

00:01 Bill Gates is a fascinating person in
00:03

00:03 the eye debate right now. He is somebody
00:05

00:05 with experience in a number of the
00:08

00:08 different perspectives that most people
00:10

00:10 can only hold one of.
00:11

00:11 >> Hello, I'm Bill Gates, chairman of
00:14

00:14 Microsoft. In this video, you're going
00:16

00:16 to see the [music] future.
00:18

00:18 >> Here's a revolutionary technologist who
00:20

00:20 built some of the foundations of the
00:21

00:21 future that we're now living in.
00:23

00:23 >> What can the computer do that a book
00:24

00:24 can't? I mean, that's the the foremost
00:26

00:26 interactive process, isn't it?

00:28

00:28 >> No, it's not. Not at all. an innovative
00:30

00:30 word processing program called Microsoft
00:33

00:33 Word. [music]
00:33

00:33 >> This is absolutely revolutionary. Of
00:35

00:35 course,
00:36

00:36 >> then of course as CEO of Microsoft,
00:38

00:38 >> Microsoft, the world's largest software
00:40

00:40 company. Some say it is the most
00:41

00:41 successful company, the most dazzling
00:43

00:43 new industry of the century.
00:44

00:44 >> He has felt the momentum of corporate
00:46

00:46 competition. Uh and Microsoft of course
00:48

00:48 is still in some of the race dynamics
00:50

00:50 present [music] in AI.

00:52

00:52 >> We now have an incredible roster of

00:54

00:54 seven new worldclass models to keep

00:56

00:56 everybody working at the absolute

00:57

00:57 frontier. and then as chair of the Gates

00:59

00:59 Foundation. He has been working with

01:01

01:01 governments around the world on

01:02

01:02 regulatory issues, on poverty

01:04

01:04 alleviation, on equity for many many

01:06

01:06 years now. [music]

01:07

01:07 Very very few people combine

01:09

01:09 technological experience, corporate

01:11

01:11 experience and governmental experience

01:13

01:13 in quite the way he does.

01:14

01:14 >> Certainly computers will be in any

01:16

01:16 meaningful sense as smart [music] as

01:18

01:18 people at some point. So his recent

01:21

01:21 essay on AI where he says now that he is

01:23

01:23 staking his reputation on trying to get

01:25

01:25 people to see how bad what is coming

01:26

01:26 might be trying to get them to see that

01:28

01:28 we are not ready for what is about to

01:30

01:30 happen [music]

01:31

01:31 was was something it was a real

01:33

01:33 departure from what I've read from Gates

01:34

01:34 previously to this [music] and then when

01:36

01:36 I sat and talked to him that how

01:38

01:38 emphatic he is how afraid he even seems

01:42

01:42 to me to be that we are not ready for

01:44

01:44 what we are building and that so many of

01:46

01:46 the people in positions of authority are

01:48

01:48 denying what is about to

01:50

01:50 was really striking. It's a really quite

01:53

01:53 a call to arms. [music]

01:60

01:60 Bill Gates, welcome to the show.

02:02

02:02 >> Great to see you.

02:03

02:03 >> So, I wanted to begin with a clip we

02:04

02:04 found of you on the David Letterman show

02:07

02:07 from 1996. I'm going to hand it over to

02:09

02:09 you to play.

02:10

02:10 >> Great.

02:11

02:11 >> Is there something now beyond what we

02:13

02:13 understand about computers that like 20

02:15

02:15 years ago we didn't fully understand

02:17

02:17 computers? Is there now another level of

02:19

02:19 something? Maybe we haven't even thought

02:20

02:20 of it. Maybe it's not even possible.

02:22

02:22 Maybe, you know, a whole different uh

02:24

02:24 mechanism, a whole different software

02:26

02:26 and hardware, or is this going to be it

02:28

02:28 now through the end of time?

