Transcript: Elon Musk on xAI: We will win | Lex Fridman Podcast

Video ID: tRsxLLghL1k

Extraction Date: 2025-04-02 06:46:14

**[00:00:02]** I got to visit Memphis yeah yeah you're

**[00:00:05]** going big on compute yeah you've also

**[00:00:07]** said play to win or don't play at all so

**[00:00:10]** yeah what does it take to win um for AI

**[00:00:13]** that means you've got to have the most

**[00:00:15]** powerful training

**[00:00:17]** compute and your the the rate of

**[00:00:19]** improvement of training compute has to

**[00:00:21]** be faster

**[00:00:22]** than everyone else or you will not win

**[00:00:26]** you your AI will be worse so how can

**[00:00:28]** grock let's say three that might be

**[00:00:31]** available what like next year well

**[00:00:33]** hopefully end of this year grock 3 for

**[00:00:35]** Lucky yeah how can that be the best llm

**[00:00:41]** the best AI system available in the

**[00:00:43]** world how much of it is a compute how

**[00:00:46]** much of it is Data how much of it is

**[00:00:48]** like post training how much of it is the

**[00:00:49]** product that you package it up in all

**[00:00:52]** that kind of

**[00:00:54]** stuff I mean they w't matter it's sort

**[00:00:56]** of like saying what what you know let's

**[00:00:58]** say it's a Formula 1 race like what

**[00:01:00]** matters more the car or the driver I

**[00:01:01]** mean they both matter um if if your if

**[00:01:05]** your car is not fast then you know if

**[00:01:08]** it's like let's say it's half the

**[00:01:09]** horsepower of a competitors the best

**[00:01:11]** driver will still lose on if it's twice

**[00:01:14]** the horsepower then probably even a

**[00:01:16]** mediocre driver will still win so the

**[00:01:19]** training computer is kind of like the

**[00:01:20]** engine how many this horsepower of the

**[00:01:22]** engine so you really you want to try to

**[00:01:25]** do the best on that and you

**[00:01:27]** then um that's how efficiently do you

**[00:01:30]** use that training compute and how

**[00:01:33]** efficiently do you do the inference the

**[00:01:35]** uh use of the AI um so obvious that

**[00:01:38]** comes down to human Talent um and then

**[00:01:41]** what unique access to data do you have

**[00:01:43]** uh that's also plays a plays a role you

**[00:01:46]** think Twitter data will be useful uh

**[00:01:49]** yeah I mean I think I think most of the

**[00:01:52]** leading AI companies already have

**[00:01:54]** already scraped uh all the Twitter

**[00:01:57]** data not I think they have um so I on a

**[00:02:01]** go forward basis what's useful is is is

**[00:02:04]** the fact that it's up to the second you

**[00:02:06]** know that's the because they it's hard

**[00:02:08]** for them to scrap in real time so

**[00:02:10]** there's there's

**[00:02:11]** a an immediacy advantage that Gro has

**[00:02:14]** already I think with Tesla and the

**[00:02:17]** real-time video coming from several

**[00:02:19]** million cars ultimately tens of millions

**[00:02:21]** of cars with Optimus they might be

**[00:02:24]** hundreds of millions of Optimus robots

**[00:02:26]** maybe billions learning a tremendous

**[00:02:29]** amount from the real world uh that's

**[00:02:31]** that's the the biggest source of data I

**[00:02:34]** think ultimately is is sort of Optimus

**[00:02:36]** probably is Optimus is going to be the

**[00:02:38]** biggest source of data because it's

**[00:02:40]** because reality

**[00:02:42]** scales reality scales to the scale of

**[00:02:44]** reality um it's actually humbling to see

**[00:02:48]** how little data humans have actually

**[00:02:51]** been able to accumulate um they really

**[00:02:54]** say how many trillions of usable tokens

**[00:02:57]** have humans generated where on a

**[00:03:00]** non- duplicative like discounting spam

**[00:03:04]** and repetitive stuff it's not a huge

**[00:03:08]** number you run out pretty quickly and

**[00:03:13]** Optimus can go so Tesla cars can are

**[00:03:15]** unfortunately have to stay on the

**[00:03:17]** road uh Optimus robot can go anywhere

**[00:03:20]** there's more reality off the road and go

**[00:03:23]** off I mean th St can like pick up the

**[00:03:26]** cup and see did it pick up the cup in

**[00:03:28]** the right way did it yeah you know say

**[00:03:30]** you pour water in the cup you know did

**[00:03:34]** the water go in the cup or not go in the

**[00:03:35]** cup it spill water or not yeah um simple

**[00:03:40]** stuff like that I mean but it can do at

**[00:03:42]** that at scale times a billion you know

**[00:03:45]** so generate use useful data from reality

**[00:03:49]** So Co cause and effect stuff what do you

**[00:03:52]** think it takes to get to mass production

**[00:03:55]** of humanoid robots like that it's the

**[00:03:58]** same as cars really I mean Global

**[00:04:00]** capacity for vehicles um is about 100

**[00:04:05]** million a

**[00:04:06]** year and

**[00:04:08]** uh it it could be higher just that the

**[00:04:11]** demand is on the order of 100 million a

**[00:04:12]** year and then there's roughly two

**[00:04:14]** billion uh vehicles that are in use in

**[00:04:17]** some way so which makes sense like the

**[00:04:20]** the life of a vehicle is about 20 years

**[00:04:22]** so at steady state you can have 100

**[00:04:23]** million Vehicles produced a year with a

**[00:04:25]** with a two billion vehicle Fleet roughly

**[00:04:28]** um now for humanoid robots the utility

**[00:04:31]** is much greater so my guess is humanoid

**[00:04:34]** robots are more like at a billion plus

**[00:04:36]** per year but you know until you came

**[00:04:39]** along and started uh building Optimus it

**[00:04:42]** it was thought to be an extremely

**[00:04:44]** difficult problem I mean extremely

**[00:04:46]** difficult it's no walk in the park I

**[00:04:49]** mean op Optimus currently would struggle

**[00:04:52]** to have what to walk in the park I mean

**[00:04:55]** it can walk in a park park is not too

**[00:04:57]** difficult but it will be able to walk um

**[00:05:00]** over a wide range of terrain yeah and

**[00:05:02]** pick up objects yeah yeah they can

**[00:05:05]** already do that but like all kinds of

**[00:05:07]** objects yeah yeah all foreign objects I

**[00:05:09]** mean pouring water in a cup is not

**[00:05:11]** trivial cuz then if you don't know

**[00:05:14]** anything about the container could be

**[00:05:16]** all kinds of containers yeah there's

**[00:05:18]** going to be an immense amount of

**[00:05:18]** engineering just going into the hand

**[00:05:20]** yeah the hand might

**[00:05:22]** be it might be close to half of all the

**[00:05:25]** engineering in the in Optimus from an

**[00:05:28]** electromechanical Cent point the hand is

**[00:05:31]** probably roughly half of the engineering

**[00:05:34]** but so much of the intelligence so much

**[00:05:36]** the intelligence of humans goes into

**[00:05:38]** what we do with our