure what those
will be but they could spread faster
than anything we've ever seen
and so that's probably something to
watch out for within that five-year
period that you're talking about and I
think that relates to this network State
concept and other things but for the
people that
get engaged by this and again we see
that's largely the youth so on the one
hand you have the youth with the AI
girlfriend on the other hand you have
the youths that want to believe in
something bigger to fill the void
and who's going to step in yeah and I
think that there there is something

about not having a shared narrative that really makes me nervous so uh you've all know a Harari talks a lot about hey the thing that makes humans so intriguing is that we're not only able to organize these really large numbers but unlike ants that have to do it in a very strict way we can do it in a very flexible way but we do it through these shared narratives now for a long time religion served as the the thing that gave people a shared narrative but as religion breaks apart and we get into this hyper personalization and it all begins to fragment then you mix that with this idea that I heard from Jordan Peterson I'm almost certain heard it from somebody else but this idea that everybody has to go through a Messianic phase where they want to really contribute to the world they want to feel like they matter and they begin glomming on to all manner of things that seem good in the abstract like climate change but when you are glomming on the climate change is your way to save those world you begin to get into the realm of well it's okay if we have to break some eggs to make The Omelette yeah yeah and it rapidly devolves into Mal so how do we when you I don't you said you haven't really thought about this but I'm super curious at least in real time how you think about the idea of how do we do we need to give people a unifying narrative and if so how do we go about it we need to tell better more positive stories about the future and these are the stories of universal education Universal Health Care you know solving the mysteries of the universe and others so I've got a lot of because that's Hope For Humanity right
and a lot of the things that we see are dystopius because you're looking at the tiger you're spotting AI is the Tiger in the bush
and it's difficult to write a tiger but maybe that's a kind of cool picture that we can make in stable diffusion in two seconds right because it does have this duality of potential outcomes and maybe it's actually all of them so what are the stories that we should tell and I

think this is Again part of the crisis
of what is the American identity what is
the American story today
whereby you've gone through many cycles
what do Americans believe and what does
America want it to be
I'm not sure what Americans want America
to be
you know I'm not sure what Chinese
people want China to be I'm not sure
what people want and I think that it's
difficult to think about what is your
objective function how are you going to
measure your life and other things
religion failed a lot of those kind of
things but it still does
religion hasn't gone away
half the world is religious right more
probably like you actually look at the
numbers sure it decreases in certain
areas particularly somewhere like
America but it's going strong around the
rest of the world
and it's just growing because they have
more kids than non-religious people
maybe that fills the Gap but how will
the religion transform with this
technology I mean yes do you think that
the countries with religion will be the
ones that propagate into the future
because they have a better shared story
well not because they propagate
literally they like procreate even if
that's how the story ends up pushing
them forward I think it could be but
then you know what is the nature of
Christianity with AI or Islam with AI
Islam is actually the one that's most
affected by AI why so Christianity you
and Shia Islam you've got like popes
you've got protestantism you've got this
every single has their own structure
Sunni Islam is based on interpretation
of texts with the interpretation having
ceased around about the 16th century
because the text became too complex
what happens when you apply AI to that
and the texts are interpretable by
anyone with all the context and nuance
and there's no centralized Authority in
Sunni Islam which is like a billion
people
that's going to be very interesting what
does that do to protestantism

you know where you don't have
necessarily a pope
what does religion look like when all of
a sudden you have a branch that is AI
enhanced to interpret texts and to tell
stories that are resonant and better
oh gosh there's a lot to think about
there right does AI become a god well
some people are trying to build an AI
God that is Agi
you look at the statements of people
trying to build AJR they're trying to
build God
because it will bring us Utopia or kill
us all this sounds very again classical
right
and they have further
they generally believe that they are
going to save the world
or destroy it
yeah we got back to the dark stuff
didn't we yeah that's a joke you make
Game of Thrones season eight you know
like come on let's do it let's bring
this technology for cool stuff uh make
the Oasis in Ready Player one minus the
mighty crunch transactions and whiny
teenagers that part I actually am
working on there you go all right so
talk to a young person out there right
now they're terrified they wanna they
wanna be future-proofed
um
what what did they do how does somebody
right now future proof themselves they
just throw themselves into this area
there are so few people actually doing
it that if you go into this area
with all your might and curiosity in a
generally open mind you can actually
have an effect on the future
because everyone in your community will
be using this everyone that you know
will be using it if you're someone that
listens to this podcast again maybe not
the people without internet but you
don't know those people you know
and so you become a shelling point you
become the expert in this area ahead of
everyone because what happens is that
anyone who gets into it now
will have almost unassailable advantage
of people who come later so kind of
seniorish thing right

because you'll see it at the start it is
the start of the biggest change I think
that we've ever seen
and again think about what you're doing
when you're typing in and seeing that
and think about a million of these
things working they're even better
it's unavoidable so I'd say just you
just have to get into it you have to get
passionate
you have to think about the bad stuff
but is that really your responsibility
right I think it is but focus on the
good stuff and focus on the potential of
what happens on this scales to make real
positive change
can be to your pocket apparently to your
community it can be to your life because
it does affect everyone that you know
so I'd go with a positive mindset leave
it to boring old guys like us to think
about all the Doom scenarios
fair enough you might where can people
follow you uh I suppose my Twitter at
email stack um follow stability AI as
well yeah that's kind of the main
mouthpiece
I love it all right everybody if you
haven't already be sure to subscribe and
deploy some AI in your life and until
next time my friends be legendary take
care peace to learn more about
artificial intelligence check out this
episode with Mo Gadot we've never
created a nuclear weapon that can create
nuclear weapons the artificial
intelligences that we're building are
capable of creating other artificial
intelligences as a matter of fact
they're encouraged to create other in
artificial intelligence