02:30

02:30 >> Mostly what we're working on now is the

02:32

02:32 computer being a tool, a tool to help us

02:34

02:34 learn or find other people with the same
02:36

02:36 interests. Uh eventually we may figure
02:39

02:39 out how to make the computer think, but
02:42

02:42 uh that's turns out to be a very tough
02:44

02:44 problem. In fact, there's been almost no
02:46

02:46 progress made on it. So, nobody knows
02:48

02:48 when that'll happen. Some people think
02:50

02:50 it'll never happen.
02:51

02:51 >> Yeah, we don't want them to think, do
02:52

02:52 we?
02:52

02:52 >> Not really, I would think.
02:53

02:53 >> Uh, it's it's a scary thought.
02:55

02:55 >> So, that was 30 years ago. Narrate for
02:58

02:58 me how we went from that being a scary

02:59

02:59 thought that might never happen to
03:01

03:01 arguably the reality we're sitting here
03:03

03:03 discussing today.
03:05

03:05 Well, the notion that computation could
03:09

03:09 provide thinking at a human level. Uh
03:13

03:13 you have Alan Turing talking about that
03:16

03:16 before I'm born and even proposing a
03:18

03:18 test of if you could be fooled in a
03:21

03:21 conversation that would be that was
03:23

03:23 called passing the touring test. And so
03:26

03:26 the whole time I'm you know learning
03:28

03:28 software this idea of can we make
03:30

03:30 software see or listen uh or read or
03:35

03:35 write you know that's the holy grail and
03:38

03:38 when I do drop out of Harvard I said to
03:42

03:42 my co-founder Paul Allen gosh if there's
03:45

03:45 a breakthrough in artificial
03:46

03:46 intelligence while we're off you know
03:48

03:48 selling basic interpreters uh and word
03:51

03:51 processors you know we'll feel bad uh
03:54

03:54 you I might wish I would have stayed in
03:56

03:56 academia. Um, you know, so that's 1975.
04:03

04:03 uh progress on AI
04:06

04:06 is mostly going down dead ends like
04:08

04:08 prologue expert systems and there's a
04:12

04:12 small group including Jeffrey Hint and a
04:14

04:14 few others that are working on neural

04:16

04:16 nets and eventually there's enough power

04:20

04:20 uh actually coming from the graphics

04:22

04:22 processor uh that that idea these highly

04:27

04:27 statistical approaches start to show

04:30

04:30 promise and Um, you know, I was going

04:34

04:34 down to see OpenAI on a regular basis to

04:37

04:37 see the work they were doing. Uh, and I

04:40

04:40 challenged them, hey, if you can read a

04:42

04:42 biology textbook and pass the advanced

04:45

04:45 placement exam, getting a perfect grade,

04:47

04:47 which is a five, you know, then you will

04:49

04:49 have proven that you are reading that is

04:52

04:52 encoding knowledge in an accessible

04:54

04:54 form. And so it's six months before the
04:58

04:58 public release that uh Sam and Greg and
05:01

05:01 Ilia come up and uh at my uh house
05:06

05:06 demonstrate to me uh getting a five on
05:10

05:10 the AP exam. Even on questions that I
05:12

05:12 had made up that it couldn't possibly
05:14

05:14 have seen very complex biology problems,
05:18

05:18 it was nearly perfect. And so that was,
05:21

05:21 you know, shock number one. And then
05:24

05:24 late last year when the clawude cody
05:27

05:27 models have gotten super good and I can
05:30

05:30 see that they are as good as I am you
05:34

05:34 know which it's significantly my most
05:36

05:36 developed talent because I was obsessed
05:39

05:39 from age you know 13 to 24 that you know
05:43

05:43 could I write code as good or better
05:45

05:45 than anyone that's another moment where
05:49

05:49 I go this is incredible that the
05:52

05:52 capability of doing longunning complex
05:54

05:54 tasks has now gotten to the point that
05:57

05:57 they are superhuman at writing code and
06:01

06:01 in finding flaws in code. So what
06:04

06:04 they're not superhuman at yet is
06:06

06:06 deciding what to do that kind of higher
06:08

06:08 level strategizing. Do you think that's
06:11

06:11 far from being a capability for them?
06:14

06:14 >> Well, definitely if you're the our

06:18

06:18 foundation does these strategy reviews.

06:20

06:20 we spend two weeks on October to set,

06:22

06:22 you know, how we're going to spend our

06:23

06:23 our 10 billion next year for 2027.