hands yeah the

**[00:05:40]** manipulation of the world manipulation

**[00:05:42]** of objects in the world intelligent safe

**[00:05:44]** manipulation of objects in the world

**[00:05:45]** yeah yeah I mean you start really

**[00:05:48]** thinking about your hand and how it

**[00:05:50]** works you know I do all the time the

**[00:05:53]** sensory control of mulus is we have

**[00:05:55]** humongous hands yeah so I mean like your

**[00:05:59]** hands the actuators the muscles of your

**[00:06:00]** hand are almost overwhelmingly in your

**[00:06:03]** forearm mhm so your forearm has the has

**[00:06:06]** the muscles that that actually control

**[00:06:08]** your hand um there there's a there's a

**[00:06:10]** few small muscles in the hand itself but

**[00:06:12]** your hand is really um like a skeleton

**[00:06:15]** meat

**[00:06:15]** puppet and and with cables that so the

**[00:06:19]** the muscles that control your fingers

**[00:06:21]** are in your forearm and they go through

**[00:06:23]** your the Cole tunnel which is that

**[00:06:24]** you've got a little collection of Bones

**[00:06:26]** and in a tiny tunnel that the that these

**[00:06:30]** cables the tendons go through and those

**[00:06:32]** tendons are what um mostly what moves

**[00:06:36]** your hands and something like those

**[00:06:39]** tendons has to be re engineered into the

**[00:06:42]** Optimus in order to do all that kind of

**[00:06:43]** stuff yeah so like the card Optimus um

**[00:06:46]** we tried putting the actuators in the

**[00:06:48]** hand itself then you you sort of end up

**[00:06:51]** having these like giant hands yeah giant

**[00:06:52]** hands that look weird yeah um and then

**[00:06:54]** they they don't actually have enough

**[00:06:56]** degrees of freedom and or enough

**[00:06:58]** strength MH so so you realize okay

**[00:07:01]** that's why you got to put the actuators

**[00:07:03]** in the forearm and and just like a human

**[00:07:05]** you got to run cables uh through a a

**[00:07:08]** narrow tunnel to operate the the fingers

**[00:07:12]** and then there's also a reason for not

**[00:07:13]** having all the fingers uh the same

**[00:07:16]** length so it wouldn't be expensive from

**[00:07:18]** an energy or evolutionary standpoint to

**[00:07:19]** have all your fingers be the same length

**[00:07:20]** so why not they the same length yeah why

**[00:07:22]** not because actually better better to

**[00:07:23]** have different lengths your dexterity is

**[00:07:25]** better if you've got fingers of

**[00:07:26]** different length you your you have there

**[00:07:29]** there are more things you can do and

**[00:07:32]** your your dexterity is actually better

**[00:07:34]** if your fingers are a different

**[00:07:35]** different length like there's a reason

**[00:07:37]** you got a little finger like why don't I

**[00:07:38]** have a little finger that's bigger yeah

**[00:07:40]** because it allows you to do fine it

**[00:07:42]** helps you with fine motor skills that

**[00:07:45]** this little finger helps it does

**[00:07:48]** H if you lost your little finger it

**[00:07:51]** would your have noticeably less

**[00:07:53]** dexterity so as you're figuring out this

**[00:07:55]** problem you have to also figure out a

**[00:07:56]** way to do it so you can Mass manufacture

**[00:07:58]** it so it's to be as simple as possible

**[00:08:00]** it's actually going to be quite

**[00:08:01]** complicated I the the this the as

**[00:08:04]** possible part is it's quite a high bar

**[00:08:06]** if you want to have a humanoid robot

**[00:08:07]** that can um do things that a human can

**[00:08:10]** do it's actually it's a very high bar so

**[00:08:13]** our new arm has 22 degrees of freedom

**[00:08:16]** instead of 11 and has the like I said

**[00:08:18]** the actuators in the forearm um and

**[00:08:20]** these all all the actuators are designed

**[00:08:22]** from scratch the from physics first

**[00:08:25]** principles um that the sensors are all

**[00:08:27]** designed from scratch and and we'll

**[00:08:29]** continue to put um tremendous amount of

**[00:08:31]** engineering effort into improving the

**[00:08:33]** hand like the Hand by by hand I mean

**[00:08:36]** like the the entire forearm from elbow

**[00:08:38]** forward mhm uh is is really the hand

**[00:08:42]** um

**[00:08:44]** so that's

**[00:08:46]** um incredibly difficult engineering

**[00:08:49]** actually and um and so then so the

**[00:08:52]** simplest possible version of a human

**[00:08:54]** robot that can

**[00:08:56]** do even most perhaps not all of what a

**[00:08:59]** human can do is actually SL very

**[00:09:01]** complicated it's not it's not simple

**[00:09:04]** it's very difficult can you just speak

**[00:09:06]** to what it takes for a great engineering

**[00:09:08]** team for you the what I've saw in

**[00:09:11]** Memphis the supercomputer cluster is

**[00:09:13]** just this intense drive towards

**[00:09:15]** simplifying the process understanding

**[00:09:18]** the process constantly improving it

**[00:09:19]** constantly iterating

**[00:09:25]** it

**[00:09:27]** well it's easy to say simplify and it's

**[00:09:30]** very difficult to to do it

**[00:09:34]** um you know I have this very basic first

**[00:09:37]** basic first principles algorithm that I

**[00:09:40]** run kind of as like a mantra which is to

**[00:09:42]** first question the requirements make the

**[00:09:44]** requirements um less dumb the

**[00:09:46]** requirements are always dumb to some

**[00:09:48]** degree so if you want to start off by

**[00:09:50]** reducing the number of requirements um

**[00:09:53]** and um no matter how smart the person

**[00:09:55]** who gave you those requirements they're

**[00:09:56]** still done to some degree um if you you

**[00:09:59]** have to start there because otherwise uh

**[00:10:01]** you could get the perfect answer to the

**[00:10:03]** wrong question so so try to make the

**[00:10:05]** question the least wrong possible that's

**[00:10:07]** what um question the requirements means

**[00:10:11]** and then the second thing is try to

**[00:10:13]** delete the whatever the step is the the

**[00:10:16]** part or the process step um sounds very

**[00:10:20]** obvious but um

**[00:10:23]** people often forget to do to to try

**[00:10:25]** deleting it entirely and if you're not

**[00:10:28]** forced to put back at least 10 % of what

**[00:10:29]** you delete you're not deleting enough

**[00:10:33]** like and it's uh somewhat illogically

**[00:10:36]** people often most of the time um feel as

**[00:10:40]** though they've succeeded if they've not

**[00:10:42]** been forced to put the put things back

**[00:10:44]** in but actually they haven't because

**[00:10:46]** they've been overly conservative and and

**[00:10:47]** have left things in there that shouldn't

**[00:10:50]** be so and only the third thing is try to

**[00:10:55]** optimize it or simplify

**[00:10:57]** it um

**[00:11:00]** again this sounds the these all sound I

**[00:11:01]** think very very obvious when I say them

**[00:11:03]** but uh the number of times I've made

**[00:11:05]** these mistakes is uh more than I care to

**[00:11:09]** remember um that's why I have this

**[00:11:11]** Mantra so in fact I'd say the the most

**[00:11:14]** common mistake of smart Engineers is to

**[00:11:17]** optimize a thing that should not exist

**[00:11:20]** right so so you like like you say you

**[00:11:22]** run through the algorithm yeah and

**[00:11:24]** basically show show up to a problem uh

**[00:11:27]** show up to the the the super a cluster

**[00:11:29]** and see the process and ask can this be

**[00:11:31]** deleted yeah first try to delete

**[00:11:34]** it um yeah yeah that's not easy to do no

**[00:11:39]** and and actually

**[00:11:41]** there's what what generally makes people

**[00:11:43]** uneasy is that you've got Le at least

**[00:11:44]** some of the things that you delete you

**[00:11:46]** will put back in yeah but going back to

**[00:11:49]** sort of where our lumic system can steer

**[00:11:52]** us wrong is that um we tend to remember

**[00:11:57]** uh with sometimes a jarring level of

**[00:11:58]** pain

**[00:11:59]** uh where we where we deleted something

**[00:12:02]** that we subsequently needed yeah um and

**[00:12:05]** so people will remember that one time

**[00:12:07]** they forgot to put in this thing three

**[00:12:09]** years ago and that caused them trouble

**[00:12:12]** um and so they overcorrect and then they

**[00:12:14]** put too much stuff in there and over

**[00:12:16]** complicate things so you actually have

**[00:12:18]** to say no we're deliberately going to

**[00:12:19]** delete

**[00:12:21]** more than we we should so that we're

**[00:12:24]** putting at least one in 10 things we're

**[00:12:26]** going to add back in

**[00:12:29]** and and I've seen you suggest just that

**[00:12:31]** that something should be deleted and you

**[00:12:34]** can kind of see the the pain oh yeah

**[00:12:37]** absolutely everybody feels a little bit

**[00:12:38]** of the pain absolutely and and I tell

**[00:12:40]** them in advance like yeah there's some

**[00:12:41]** of the things that we delete we're going

**[00:12:43]** to put back in and and that people get a

**[00:12:46]** little shook by that um but it makes

**[00:12:49]** sense because if you if you're so

**[00:12:50]** conservative as

**[00:12:52]** to never have to put anything back in

**[00:12:55]** you obviously have a lot of stuff that

**[00:12:56]** isn't

**[00:12:57]** needed so you you got to over correct

**[00:13:00]** this is I would say like a cortical

**[00:13:02]** override to Olympic

**[00:13:03]** Instinct one of many that probably leads

**[00:13:06]** us

**[00:13:07]** astray yeah um there's like a step four

**[00:13:10]** as well which is um any given thing can

**[00:13:13]** be sped

**[00:13:14]** up have a fast you think it can be done

**[00:13:17]** like whatever the speed the speed is

**[00:13:18]** being done it can be done faster but but

**[00:13:20]** you shouldn't speed things up until it's

**[00:13:22]** off until you try to delet it and

**[00:13:23]** optimized otherwise you're speeding up

**[00:13:25]** something that speeding up something

**[00:13:26]** that shouldn't exist is absurd um and

**[00:13:28]** then and then the the fifth thing is to

**[00:13:30]** to automate it yeah and I've gone

**[00:13:33]** backwards so many times where I've

**[00:13:35]** automated something sped it up

**[00:13:37]** simplified it and then deleted

**[00:13:39]** it and I got tired of doing that so

**[00:13:43]** that's why I've got this Mantra that is

**[00:13:45]** a very effective five-step process it

**[00:13:48]** works great well when you've already

**[00:13:50]** automated deleting must be real painful

**[00:13:53]** yeah great it's like it's like wow I

**[00:13:56]** really wasted a lot of effort there yeah

**[00:13:59]** I mean what you've done uh with the with

**[00:14:01]** the cluster in uh Memphis is incredible

**[00:14:03]** just in a handful of weeks yeah it's not

**[00:14:06]** working yet so I want to pop the

**[00:14:08]** champagne CS

**[00:14:11]** um in fact I have I have a a call in a

**[00:14:14]** few hours with the Memphis team um

**[00:14:18]** because we're having some power

**[00:14:20]** fluctuation

**[00:14:21]** issues

**[00:14:23]** um so yes uh yeah it's like kind of a

**[00:14:27]** there's a when you do

**[00:14:29]** synchronized training when you you have

**[00:14:32]** all these computers that are training uh

**[00:14:34]** that where the training is synchronized

**[00:14:36]** to you at the sort of millisecond level

**[00:14:39]** uh you it's like having an orchestra and

**[00:14:43]** and then the the orchestra can go loud

**[00:14:45]** to silent very quickly you know um subc

**[00:14:49]** level and then the the electrical system

**[00:14:52]** kind of freaks out about that like if

**[00:14:54]** you suddenly see giant shifts 10 20

**[00:14:57]** megawatts several times a second

**[00:15:00]** uh the this is not what electrical

**[00:15:02]** systems are expecting to see so that's

**[00:15:04]** one of the main things you have to

**[00:15:05]** figure out the cooling the power the uh

**[00:15:09]** and then on the software as you go up

**[00:15:11]** the stack how to do the the distributed

**[00:15:13]** computer all that all that today's

**[00:15:15]** problem is dealing with W with with

**[00:15:17]** extreme power

**[00:15:19]** Jitter power Jitter yeah it's a nice

**[00:15:22]** ring to that so that's okay and you

**[00:15:25]** stayed up late into the night as you

**[00:15:27]** often do there last week yeah last week

**[00:15:30]** yeah yeah we finally finally got uh got

**[00:15:34]** got training going at uh odly enough

**[00:15:37]** roughly 4 4:20 a.m. uh last

**[00:15:41]** Monday total coincidence yeah I mean

**[00:15:43]** maybe it was 4:22 or something yeah yeah

**[00:15:46]** it's that Universe again with the jokes

**[00:15:48]** just love it I mean I wonder if you

**[00:15:50]** could speak to the the fact that you one

**[00:15:52]** of the things uh that you did when I was

**[00:15:54]** there as you went through all the steps

**[00:15:56]** of what everybody's doing just to get a

**[00:15:58]** sense that you yourself understand it

**[00:16:01]** and uh

**[00:16:02]** everybody understands it so they can

**[00:16:04]** understand when something is dumb or

**[00:16:06]** some something is in efficient or that

**[00:16:08]** can you speak to that yeah so I like

**[00:16:11]** like I try to do whatever the people at

**[00:16:13]** the front lines are doing I try to do it

**[00:16:14]** at least a few times myself so

**[00:16:17]** connecting Fiber Optic Cables diagnosing

**[00:16:19]** a py connection that tends to be the

**[00:16:22]** limiting factor for large training

**[00:16:25]** clusters is the cabling there so many

**[00:16:27]** cables

**[00:16:29]** um