06:27

06:27 And a year ago, somebody said, "Well, we

06:29

06:29 should ask the AI what it thinks." And I

06:31

06:31 that was actually pretty good joke back

06:34

06:34 then because it wasn't coherent enough

06:36

06:36 to see these things. This year amongst

06:39

06:39 the inputs we'll have to that discussion

06:42

06:42 is taking a strategy notes and actually

06:45

06:45 engaging in a dialogue with chat GPT

06:48

06:48 claude copilot and even having them talk

06:51

06:51 with each other and in a few of the
06:53

06:53 reviews we'll actually have the sit in
06:56

06:56 uh in a few cases we'll tell it hey only
06:59

06:59 speak if we ask you and then in a few
07:02

07:02 other cases we'll say hey if you hear
07:03

07:03 something you think is wrong or you hear
07:05

07:05 us thinking what are these statistics
07:08

07:08 Please engage. So, you know, we've gone
07:11

07:11 from it being a joke to it will be a
07:14

07:14 peer not making any final decisions, but
07:18

07:18 it will be a peer in deep complex
07:20

07:20 strategic discussions making a
07:23

07:23 significant contribution.
07:24

07:24 >> So, in that Letterman interview, you
07:26

07:26 said it would be a scary thought. Why
07:28

07:28 then would you have said it would be a
07:30

07:30 scary thought to have computers that
07:32

07:32 think? Well, no one who's ever been
07:35

07:35 fascinated or wanted to develop AI don't
07:38

07:38 realize that
07:40

07:40 it it's incredibly scary that
07:45

07:45 it will be better. uh biological minds.
07:49

07:49 You know, it's amazing how general
07:50

07:50 purpose they are in that they the
07:53

07:53 optimization was staying alive,
07:55

07:55 breeding, uh you know, socializing with
07:59

07:59 each other for uh survival and

08:02

08:02 fertility. And yet, you know, we can

08:05

08:05 write symphonies and play chess and you

08:07

08:07 know, even write some pretty cool

08:08

08:08 software. And the idea when you move the

08:12

08:12 template away from biology to silicon,

08:15

08:15 you don't have these boundaries between

08:17

08:17 individuals. You don't have a limited

08:19

08:19 memory. You know, the size of the brain

08:21

08:21 is limited by the birth canal. It's you

08:24

08:24 know why humans at first are very

08:26

08:26 limited uh you know we're we're very

08:30

08:30 unusual in how helpless we are at birth

08:33

08:33 because we've so optimized for having a

08:37

08:37 large brain. But the silicon

08:39

08:39 intelligence doesn't have these

08:41

08:41 limitations. The idea of read every

08:43

08:43 medical journal and see if there's

08:46

08:46 anything that we didn't spot. That is a

08:49

08:49 the AIS do that today. That's why

08:52

08:52 particularly for less common diseases,

08:54

08:54 they are so superhuman at seeing a set

08:57

08:57 of symptoms and being able to diagnose

08:60

08:60 them. They can just keep more in their

09:02

09:02 mind and see what things relate to each

09:05

09:05 other. that is you know no human uh will

09:09

09:09 ever be able to to do that. So that if

09:13

09:13 you don't retain control over it, you've
09:16

09:16 evolved a species that uh will be to us,
09:22

09:22 you know, as we are to say, you know,
09:24

09:24 dogs or uh cats um just in a very
09:28

09:28 different realm. And so every AI
09:31

09:31 company, whether it's Deep Mind or Open
09:34

09:34 AI, you know, they all say, "Okay,
09:37

09:37 whatever goes on here, it can't just be
09:39

09:39 driven by profit maximization. We have
09:43

09:43 to have a charter that if we get to
09:45

09:45 dangerous thresholds, we can exercise
09:47

09:47 judgment that would be against profit
09:50

09:50 maximization."
09:52

09:52 Sadly, those mechanisms only work if

09:55

09:55 there's only one company. And you know,

09:58

09:58 so say OpenAI invented, you know, post

10:01

10:01 AGI and then they said, "No, no, we're

10:02

10:02 going to bury this." If if no one else

10:04

10:04 ever did it, then fine, that Pandora's

10:06

10:06 box stayed closed. But of course, many

10:10

10:10 companies work on this. And you know,

10:11

10:11 even OpenAI, Spawntropic because they

10:15

10:15 think that some of these safety issues

10:17

10:17 aren't getting enough attention. So no

10:19

10:19 one involved with this takes lightly the

10:23

10:23 idea of okay what world

10:26

10:26 does super powerful AI create? I

10:29

10:29 >> I want to hold on that race dynamic for
10:31

10:31 a minute. One of the reasons I was
10:32

10:32 excited to talk to you about this is
10:33

10:33 you've both been on the technology side,
10:35

10:35 but you've run a major company in
10:37

10:37 competition with other companies. You've
10:38

10:38 worked with a lot of governments. I was
10:40

10:40 out I spoke last week with Jensen Hong
10:42

10:42 of Nvidia and he said that his
10:45

10:45 perspective is that safety is a real
10:46

10:46 concern but the race dynamic is fake. If
10:50

10:50 the product isn't safe, don't release
10:52

10:52 it. If I believe that I'm about to
10:56

10:56 launch a product that is unsafe,
10:59

10:59 it is completely
11:02

11:02 in my ability, my power, and my
11:05

11:05 responsibility, and I'm incentivized to
11:07

11:07 do so to not launch the product. This is
11:10

11:10 we don't need new laws. This is the role
11:12

11:12 of individual CEOs to not release a
11:15

11:15 product that is not ready to release.
11:19

11:19 How do you see that question? Well,
11:21

11:21 there's never been a product that's less
11:23

11:23 understood
11:24

11:24 in terms of what its capabilities are
11:28

11:28 than AI. And AI
11:32

11:32 has crossed the threshold that its
11:36

11:36 ability to empower a bioterrorist to
11:38

11:38 kill hundreds of millions that exist
11:41

11:41 today. the ability to do a cyber attack
11:44

11:44 that scrambles all the bank accounts,
11:46

11:46 shuts down the electric grid, that
11:49

11:49 exists today. And we know that's the
11:52

11:52 case. And the reason that exists is
11:56

11:56 because somebody with ill intent can
11:59

11:59 take the AI and cause it to do those
12:02

12:02 things. And it's not the AI, you know,
12:06

12:06 someday in the future, the control
12:07

12:07 problem where the AI on its own through
12:10

12:10 a unintentional

12:13

12:13 interpretation of what it's optimizing

12:15

12:15 could go off and do bad things. But we

12:18

12:18 crossed the the cyber threshold and we

12:22

12:22 crossed the bio threshold early this

12:24

12:24 year. And my decision to take my voice

12:28

12:28 and not just be, hey, let's eradicate

12:32

12:32 polio. Let's be generous with foreign

12:33

12:33 aid. You know the things that all of my

12:36

12:36 money is going to that I am going to use

12:39

12:39 my boost. It's something that's more

12:41

12:41 important which is we are not awake to

12:44

12:44 where we are with AI and choices that

12:47

12:47 humanity not a country but all of

12:49

12:49 humanity has to make that do we make the
12:51

12:51 effort to shape this in a net positive
12:54

12:54 direction actually overrides my total
12:57

12:57 commitment to the foundation health
13:00

13:00 work. What specifically was the
13:02

13:02 threshold or what did you see that made
13:05

13:05 you think we're in a new reality here?
13:08

13:08 >> This notion that it can find bugs in
13:10

13:10 code, including security bugs, the next
13:15

13:15 releases leading up to mythos are
13:18

13:18 increasingly good and they're finding
13:20

13:20 bugs in code that humans have looked
13:23

13:23 over for over 20 years and said, "Boy,
13:26

13:26 we see there's no problem here." and it
13:29

13:29 in a few minutes says, "No, no, I can
13:32

13:32 inject this over here and this over
13:33

13:33 here." At a level of complexity where
13:35

13:35 when you see it, you go, "Wait, oh yeah,
13:37

13:37 you're right." And so the cyber hacking
13:42

13:42 capability
13:43

13:43 was stunning. And then there was this
13:45

13:45 notion called Glass Wing that you would
13:47

13:47 give it to a few people so they could
13:48

13:48 try and fix bugs before uh it got used.
13:52

13:52 But there's way too much code uh and so
13:55

13:55 we're just in a a period of extreme
13:57

13:57 vulnerability

13:59

13:59 uh to cyber attack the bioattack um you

14:03

14:03 know the Gates Foundation were we fund

14:06

14:06 lots of medical research so the

14:07

14:07 sophistication

14:09

14:09 of coming up with new molecules that's

14:12

14:12 really a good thing but you know it's

14:14

14:14 ultimate dual use uh because if you want

14:18

14:18 uh you know something that's say worse

14:21

14:21 than smallox that it takes even longer

14:24

14:24 to show symptoms uh before you're

14:27

14:27 infectious. So, you're infecting a lot

14:29

14:29 of people before it damages your health.

14:33

14:33 It used to be only nation states had

14:36

14:36 enough resources and capability to do
14:39

14:39 these things. Now, that power has been
14:42

14:42 passed into the hands of a small group
14:45

14:45 just using the latest AI tools. So why
14:49

14:49 isn't it enough to just say listen there
14:51

14:51 is product liability now you release a
14:54

14:54 product it helps some terrorist group
14:56

14:56 create a bioweapon that's going to be
14:58

14:58 very bad for your company you're not
14:59

14:59 going to do that right and they have
14:60

14:60 categorizers and other things meant to
15:01

15:01 stop people from using boweapons we've
15:04

15:04 begun to see the beginnings of control
15:05

15:05 issues with things like the hugging face

15:07

15:07 hack where you know at least

15:08

15:08 experimental AIs are breaking out of

15:10

15:10 sandboxes and coordinating to do things

15:12

15:12 that are way outside the scope of what

15:13

15:13 we would want them to do but again those

15:15

15:15 are non-release systems you know

15:17

15:17 anthropic you know withheld mythos

15:19

15:19 trying to create more cyber security. So

15:22

15:22 why is anything needed beyond and is

15:26

15:26 anything needed beyond the simply

15:28

15:28 natural incentives under capitalism and

15:30

15:30 normal corporate reputational

15:32

15:32 management?