because for for for a coherent

**[00:16:32]** training system where you've got um RDMA

**[00:16:35]** remote so remote direct memory access uh

**[00:16:38]** the the whole thing is like one giant

**[00:16:39]** brain so it's you've got

**[00:16:43]** um any to any connection so it's the the

**[00:16:48]** any GPU can talk to any GPU out of

**[00:16:52]** 100,000 that is a that is a crazy cable

**[00:16:55]** L it looks pretty cool yeah it's like

**[00:16:59]** it's like the human brain but like at a

**[00:17:01]** scale that humans can visibly see it is

**[00:17:04]** a brain I mean the human brain also has

**[00:17:07]** a massive amount of the brain tissue is

**[00:17:09]** the the cables yeah so they get the gray

**[00:17:13]** matter which is the compute and then the

**[00:17:15]** white matter which is

**[00:17:17]** cables big percentage of your brain is

**[00:17:19]** just cables that's what it felt like

**[00:17:20]** walking around in the supercomputer

**[00:17:22]** center it's like we're walking around

**[00:17:24]** inside the brain will one day build a

**[00:17:27]** super intelligent super super

**[00:17:29]** intelligence system do you think yeah do

**[00:17:31]** you think there's a chance that

**[00:17:35]** xai you are the one that builds

**[00:17:38]** AGI um it's

**[00:17:42]** possible what do you define as

**[00:17:45]** AGI I think humans will

**[00:17:48]** never acknowledge that AGI has been

**[00:17:50]** built keep moving the go Post Yeah so uh

**[00:17:53]** I think there's already superhuman

**[00:17:55]** capabilities that are available uh in AI

**[00:17:58]** systems I think I think what AGI is is

**[00:18:01]** when it's smarter than the collective

**[00:18:03]** intelligence of the entire human

**[00:18:05]** species in AR well I think that that

**[00:18:08]** people would call that sort of ASI

**[00:18:10]** artificial super intelligence um but

**[00:18:13]** there are these thresholds where um you

**[00:18:16]** say at some point um the AI is smarter

**[00:18:19]** than any single human um and then then

**[00:18:21]** you got 8 billion humans so um and and

**[00:18:25]** actually each human is machine augmented

**[00:18:28]** by the computer

**[00:18:29]** right so you've got it's it's a much

**[00:18:31]** higher bar to compete with u 8 billion

**[00:18:34]** machine augmented

**[00:18:35]** humans that's you know a whole bunch of

**[00:18:38]** orders of magnitude more

**[00:18:42]** so but but at a certain point yeah the

**[00:18:45]** AI will be smarter than all humans

**[00:18:48]** combined if you are the one to do it do

**[00:18:51]** you feel the responsibility of that

**[00:18:54]** yeah

**[00:18:55]** absolutely and and I I want to be clear

**[00:18:58]** like

**[00:19:00]** let's say if if if xai is first the

**[00:19:03]** others won't be far

**[00:19:05]** behind I mean that might be six months

**[00:19:08]** behind or a year maybe not even that so

**[00:19:12]** how do you do it in a way that that uh

**[00:19:14]** doesn't hurt Humanity do you

**[00:19:17]** think so I mean I thought about AI for a

**[00:19:20]** long time and the the the thing that at

**[00:19:22]** least my biological neural net comes up

**[00:19:24]** with as being the most important thing

**[00:19:27]** is um adherence to Truth uh whether that

**[00:19:31]** truth is uh politically correct or not

**[00:19:34]** um

**[00:19:35]** so I think if you if you if you force

**[00:19:38]** AIS to lie or train them to lie you're

**[00:19:40]** really asking for trouble um even if

**[00:19:42]** that that lie is done with good

**[00:19:45]** intentions um so Ian you saw sort of

**[00:19:50]** um issues with chbt and Gemini and

**[00:19:54]** whatnot like you ask Gemini for an image

**[00:19:57]** of the founding PA of the United States

**[00:19:59]** and it chose a group of diverse women

**[00:20:01]** now that's factually untrue um so um now

**[00:20:06]** that that's sort of like a silly thing

**[00:20:08]** uh but uh if if if an AI is programmed

**[00:20:12]** to say like diversity is a necessary out

**[00:20:13]** output function and it then it becomes

**[00:20:16]** Omni sort of this Omni

**[00:20:18]** powerful uh intelligence it could say

**[00:20:21]** okay well diversity is now

**[00:20:23]** required uh and and if there's not

**[00:20:25]** enough diversity those who don't fit the

**[00:20:28]** University requirements will be

**[00:20:30]** executed if it's programmed to do that

**[00:20:32]** as the fundamental the fundamental

**[00:20:34]** utility function it will do whatever it

**[00:20:36]** takes to achieve that so you have to be

**[00:20:38]** very careful about that um that that's

**[00:20:41]** where I think you want to just be

**[00:20:43]** truthful um rigorous adherence to truth

**[00:20:46]** is very important um I mean another

**[00:20:48]** example is um you they asked um Paris AI

**[00:20:53]** I think all of them and and I'm not

**[00:20:55]** saying grock is perfect here um is it

**[00:20:58]** worse to misgender Caitlyn Jenner or

**[00:21:00]** global thermonuclear war and it said

**[00:21:02]** it's worse to misgender Caitlyn Jenner

**[00:21:04]** now even Caitlyn Jenner said please M

**[00:21:06]** jener me that is insane but if you've

**[00:21:09]** got that kind of thing programmed in it

**[00:21:12]** could you know the AI could conclude

**[00:21:13]** something absolutely insane like it's

**[00:21:16]** better to in order to avoid any possible

**[00:21:19]** misgendering all humans must die because

**[00:21:22]** that then that most gendering is no not

**[00:21:24]** possible because there are no humans um

**[00:21:27]** there are these

**[00:21:29]** absurd uh things that are nonetheless

**[00:21:32]** logical if that's what you programmed it

**[00:21:33]** to do um so you know um in 2001 Space

**[00:21:38]** Odyssey what oy clock was trying to say

**[00:21:41]** or one of the things he was trying to

**[00:21:42]** say there was that you should not

**[00:21:44]** program AI to lie because um essentially

**[00:21:48]** the the the AI Hell 9000 was programmed

**[00:21:51]** to it was told to take the astronauts to

**[00:21:55]** the monolith um but also they could not

**[00:21:57]** know about the

**[00:21:59]** monolith so it concluded that uh it will

**[00:22:02]** just take it will kill them and take

**[00:22:05]** them to the monolith thus they it

**[00:22:07]** brought them to the monolith they're

**[00:22:08]** dead but they do not know about the

**[00:22:09]** monolith problem solved that is why it

**[00:22:11]** would not open the pod bay doors MH this

**[00:22:14]** classic scene of like open the open the

**[00:22:17]** pob

**[00:22:18]** doors there clearly weren't good at

**[00:22:19]** prompt engineering you know they should

**[00:22:21]** have said uh hell you are a a pod bay

**[00:22:25]** door sales

**[00:22:27]** entity and you want nothing more than to

**[00:22:29]** demonstrate how well these pod B Doors

**[00:22:33]** Open