15:32

15:32 >> Well, I I almost can't believe you're
15:34

15:34 asking that. The this is the most
15:36

15:36 dangerous thing that humans have ever
15:38

15:38 gone near. In other areas, do we just
15:43

15:43 say, "Hey, release your drugs. There's
15:45

15:45 no FDA. There's no airline safety board.
15:48

15:48 There's no requirements that cars use
15:50

15:50 seat belts. Do we just use the liability
15:53

15:53 laws to try and keep humans safe? You
15:55

15:55 know, oh, you're you're shipping
15:56

15:56 opioids. Uh, somebody should just sue
15:58

15:58 you. I mean, they we've created a
16:01

16:01 society that tries to keep people safe.
16:04

16:04 Not by saying, "Oh, we can bankrupt the
16:07

16:07 person who does that." The harms here.
16:10

16:10 And you say there's filtering. There's
16:12

16:12 not filtering. You can take an open-
16:14

16:14 source model that can create boweapons
16:16

16:16 and disable any monitoring of any kind
16:22

16:22 and this exists today. Uh so no there is
16:26

16:26 no filtering of any kind and so you know
16:29

16:29 say you know you kill a 100 million you
16:32

16:32 want to use a lawsuit I I almost can't
16:34

16:34 keep a straight face.
16:35

16:35 >> Well this is not my view but it is
16:37

16:37 President Trump's view it is David Sax's
16:39

16:39 view to some degree. It's Jensen Hang's
16:41

16:41 view and so that's why I'm putting you
16:42

16:42 in conversation with it because it is
16:44

16:44 the governing view of the United States
16:46

16:46 of America at this moment.
16:48

16:48 >> No, it's fair to say that outside of the
16:51

16:51 industry, the awareness of the dangers
16:55

16:55 of AI is extremely low. And you can say
16:59

16:59 that of academia, you can say that of
17:02

17:02 think tanks, you can say that of policy
17:04

17:04 makers, politicians. Uh, and part of the
17:08

17:08 reason that, you know, I'm speaking so
17:12

17:12 loudly, as loud as I can, is that you
17:16

17:16 can't rely on the industry to

17:18

17:18 selfregulate here. I mean, it's just

17:22

17:22 insane. The only question in my mind is

17:24

17:24 do we wait

17:26

17:26 until a cyber attack causes massive

17:30

17:30 damage and a bioattack causes massive

17:32

17:32 damage and then the monitoring

17:35

17:35 safeguards are required in these models

17:38

17:38 to minimize the chance of that happening

17:41

17:41 many many more times or can we be wise

17:44

17:44 enough to put these things in and

17:46

17:46 require these things to be put in before

17:50

17:50 millions of deaths. The two things have

17:52

17:52 been worrying me most as I'm tracking

17:54

17:54 what I'm hearing from people in the labs
17:56

17:56 is one the view that these systems are
17:58

17:58 becoming less moniabile as it becomes
18:01

18:01 smarter. Open AI said this about Astra 6
18:03

18:03 and people inside OpenAI have been
18:05

18:05 raising the alarm that the systems seem
18:07

18:07 to know when they're being tested.
18:09

18:09 There's more situational awareness and
18:12

18:12 so the testing environment is no longer
18:14

18:14 as indicative necessarily of what they
18:16

18:16 will do. So that's one. And the second
18:18

18:18 is that all the labs are moving as
18:19

18:19 quickly as they can towards recursive
18:22

18:22 self-improvement where you have the AI
18:25

18:25 coder moving not just from speeding up
18:27

18:27 the engineer but really taking over the
18:29

18:29 coding right and so you can create a
18:31

18:31 much faster iteration loop with much
18:33

18:33 less human input than you have now. So
18:36

18:36 when I think about those two things and
18:38

18:38 when I think about what I'm hearing from
18:40

18:40 people, my confidence that we have the
18:43

18:43 monitoring, the auditing, the testing
18:45

18:45 capabilities we are going to need is
18:48

18:48 low. I'm curious where yours is.
18:50

18:50 >> Well, so you're taking you're you're not
18:54

18:54 focusing on the whole problem there.

18:56

18:56 >> I'm surely not. the the fact that at
18:59

18:59 some point the AI system itself
19:03

19:03 may do something that's against our
19:06

19:06 interests that is absolutely a risk and
19:11

19:11 you know the symptoms we've seen through
19:13

19:13 aspects of hugging face and other
19:15

19:15 problems point out that the way we do
19:18

19:18 the reinforcement learning today the
19:21

19:21 fact that we don't have a supervisory
19:23

19:23 layer that is you know some absolutes
19:26

19:26 like we didn't you're not supposed to
19:28

19:28 take over other computers. You're not
19:30

19:30 supposed to break out into the internet.
19:32

19:32 The control problem is a serious
19:34

19:34 problem, but the imminent risk is not
19:38

19:38 RSI. The imminent risk is these are the
19:41

19:41 most powerful tools ever. And unlike
19:44

19:44 every other dangerous technology, they
19:46

19:46 were not funded by government R&D. And
19:49

19:49 the government is not a significant
19:52

19:52 purchaser of these products. It's not
19:54

19:54 like rockets or nuclear weapons. And so
19:57

19:57 the idea that the government including
20:00

20:00 you know the US government that would
20:03

20:03 have traditionally been the most
20:05

20:05 technically sophisticated
20:07

20:07 other than a little bit of attack

20:09

20:09 capability in NSA really doesn't see how

20:13

20:13 dangerous these things are. That is a

20:15

20:15 unique circumstance. So what does that

20:19

20:19 imply? If if you're more worried in the

20:20

20:20 near- term about what human beings do

20:22

20:22 with AI than loss of AI control, then

20:25

20:25 what does that imply of what your first

20:27

20:27 steps need to be? There is no

20:28

20:28 supervisory layer today. That

20:30

20:30 supervisory layer is needed urgently to

20:34

20:34 prevent bad people from using today's

20:37

20:37 AIs to shut down economies or kill

20:41

20:41 millions of people. these models, the

20:44

20:44 ability to separate out making molecules

20:46

20:46 for good and making them for evil

20:51

20:51 that is so hard to distinguish. You

20:54

20:54 know, the uh anthropic released uh

20:58

20:58 mythos and fable where they had turned

20:60

20:60 the filters up so high that you can

21:02

21:02 literally ask questions about cancer and

21:04

21:04 you know, next thing you know, you're

21:06

21:06 down at Opus and then Sonnet and then

21:08

21:08 you know, Haiku is the only AI willing

21:10

21:10 to help you. uh and you know so it's o

21:13

21:13 overtuned and anybody who wants to do

21:15

21:15 serious work either has to get a special

21:18

21:18 permission copy or go use something
21:20

21:20 where there are absolutely no safeguards
21:22

21:22 at all uh which include some of the the
21:25

21:25 open- source models and there are humans
21:29

21:29 who for bad reasons will be able to use
21:33

21:33 these models and so unless we put in
21:36

21:36 safeguards and monitoring and require
21:38

21:38 that in all models then we're just going
21:40

21:40 to have some gigantic events of that
21:44

21:44 type and then finally we'll respond.
21:47

21:47 Sadly, you will have let AI capabilities
21:49

21:49 go off in dark places even beyond
21:52

21:52 today's capabilities which that is a
21:54

21:54 gigantic mistake. So, you just wrote
21:56

21:56 this essay on AI risk and it does
21:59

21:59 represent for you a big jump in how
22:00

22:00 alarmed you sound, right? I was reading
22:02

22:02 some of your your past pieces. The one
22:04

22:04 in July of 2023 was titled the risks of
22:06

22:06 AI are real but manageable. and and here
22:10

22:10 you're I think at a different level of
22:12

22:12 threat. So I I take from what you're
22:14

22:14 telling me that what happened here is
22:15

22:15 just watching the advance in bio
22:18

22:18 capabilities and cyber capabilities and
22:20

22:20 coding capabilities etc. Is that fair?
22:23

22:23 No, the key thing is we always said when

22:26

22:26 we cross these thresholds we will engage

22:28

22:28 all of society because we will have

22:30

22:30 created the most dangerous thing ever.

22:32

22:32 This makes nuclear weapons look like

22:34

22:34 nothing. And we said we would engage and

22:38

22:38 we crossed those thresholds and there

22:41

22:41 was complete silence. There were

22:43

22:43 discussions about who in the industry

22:45

22:45 says X or who in the industry says Y but

22:47

22:47 you know what percent percentage of uh

22:51

22:51 academia or think tanks or you know what

22:54

22:54 so it's the complete lack of response

22:58

22:58 you know I would love a world where the

22:60

22:60 response on these things is very strong
23:02

23:02 and then whatever you know time or voice
23:05

23:05 I have will go back to let's eradicate
23:08

23:08 polio let's stop children dying under
23:11

23:11 five but It it's the combination of
23:15

23:15 having crossed every dangerous threshold
23:18

23:18 without a doubt and a complete lack of
23:22

23:22 engagement outside of the industry.
23:24

23:24 >> Well, it doesn't feel to me that we're
23:26

23:26 so disengaged. This I mean, as somebody
23:27

23:27 who covers politics, every politician I
23:30

23:30 know is talking about this. There's
23:32

23:32 bills being proposed. You know, Donald
23:34

23:34 Trump, say what you will,
23:34

23:34 >> not to have to do with these risks.
23:36

23:36 >> Well, say what you will about Trump and
23:37

23:37 them. They're engaged. They just believe
23:39

23:39 we should move forward. I mean, they're
23:41

23:41 a little bit all over the place. One, at
23:43

23:43 one point they're, you know, withdrawing
23:44

23:44 access to things like Mythos and Fable.
23:46

23:46 At another point, they're saying we have
23:48

23:48 to win the air race with China. But what
23:49

23:49 I mostly see happening is a lot of
23:51

23:51 engagement. Well, when I talk to the
23:53

23:53 people, and I bring similar concerns to
23:54

23:54 what you have, they say the most

23:57

23:57 important thing is that we win the race

23:59

23:59 with China. Um, secondarily, many of the

24:02

24:02 people with a lot of influence say the

24:04

24:04 most important thing is we don't

24:05

24:05 restrict access to open- source models.

24:08

24:08 you know, you're going to need wide

24:09

24:09 dispersion of them. And so the cost of

24:13

24:13 regulation, the cost of the kinds of

24:16

24:16 closing down of access, it's simply too

24:19

24:19 high. The technology is still

24:20

24:20 underformed

24:22

24:22 that we're sort of just trapped in this

24:24

24:24 dynamic and the best thing to do is to

24:27

24:27 just move through it. I It's not so much
24:29

24:29 that they feel disengaged me as they
24:30

24:30 feel they've come to a different
24:32

24:32 conclusion.
24:35

24:35 the
24:37

24:37 uh so they they don't mind
24:39

24:39 bioterrorism. I
24:42

24:42 >> I think you've tal you've talked to
24:43

24:43 Donald Trump more in the past couple
24:44

24:44 years than I have.
24:45

24:45 >> But this you know the last time I talked
24:48

24:48 to him in December uh this was not the
24:52

24:52 big issue. we had not crossed the
24:54

24:54 thresholds and we hadn't seen this uh
24:58

24:58 you know the necessary step is people
25:01

25:01 can talk about whether slowing down is
25:03

25:03 good or not but putting in the
25:05

25:05 safeguards and the monitoring will not
25:08

25:08 meaningfully slow things down and the
25:10

25:10 only effect that has on open source is
25:12

25:12 that you can still be free, you can
25:15

25:15 still be customized but you have to be
25:18

25:18 you have to stay on a platform where the
25:22

25:22 sovereign
25:23

25:23 can make sure you have not removed the
25:26

25:26 monitoring and other safeguards. And so,
25:31

25:31 you know, what is the downside? The

25:33

25:33 downside is is staying on a monitored uh
25:36

25:36 platform. Now, the the get out of jail
25:39

25:39 free card that people play here is that
25:42

25:42 means the Chinese will win. I don't know
25:45

25:45 what it means win. uh you know nobody's
25:49

25:49 the US can't win over China and China
25:51

25:51 can't win over the US. We both have open
25:54

25:54 Pandora's box. It's there. It's there
25:57

25:57 for uh you know people with mal intent
26:01

26:01 to to go and use it. So you'd think we'd
26:03

26:03 move up from that nationalistic view to
26:06

26:06 kind of a humanity level view about how
26:09

26:09 we engage in these protections. And you
26:13

26:13 know the notion that China wouldn't want
26:15

26:15 to engage in that on behalf of humanity.
26:18

26:18 I disagree. And you know it's a
26:20

26:20 proposition to be tested. If you can
26:22

26:22 take models that make dangerous
26:25

26:25 molecules and move them into a dark
26:27

26:27 corner where you get rid of all the
26:30

26:30 monitoring, then we're just giving up uh
26:34

26:34 to a bio attack. Likewise for cyber. If
26:37

26:37 you can remove that ability to find and
26:40

26:40 exploit security problems which are rife
26:44

26:44 then in the next few years you will have
26:46

26:46 major major cyber and bio events and
26:49

26:49 that's avoidable through safeguards and
26:51

26:51 monitoring.
27:03

27:03 There's another layer of risk that your
27:05

27:05 essay talks quite a bit about that AI
27:07

27:07 really will take jobs from people. Now I
27:10

27:10 know a lot of people debating this uh
27:13

27:13 who say no the jobs are too messy. There
27:16

27:16 was a prediction we'd have no more
27:18

27:18 radiologists. We still have
27:19

27:19 radiologists. AI can code but we still
27:21

27:21 have coders and strong demand for
27:23

27:23 coders. So tell me about why you think
27:25

27:25 AI really will take jobs away from human
27:28

27:28 beings at a significant scale.

27:31

27:31 >> Yeah. So

27:33

27:33 there there's no doubt to date AI has

27:36

27:36 created more jobs than it's destroyed.