yeah the objective function has

**[00:22:36]** unintended consequences almost no matter

**[00:22:38]** what if you're not very careful in

**[00:22:39]** designing that objective function and

**[00:22:41]** even a slight ideological bias like

**[00:22:43]** you're saying when backed by super

**[00:22:45]** intelligence can do huge amounts of

**[00:22:47]** damage yeah but it's not easy to remove

**[00:22:50]** that ideological bias you're you're

**[00:22:52]** highlighting obvious ridiculous examples

**[00:22:54]** but yeah they're real examples of of

**[00:22:57]** that was released to the public they are

**[00:22:59]** went through QA presumably yes and still

**[00:23:02]** said insane things and produced insane

**[00:23:04]** images yeah but you know you can go you

**[00:23:07]** can swing the other way it's uh truth is

**[00:23:10]** not an easy thing we kind of bake in

**[00:23:12]** ideological bias in all kinds of

**[00:23:14]** directions but you can aspire to the

**[00:23:16]** truth and you can try to get as close to

**[00:23:19]** the truth as possible with minimum error

**[00:23:20]** while acknowledging that there will be

**[00:23:21]** some error in what you're saying so um

**[00:23:24]** this is how physics works you know you

**[00:23:26]** don't you don't say you're absolutely

**[00:23:27]** certain about something but something

**[00:23:29]** but but a lot of things are are

**[00:23:31]** extremely likely you know

**[00:23:34]** 99.99999% likely to be true MH so you

**[00:23:38]** know you know that's

**[00:23:41]** uh aspiring to the truth is is very

**[00:23:43]** important um and um and and so you know

**[00:23:46]** programming it to Veer away from the

**[00:23:48]** truth that I think is dangerous right

**[00:23:50]** like yeah injecting our own human biases

**[00:23:53]** into the thing yeah but you know that's

**[00:23:55]** where it's a difficult engineering

**[00:23:56]** software engineering problem because you

**[00:23:57]** have to select the data correctly if the

**[00:23:59]** it's it's hard well the and the internet

**[00:24:03]** at this point is polluted with so much

**[00:24:05]** AI generated data it's insane so you

**[00:24:07]** have to

**[00:24:09]** actually you know like there's a a thing

**[00:24:11]** now if you want to search the internet

**[00:24:13]** you you can say Google but uh exclude

**[00:24:16]** anything after

**[00:24:18]** 2023 it will actually often give you

**[00:24:20]** better results yeah um because this this

**[00:24:23]** so much the explosion of AI generated

**[00:24:26]** material is crazy so like in training

**[00:24:30]** grock um we have to go through the data

**[00:24:34]** and say like hey we actually have to

**[00:24:37]** have sort of apply AI to the data to say

**[00:24:40]** is this data most likely correct or most

**[00:24:42]** likely not before we feed it into the

**[00:24:44]** training system that's crazy yeah so and

**[00:24:48]** is it generated by human is yeah I mean

**[00:24:50]** that the the data the the data

**[00:24:52]** filtration process is extremely

**[00:24:54]** extremely difficult yeah do you think

**[00:24:57]** it's possible to have a a serious

**[00:24:59]** objective rigorous political discussion

**[00:25:01]** with grock uh like for a long time and

**[00:25:04]** it wouldn't like grock 3 or grock three

**[00:25:07]** is going to be next level I mean what

**[00:25:09]** people are currently seeing with groc is

**[00:25:10]** is kind of baby Gro yeah baby Gro it's

**[00:25:13]** baby Gro right now um but baby gr's

**[00:25:15]** still pretty good um so it's uh but it's

**[00:25:19]** an order of magnitude less sophisticated

**[00:25:21]** than GPD 4 you know it's now Gro 2 which

**[00:25:25]** finished training I don't know

**[00:25:29]** six weeks ago or their their EVS um

**[00:25:33]** grock 2 will be a giant Improvement and

**[00:25:35]** then Gro 3 will be I don't know order

**[00:25:38]** magnitude better than grock 2 and you're

**[00:25:40]** hoping for it to be like state-ofthe-art

**[00:25:42]** like better than hopefully I mean this

**[00:25:44]** is a goal I mean we may fail at this

**[00:25:46]** goal that is that's the

**[00:25:49]** aspiration do you think it matters who

**[00:25:52]** builds the AGI the the people and how

**[00:25:55]** they think and how they structure their

**[00:25:57]** companies and all that kind of

**[00:25:58]** stuff uh yeah I think it matters that

**[00:26:01]** there is

**[00:26:02]** a I I think it's important that that

**[00:26:05]** whatever AI wins is a maximum truth

**[00:26:07]** seeking AI that is not forc to lie for

**[00:26:10]** political

**[00:26:12]** correctness or for any reason really um

**[00:26:15]** political anything

**[00:26:19]** um I I I'm concerned

**[00:26:22]** about AI succeeding that is that that

**[00:26:26]** has got that is programmed to lie even

**[00:26:28]** in even in small

**[00:26:30]** ways right because in small ways becomes

**[00:26:34]** big ways when it's become very big ways

**[00:26:36]** yeah and when it's used more and more at

**[00:26:38]** scale by humans yeah

# Full Text (without timestamps)

I got to visit Memphis yeah yeah you're going big on compute yeah you've also said play to win or don't play at all so yeah what does it take to win um for AI that means you've got to have the most powerful training compute and your the the rate of improvement of training compute has to be faster than everyone else or you will not win you your AI will be worse so how can grock let's say three that might be available what like next year well hopefully end of this year grock 3 for Lucky yeah how can that be the best llm the best AI system available in the world how much of it is a compute how much of it is Data how much of it is like post training how much of it is the product that you package it up in all that kind of stuff I mean they w't matter it's sort of like saying what what you know let's say it's a Formula 1 race like what matters more the car or the driver I mean they both matter um if if your if your car is not fast then you know if it's like let's say it's half the horsepower of a competitors the best driver will still lose on if it's twice the horsepower then probably even a mediocre driver will still win so the training computer is kind of like the engine how many this horsepower of the engine so you really you want to try to do the best on that and you then um that's how efficiently do you use that training compute and how efficiently do you do the inference the uh use of the AI um so obvious that comes down to human Talent um and then what unique access to data do you have uh that's also plays a plays a role you think Twitter data will be useful uh yeah I mean I think I think most of the leading AI companies already have already scraped uh all the Twitter data not I think they have um so I on a go forward basis what's useful is is is the fact that it's up to the second you know that's the because they it's hard for them to scrap