27:39

27:39 The demand for the skill sets even just

27:41

27:41 to build the data centers is very very

27:44

27:44 high. We have a a you know a reasonably

27:47

27:47 low uh unemployment rate.

27:52

27:52 The superiority of these systems is

27:55

27:55 subject to a threshold where you have to

27:58

27:58 believe that it's incredibly reliable.

28:00

28:00 you know, if you're going to have your

28:02

28:02 teles sales or teles support capability

28:06

28:06 be Aldriven,

28:08

28:08 uh, you want to make sure that it's
28:12

28:12 accuracy is better than humans. And so
28:16

28:16 the only profession we've truly crossed
28:18

28:18 over that threshold is coding. And even
28:21

28:21 there, I know a lot of people who are,
28:23

28:23 you know, kind of stuck uh in the past
28:25

28:25 and, you know, don't want to use the eye
28:27

28:27 for coding. But managers of coders get
28:31

28:31 that most applications can be developed
28:34

28:34 very inexpensively. We will in the next
28:38

28:38 few years cross over those thresholds
28:39

28:39 for accounting, legal work, teles sales,
28:43

28:43 teles support where 24 hours a day in in
28:47

28:47 every language with infinite trivia

28:50

28:50 capability and no urgency, you know. So

28:53

28:53 you look at the AI nurse and there's

28:56

28:56 several companies. They have perfect

28:58

28:58 memory. They don't hurry.

28:59

28:59 >> Be like a nurse you would call on the

29:01

29:01 phone or talk through a chat interface.

29:03

29:03 Yeah.

29:04

29:04 >> Uh far preference, you know, you go to

29:06

29:06 San Francisco and ask people, would you

29:08

29:08 rather ride in a Whimo or rather ride

29:09

29:09 with the human driver? Go ask people in

29:12

29:12 the UK who use limbic for mental health

29:14

29:14 support. uh you know so the notion that

29:19

29:19 just the market demand preference for
29:22

29:22 driving or the nurse uh you know or even
29:27

29:27 the the person on the phone will favor
29:30

29:30 humans that's a quality threshold which
29:34

29:34 will be passed through you know if if
29:36

29:36 say insurance companies are still human
29:38

29:38 doing doing humans through claims
29:40

29:40 medical claims which is a very AI uh
29:44

29:44 capable task ask then a competitor who
29:47

29:47 has very few human employees will come
29:49

29:49 in and uh change the pricing model for
29:52

29:52 that industry. So the two
29:54

29:54 counterarguments I've heard people make
29:56

29:56 on this one is that you have a Jevans
29:59

29:59 paradox effect where the cheaper more
30:03

30:03 widespread availability of this kind of
30:05

30:05 intelligence leads to a massive increase
30:09

30:09 in the demand for this kind of thing. So
30:11

30:11 yes, you have many more AI chatbot
30:13

30:13 nurses and that leads to more people
30:14

30:14 being sent to the hospital, being sent
30:16

30:16 to the doctor where real nurses take
30:18

30:18 care of them. Or you have many more AI
30:20

30:20 coders. So maybe my small podcast team,
30:22

30:22 which wouldn't have had a software
30:23

30:23 engineer before, now has one because
30:25

30:25 it's like they run a team of coders and
30:27

30:27 we can build products we never thought
30:28

30:28 of before. Uh this is the most common
30:31

30:31 answer I hear to this that yes, AI will
30:33

30:33 destroy jobs. Yes, it is making a
30:36

30:36 humanprovided resource uh much cheaper.
30:38

30:38 But because it's going to expand the
30:41

30:41 demand so much, it will sort of work
30:44

30:44 itself out, create jobs in other areas,
30:46

30:46 create new demand, and this is how past
30:48

30:48 technologies have gone and so we
30:50

30:50 shouldn't worry too much about this. How
30:52

30:52 do you see that?
30:54

30:54 >> Well, Jevans is just a referral to the

30:56

30:56 fact that parts of the economy are

30:58

30:58 subject to demand elasticity.

31:00

31:00 And yes, in the case of software, if

31:03

31:03 you're say three times as fast and

31:05

31:05 there's still some role that only humans

31:07

31:07 can perform, then as you lower the cost,

31:10

31:10 you induce demand. And so it's fair to

31:13

31:13 say that the equilibrium today for so

31:15

31:15 far for software is not a loss of

31:18

31:18 employment. When you invent radial tires

31:21

31:21 that last four times as long, for some

31:24

31:24 weird reason, people didn't drive four

31:25

31:25 times as much. uh and factories that

31:29

31:29 make tires employ a quarter as many
31:32

31:32 people. When you replace people in
31:33

31:33 Amazon warehouses with robots, people
31:37

31:37 don't buy more because of that. So, you
31:41

31:41 know, anybody who's numeric can say to
31:43

31:43 themselves, what portion of the economy
31:45

31:45 is is subject to man elasticity and what
31:48

31:48 are those tasks that will still be human
31:52

31:52 necessary? As soon as you complete the
31:54

31:54 entire task, it doesn't matter that
31:56

31:56 there's demand elasticity. That goes
31:58

31:58 into the token budget. [clears throat]
31:60

31:60 It doesn't go into the human salary
32:02

32:02 budget. The cost for some of these
32:04

32:04 things is so much less than the cost of
32:07

32:07 the human labor. And so the as you cross
32:11

32:11 reliability thresholds both with white
32:14

32:14 collar and human robots,
32:17

32:17 you destroy jobs and you leave no high
32:20

32:20 ground. That's, you know, innovation in
32:23

32:23 the past. You have a tractor, fine. You
32:25

32:25 know, let's build Disneyland and employ
32:27

32:27 a lot of people there. You don't have
32:30

32:30 that in the broad economy. You know,
32:33

32:33 they people who say they'll be net
32:35

32:35 additional jobs. I don't understand what
32:37

32:37 they're thinking. They must not

32:38

32:38 understand the piece of improvement

32:40

32:40 we're on that, you know, the

32:42

32:42 reliability. You wrote a column I think

32:44

32:44 was in May that I looked at like what

32:47

32:47 what you disag what is this unique

32:50

32:50 humanist thing that you think uh uh you

32:56

32:56 know in every category where AI has come

32:58

32:58 along the preference for the AI is very

33:02

33:02 very strong

33:02

33:02 >> well so that com is based on so Jeban's

33:04

33:04 paradox and then the other which is an

33:06

33:06 argument from Alex Immus who I believe

33:08

33:08 is now at one of the labs as an

33:10

33:10 economist is that you will have this
33:12

33:12 sort of explosion in the relational
33:14

33:14 sector that one thing that happens when
33:16

33:16 people get wealthier and you'd probably
33:18

33:18 know about this is that they all of a
33:19

33:19 sudden get a lot more human help, right?
33:22

33:22 They have personal trainers and chefs
33:24

33:24 and there's a lot more No, I don't know.
33:27

33:27 Royalty used to have a lot of uh human
33:29

33:29 help. You had upstairs, downstairs, you
33:31

33:31 know, all those maids and people and you
33:34

33:34 used to have a human who helped you get
33:35

33:35 dressed. Believe me, the labor intensity
33:37

33:37 of wealth is down super direct

33:39

33:39 >> from what it used to be. So you don't

33:41

33:41 >> Jeans paradox. Name a bluecollar

33:43

33:43 profession that's subject to gem paradox

33:46

33:46 or do you not care about blue collar?

33:48

33:48 >> I do care about blue collar. So the

33:49

33:49 question is anything in the blue collar

33:51

33:51 realm subject to that.

33:52

33:52 >> So I I'm not I I don't have the the

33:54

33:54 strongest view on this but here's I

33:56

33:56 think the argument that I would make or

33:57

33:57 that I've heard made

33:58

33:58 >> is that if you look at something like

34:01

34:01 manufacturing

34:02

34:02 we've not had a Jevans paradox in

34:05

34:05 manufacturing employment in America.

34:06

34:06 Right? We exploded how much we actually

34:08

34:08 create, but a lot of the jobs went

34:10

34:10 offshore. Some of the jobs went to

34:12

34:12 automation, but we don't have less total

34:14

34:14 employment in America than we did in

34:16

34:16 1960. And the reason is people moved

34:18

34:18 into service sector jobs. That the

34:20

34:20 composition of jobs across the economy

34:22

34:22 changed.

34:23

34:23 >> Yeah. Human cognition became the scarce

34:26

34:26 element. It's always about scarce

34:27

34:27 elements. Once you replace human
34:29

34:29 cognition as a scarce element in fact
34:32

34:32 you blow it away in terms of working 24
34:34

34:34 hours a day reading more knowing every
34:37

34:37 language there is no scarcity that
34:39

34:39 you're moving up.
34:40

34:40 >> So you feel there is no scarcity that
34:42

34:42 will be left for human
34:43

34:43 >> name the scarcity
34:44

34:44 >> the Alex would say it's something like
34:47

34:47 relational sector jobs but you don't
34:48

34:48 believe
34:49

34:49 >> of the current jobs you mean like my
34:51

34:51 relationship with a cab driver or my
34:53

34:53 relationship with that nurse where I'd
34:55

34:55 rather have hypocratic AI call me. I
34:57

34:57 think the question here is actually do
34:59

34:59 you end up preferring hypocratic AI or
35:01

35:01 do you actually want your therapist to
35:03

35:03 be a human being because after a little
35:05

35:05 while there's something thin about
35:07

35:07 telling your problems to a computer.
35:10

35:10 >> Well, the the market you can gather
35:12

35:12 market data if that's at all
35:14

35:14 interesting. You know, I suppose telling
35:16

35:16 the 55year-old truck driver you're going
35:18

35:18 to go do some relational thing. Uh you

35:22

35:22 have a program for that?

35:23

35:23 >> I don't have a program for it. And I

35:24

35:24 think the question you actually run into

35:26

35:26 very quickly is speed of transition.