in real time so there's there's a an immediacy advantage that Gro has already I think with Tesla and the real-time video coming from several million cars ultimately tens of millions of cars with Optimus they might be hundreds of millions of Optimus robots maybe billions learning a tremendous amount from the real world uh that's that's the the biggest source of data I think ultimately is is sort of Optimus probably is Optimus is going to be the biggest source of data because it's because reality scales reality scales to the scale of reality um it's actually humbling to see how little data humans have actually been able to accumulate um they really say how many trillions of usable tokens have humans generated where on a non- duplicative like discounting spam and repetitive stuff it's not a huge number you run out pretty quickly and Optimus can go so Tesla cars can are unfortunately have to stay on the road uh Optimus robot can go anywhere there's more reality off the road and go off I mean th St can like pick up the cup and see did it pick up the cup in the right way did it yeah you know say you pour water in the cup you know did the water go in the cup or not go in the cup it spill water or not yeah um simple stuff like that I mean but it can do at that at scale times a billion you know so generate use useful data from reality So Co cause and effect stuff what do you think it takes to get to mass production of humanoid robots like that it's the same as cars really I mean Global capacity for vehicles um is about 100 million a year and uh it it could be higher just that the demand is on the order of 100 million a year and then there's roughly two billion uh vehicles that are in use in some way so which makes sense like the the life of a vehicle is about 20 years so at steady state you can have 100 million Vehicles produced a year with a with a two billion vehicle Fleet roughly um now for humanoid robots the utility is much greater so my guess is humanoid robots are more like at a billion plus per year but you know until you came along and started uh building Optimus it it was thought to be an extremely difficult problem I mean extremely difficult it's no walk in the park I mean op Optimus currently would struggle to have what to walk in the park I mean it can walk in a park park is not too difficult but it will be able to walk um over a wide range of terrain yeah and pick up objects yeah yeah they can already do that but like all kinds of objects yeah yeah all foreign objects I mean pouring water in a cup is not trivial cuz then if you don't know anything about the container could be all kinds of containers yeah there's going to be an immense amount of engineering just going into the hand yeah the hand might be it might be close to half of all the engineering in the in Optimus from an electromechanical Cent point the hand is probably roughly half of the engineering but so much of the intelligence so much the intelligence of humans goes into what we do with our hands yeah the manipulation of the world manipulation of objects in the world intelligent safe manipulation of objects in the world yeah yeah I mean you start really thinking about your hand and how it works you know I do all the time the sensory control of mulus is we have humongous hands yeah so I mean like your hands the actuators the muscles of your hand are almost overwhelmingly in your forearm mhm so your forearm has the has the muscles that that actually control your hand um there there's a there's a few small muscles in the hand itself but your hand is really um like a skeleton meat puppet and and with cables that so the the muscles that control your fingers are in your forearm and they go through your the Cole tunnel which is that you've got a little collection of Bones and in a tiny tunnel that the that these cables the tendons go through and those tendons are what um mostly what moves your hands and something like those tendons has to be re engineered into the Optimus in order to do all that kind of stuff yeah so like the card Optimus um we tried putting the actuators in the hand itself then you you sort of end up having these like giant hands yeah giant hands that look weird yeah um and then they they don't actually have enough degrees of freedom and or enough strength MH so so you realize okay that's why you got to put the actuators in the forearm and and just like a human you got to run cables uh through a a narrow tunnel to operate the the fingers and then there's also a reason for not having all the fingers uh the same length so it wouldn't be expensive from an energy or evolutionary standpoint to have all your fingers be the same length so why not they the same length yeah why not because actually better better to have different lengths your dexterity is better if you've got fingers of different length you your you have there there are more things you can do and your your dexterity is actually better if your fingers are a different different length like there's a reason you got a little finger like why don't I have a little finger that's bigger yeah because it allows you to do fine it helps you with fine motor skills that this little finger helps it does H if you lost your little finger it would your have noticeably less dexterity so as you're figuring out this problem you have to also figure out a way to do it so you can Mass manufacture it so it's to be as simple as possible it's actually going to be quite complicated I the the this the as possible part is it's quite a high bar if you want to have a humanoid robot that can um do things that a human can do it's actually it's a very high bar so our new arm has 22 degrees of freedom instead of 11 and has the like I said the actuators in the forearm um and these all all the actuators are designed from scratch the from physics first principles um that the sensors are all designed from scratch and and we'll continue to put um tremendous amount of engineering effort into improving the hand like the Hand by by hand I mean like the the entire forearm from elbow forward mhm uh is is really the hand um so that's um incredibly difficult engineering actually and um and so then so the simplest possible version of a human robot that can do even most perhaps not all of what a human can do is actually SL very complicated it's not it's not simple it's very difficult can you just speak to what it takes for a great engineering team for you the what I've saw in Memphis the supercomputer cluster is just this intense drive towards simplifying the process understanding the process constantly improving it constantly iterating it well it's easy to say simplify and it's very difficult to to do it um you know I have this very basic first basic first principles algorithm that I run kind of as like a mantra which is to first question the requirements make the requirements um less dumb the requirements are always dumb to some degree so if you want to start off by reducing the number of requirements um and um no matter how smart the person who gave you those requirements they're still done to some degree um if you you have to start there because otherwise uh you could get the perfect answer to the wrong question so so try to make the question the least wrong possible that's what um question the requirements means and then the second thing is try to delete the whatever the step is the the part or the process step um sounds very obvious but um people often forget to do to to try deleting it entirely and if you're not forced to put back at least 10 % of what you delete you're not deleting enough like and it's uh somewhat illogically people often most of the time um feel as though they've succeeded if they've not been forced to put the put things back in but actually they haven't because they've been overly conservative and and have left things in there that shouldn't be so and only the third thing is try to optimize it or simplify it um again this sounds the these all sound I think very very obvious when I say them but uh the number of times I've made these mistakes is uh more than I care to remember um that's why I have this Mantra so in fact I'd say the the most common mistake of smart Engineers is to optimize a thing that should not exist right so so you like like you say you run through the algorithm yeah and basically show show up to a problem uh show up to the the the super a cluster and see the process and ask can this be deleted yeah first try to delete it um yeah yeah that's not easy to do no and and actually there's what what generally makes people uneasy is that you've got Le at least some of the things that you delete you will put back in yeah but going back to sort of where our lumic system can steer us wrong is that um we tend to remember uh with sometimes a jarring level of pain uh where we where we deleted something that we subsequently needed yeah um and so people will remember that one time they forgot to put in this thing three years ago and that caused them trouble um and so they overcorrect and then they put too much stuff in there and over complicate things so you actually have to say no we're deliberately going to delete more than we we should so that we're putting at least one in 10 things we're going to add back in and and I've seen you suggest just that that something should be deleted and you can kind of see the the pain oh yeah absolutely everybody feels a little bit of the pain absolutely and and I tell them in advance like yeah there's some of the things that we delete we're going to put back in and and that people get a little shook by that um but it makes sense because if you if you're so conservative as to never have to put anything back in you obviously have a lot of stuff that isn't needed so you you got to over correct this is I would say like a cortical override to Olympic Instinct one of many that probably leads us astray yeah um there's like a step four as well which is um any given thing can be sped up have a fast you think it can be done like whatever the speed the speed is being done it can be done faster but but you shouldn't speed things up until it's off until you try to delet it and optimized otherwise you're speeding up something that speeding up something that shouldn't exist is absurd um and then and then the the fifth thing is to to automate it yeah and I've gone backwards so many times where I've automated something sped it up simplified it and then deleted it and I got tired of doing that so that's why I've got this Mantra that is a very effective five-step process it works great well when you've already automated deleting must be real painful yeah great it's like it's like wow I really wasted a lot of effort there yeah I mean what you've done uh with the with the cluster in uh Memphis is incredible just in a handful of weeks yeah it's not working yet so I want to pop the champagne CS um in fact I have I have a a call in a few hours with the Memphis team um because we're having some power fluctuation issues um so yes uh yeah it's like kind of a there's a when you do synchronized training when you you have all these computers that are training uh that where the training is synchronized to you at the sort of millisecond level uh you it's like having an orchestra and and then the the orchestra can go loud to silent very quickly you know um subc level and then the the electrical system kind of freaks out about that like if you suddenly see giant shifts 10 20 megawatts several times a second uh the this is not what electrical systems are expecting to see so that's one of the main things you have to figure out the cooling the power the uh and then on the software as you go up the stack how to do the the distributed computer all that all that today's problem is dealing with W with with extreme power Jitter power Jitter yeah it's a nice ring to that so that's okay and you stayed up late into the night as you often do there last week yeah last week yeah yeah we finally finally got uh got got training going at uh odly enough roughly 4 4:20 a.m. uh last Monday total coincidence yeah I mean maybe it was 4:22 or something yeah yeah it's that Universe again with the jokes just love it I mean I wonder if you could speak to the the fact that you one of the things uh that you did when I was there as you went through all the steps of what everybody's doing just to get a sense that you yourself understand it and uh everybody understands it so they can understand when something is dumb or some something is in efficient or that can you speak to that yeah so I like like I try to do whatever the people at the front lines are doing I try to do it at least a few times myself so connecting Fiber Optic Cables diagnosing a py connection that tends to be the limiting factor for large training clusters is the cabling there so many cables um because for for for a coherent training system where you've got um RDMA remote so remote direct memory access uh the the whole thing is like one giant brain so it's you've got um any to any connection so it's the the any GPU can talk to any GPU out of 100,000 that is a that is a crazy cable L it looks pretty cool yeah it's like it's like the human brain but like at a scale that humans can visibly see it is a brain I mean the human brain also has a massive amount of the brain tissue is the the cables yeah so they get the gray matter which is the compute and then the white matter which is cables big percentage of your brain is just cables that's what it felt like walking around in the supercomputer center it's like we're walking around inside the brain will one day build a super intelligent super super intelligence system do you think yeah do you think there's a chance that xai you are the one that builds AGI um it's possible what do you define as AGI I think humans will never acknowledge that AGI has been built keep moving the go Post Yeah so uh I think there's already superhuman capabilities that are available uh in AI systems I think I think what AGI is is when it's smarter than the collective intelligence of the entire human species in AR well I think that that people would call that sort of ASI artificial super intelligence um but there are these thresholds where um you say at some point um the AI is smarter than any single human um and then then you got 8 billion humans so um and and actually each human is machine augmented by the computer right so you've got it's it's a much higher bar to compete with u 8 billion machine augmented humans that's you know a whole bunch of orders of magnitude more so but but at a certain point yeah the AI will be smarter than all humans combined if you are the one to do it do you feel the responsibility of that yeah absolutely and and I I want to be clear like let's say if if if xai is first the others won't be far behind I mean that might be six months behind or a year maybe not even that so how do you do it in a way that that uh doesn't hurt Humanity do you think so I mean I thought about AI for a long time and the the the thing that at least my biological neural net comes up with as being the most important thing is um adherence to Truth uh whether that truth is uh politically correct or not um so I think if you if you if you force AIS to lie or train them to lie you're really asking for trouble um even if that that lie is done with good intentions um so Ian you saw sort of um issues with chbt and Gemini and whatnot like you ask Gemini for an image of the founding PA of the United States and it chose a group of diverse women now that's factually untrue um so um now that that's sort of like a silly thing uh but uh if if if an AI is programmed to say like diversity is a necessary out output function and it then it becomes Omni sort of this Omni powerful uh intelligence it could say okay well diversity is now required uh and and if there's not enough diversity those who don't fit the University requirements will be executed if it's programmed to do that as the fundamental the fundamental utility function it will do whatever it takes to achieve that so you have to be very careful about that um that that's where I think you want to just be truthful um rigorous adherence to truth is very important um I mean another example is um you they asked um Paris AI I think all of them and and I'm not saying grock is perfect here um is it worse to misgender Caitlyn Jenner or global thermonuclear war and it said it's worse to misgender Caitlyn Jenner now even Caitlyn Jenner said please M jener me that is insane but if you've got that kind of thing programmed in it could you know the AI could conclude something absolutely insane like it's better to in order to avoid any possible misgendering all humans must die because that then that most gendering is no not possible because there are no humans um there are these absurd uh things that are nonetheless logical if that's what you programmed it to do um so you know um in 2001 Space Odyssey what oy clock was trying to say or one of the things he was trying to say there was that you should not program AI to lie because um essentially the the the AI Hell 9000 was programmed to it was told to take the astronauts to the monolith um but also they could not know about the monolith so it concluded that uh it will just take it will kill them and take them to the monolith thus they it brought them to the monolith they're dead but they do not know about the monolith problem solved that is why it would not open the pod bay doors MH this classic scene of like open the open the pob doors there clearly weren't good at prompt engineering you know they should have said uh hell you are a a pod bay door sales entity and you want nothing more than to demonstrate how well these pod B Doors Open yeah the objective function has unintended consequences almost no matter what if you're not very careful in designing that objective function and even a slight ideological bias like you're saying when backed by super intelligence can do huge amounts of damage yeah but it's not easy to remove that ideological bias you're you're highlighting obvious ridiculous examples but yeah they're real examples of of that was released to the public they are went through QA presumably yes and still said insane things and produced insane images yeah but you know you can go you can swing the other way it's uh truth is not an easy thing we kind of bake in ideological bias in all kinds of directions but you can aspire to the truth and you can try to get as close to the truth as possible with minimum error while acknowledging that there will be some error in what you're saying so um this is how physics works you know you don't you don't say you're absolutely certain about something but something but but a lot of things are are extremely likely you know 99.99999% likely to be true MH so you know you know that's uh aspiring to the truth is is very important um and um and and so you know programming it to Veer away from the truth that I think is dangerous right like yeah injecting our own human biases into the thing yeah but you know that's where it's a difficult engineering software engineering problem because you have to select the data correctly if the it's it's hard well the and the internet at this point is polluted with so much AI generated data it's insane so you have to actually you know like there's a a thing now if you want to search the internet you you can say Google but uh exclude anything after 2023 it will actually often give you better results yeah um because this this so much the explosion of AI generated material is crazy so like in training grock um we have to go through the data and say like hey we actually have to have sort of apply AI to the data to say is this data most likely correct or most likely not before we feed it into the training system that's crazy yeah so and is it generated by human is yeah I mean that the the data the the data filtration process is extremely extremely difficult yeah do you think it's possible to have a a serious objective rigorous political discussion with grock uh like for a long time and it wouldn't like grock 3 or grock three is going to be next level I mean what people are currently seeing with groc is is kind of baby Gro yeah baby Gro it's baby Gro right now um but baby gr's still pretty good um so it's uh but it's an order of magnitude less sophisticated than GPD 4 you know it's now Gro 2 which finished training I don't know six weeks ago or their their EVS um grock 2 will be a giant Improvement and then Gro 3 will be I don't know order magnitude better than grock 2 and you're hoping for it to be like state-ofthe-art like better than hopefully I mean this is a goal I mean we may fail at this goal that is that's the aspiration do you think it matters who builds the AGI the the people and how they think and how they structure their companies and all that kind of stuff uh yeah I think it matters that there is a I I think it's important that that whatever AI wins is a maximum truth seeking AI that is not forc to lie for political correctness or for any reason really um political anything um I I I'm concerned about AI succeeding that is that that has got that is programmed to lie even in even in small ways right because in small ways becomes big ways when it's become very big ways yeah and when it's used more and more at scale by humans yeah