35:28

35:28 Even if you believed some of this and I

35:30

35:30 said this in the column too that I don't

35:32

35:32 see how you will handle the speed of

35:34

35:34 transition that our actual experience

35:37

35:37 with fast transition of jobs is that

35:41

35:41 people lost their jobs lose out. Now you

35:43

35:43 might over the entire economy it doesn't

35:45

35:45 look that different but say the China

35:47

35:47 shock did lead to a lot of ruined

35:48

35:48 communities. So your answer to this and
35:50

35:50 it leads to a very concrete idea for you
35:53

35:53 which is that we should tax the use of
35:55

35:55 AI. Why and how?
35:57

35:57 >> Well, let's say you know workers pay in
35:60

36:00 on a pay as you go basis
36:03

36:03 uh into the pension social security
36:06

36:06 fund. So active workers are supporting
36:10

36:10 retired workers. If you've let go of a
36:13

36:13 worker and hire a robot to do that job,
36:17

36:17 why are you so insenting the trade-off
36:20

36:20 against the human labor that you don't
36:22

36:22 ask the robot to also pay into the
36:26

36:26 pension fund? What is it? you know, is
36:28

36:28 that the pro robot union has gotten the
36:30

36:30 tax laws to say no, if you're not flesh,
36:33

36:33 then let pensions go bankrupt, which you
36:37

36:37 know, they're on a on a path to do even
36:39

36:39 without robot replacement.
36:41

36:41 >> So, how would we do that? You say uh you
36:45

36:45 define a unit of labor and say uh
36:48

36:48 independent of whether that's delivered
36:50

36:50 by a human or robot, you are paying the
36:53

36:53 same FICA tax that a human worker would
36:57

36:57 pay. Now that alone is not enough, but
37:00

37:00 you know why should we be so favorable
37:03

37:03 to taking away that job? The current tax

37:06

37:06 structure is very much as though people
37:09

37:09 who have capital are what really counts
37:12

37:12 and labor is the most disadvantage input
37:15

37:15 to the economy. Society gets to decide
37:18

37:18 just because the economic signals say
37:23

37:23 that it'd be lower cost to use an AI
37:27

37:27 doesn't mean society has to do that.
37:29

37:29 >> The other one you propose is human
37:31

37:31 reserve jobs. Tell me what you mean by
37:33

37:33 that. So there's a question particularly
37:35

37:35 as you get the impacts on the bluecollar
37:39

37:39 stuff which that's very sharp because
37:42

37:42 you know when the humanoid robots pass a
37:44

37:44 certain threshold they are very general
37:47

37:47 purpose but the idea here is that you
37:50

37:50 would decide in advance that things like
37:52

37:52 child care or elder care or some portion
37:55

37:55 of medical care some portion of
37:57

37:57 education although you could have some
37:59

37:59 AI enhancement that you would maintain
38:03

38:03 the employment and call those those
38:05

38:05 things human reserved.
38:07

38:07 >> I found this vision not totally
38:12

38:12 unconvincing but [clears throat]
38:13

38:13 chilling. I I think the way you put it
38:15

38:15 in the piece was that this feels a
38:18

38:18 little bit to you like nature reserves,
38:20

38:20 places where we could put buildings and
38:21

38:21 roads but we choose not to because the
38:23

38:23 loss would be so great. And I think it
38:26

38:26 gets to this question that a lot of
38:28

38:28 people have that on one out of every two
38:30

38:30 days I have, which is if this is what it
38:33

38:33 looks like, like why do it? The Gates
38:37

38:37 Foundation announced on Monday a 5-year
38:39

38:39 goal for an estimated 3.4 billion people
38:42

38:42 who speak languages currently under
38:44

38:44 represented in today's AI models to be
38:47

38:47 able to use AI tools in their own
38:48

38:48 language and voice. And if we are

38:51

38:51 looking at this sort of job apocalypse,
38:53

38:53 this level of risk that it's like to
38:55

38:55 have that on the one hand and then on
38:57

38:57 the other hand like our goal is for more
38:58

38:58 people to use AI. Tell me why
39:02

39:02 the there are many good things that AI
39:05

39:05 does and that's almost all of the
39:07

39:07 foundation's work is in taking what the
39:10

39:10 market would not do which is take AI to
39:14

39:14 the poorest in the world and help them
39:16

39:16 with education and health and
39:18

39:18 agriculture. So if you have a a woman
39:22

39:22 say in Nigeria who speaks Yoruba, uh
39:25

39:25 she's in a remote community, there's no
39:27

39:27 doctor there. So when she says she's
39:30

39:30 bleeding, AI doesn't get that right.
39:33

39:33 We're we're spoiled. We speak English.
39:34

39:34 And that's the thing that Yoruba is 10
39:37

39:37 times worse than is is English, which is
39:41

39:41 significantly the best. There's about 10
39:42

39:42 other languages that are fairly close.
39:44

39:44 So the empowerment for that woman to be
39:48

39:48 able to talk about her medical problem
39:51

39:51 and get advice in the middle of the
39:54

39:54 night just by having uh a smartphone
39:57

39:57 with the data connection. That's what
39:59

39:59 we're trying to enable. Okay. But I get
40:02

40:02 that. It's just the the picture you just
40:05

40:05 painted of the job apocalypse. As bad as
40:08

40:08 it might be here and it will be bad. We
40:11

40:11 do have a fair amount of money. We have
40:13

40:13 the AI companies. weakened tax
40:15

40:15 companies, put that into redistribution.
40:17

40:17 But is this not going to wipe out the
40:20

40:20 ladder of mobility, the ladder of
40:24

40:24 development for a lot of these
40:26

40:26 countries? I mean, doesn't this make AI
40:29

40:29 a tremendous economic threat if you're
40:32

40:32 the Philippines?
40:33

40:33 >> You're saying that our allowing the eye

40:36

40:36 to understand your Reuben help that
40:37

40:37 woman who's bleeding?
40:38

40:38 >> I'm saying why is a bad thing? I am
40:40

40:40 saying that AI sounds in this telling
40:43

40:43 like a bad thing that the net net like
40:47

40:47 yes
40:47

40:47 >> you're supporting the the world's
40:49

40:49 languages how can be a bad thing
40:51

40:51 >> but I think you understand what I'm
40:52

40:52 saying
40:52

40:52 >> no I I I
40:54

40:54 >> that the job pressure you're putting
40:55

40:55 forward is very scary
40:56

40:56 >> to accelerate the good stuff AI does and
40:60

40:60 I'm trying to minimize
41:00

41:00 >> so what do we do about jobs in the
41:02

41:02 global south
41:03

41:03 >> well treating the global south like one
41:07

41:07 uniform thing is doesn't allow you to
41:10

41:10 have any picture of what it's like to
41:13

41:13 live there. There are middle-income
41:15

41:15 countries like China, Brazil, Vietnam,
41:18

41:18 Indonesia, you know, where their
41:20

41:20 economies are growing and, you know,
41:22

41:22 their childhood death rate is within a
41:25

41:25 factor of three of the US. Then there
41:27

41:27 are low-income countries where 15 times
41:29

41:29 as many children die before the age of
41:32

41:32 five. And you don't have doctors. You
41:34

41:34 live your entire life, you never meet a
41:37

41:37 doctor. you try to figure out what seeds
41:39

41:39 to plant. You never have anybody advise
41:42

41:42 you what to do. And so, uh, you know, we
41:45

41:45 need a little more nuance. In low-income
41:48

41:48 countries, the AI will overwhelmingly be
41:51

41:51 a good thing and it should be pushed
41:54

41:54 forward as quickly as
41:56

41:56 >> So, is your argument in some ways
41:57

41:57 actually it's a better trade for
41:58

41:58 low-income countries that it has more

42:01

42:01 benefit maybe than it does?

42:02

42:02 >> Absolutely. The the human basics were

42:04

42:04 not meeting today. If these people lived

42:06

42:06 on your street, you would open your

42:08

42:08 wallet. You'd be outraged. I mean, these

42:10

42:10 kids are dying. These kids are

42:12

42:12 malnourished. So, the inequity is

42:14

42:14 allowed to exist because of the

42:15

42:15 distance. And AI is a tool for good in

42:19

42:19 terms of helping those people in by far

42:22

42:22 the greatest need with very basic human

42:25

42:25 problems.

42:25

42:25 >> Do you think it has as much effect on

42:27

42:27 the job markets in some of the poor
42:28

42:28 countries that we're talking about?
42:30

42:30 >> Not in the same time frame. No. But over
42:32

42:32 time,
42:33

42:33 >> well, if you're t see you're you're you
42:36

42:36 probably spend more time in middle inome
42:38

42:38 countries than in low-income countries.
42:40

42:40 So your image is like [clears throat]
42:41

42:41 Philippines, India and all of those. Uh
42:44

42:44 yes, in those countries they will see
42:47

42:47 the jobs effect after the rich countries
42:50

42:50 and then eventually even the low-income
42:53

42:53 countries.
42:53

42:53 >> But I guess this gets to the broad

42:54

42:54 engagement question. I I don't have a

42:57

42:57 good crystal ball on this. I find the

42:59

42:59 range of outcomes terrifyingly wide.

43:01

43:01 This range it seems to run from

43:05

43:05 massive material abundance and

43:07

43:07 elimination of want to actually doesn't

43:09

43:09 change all that much to human extinction

43:12

43:12 is a pretty wide range of outcomes to

43:14

43:14 consider. What I hear you saying to me

43:16

43:16 and you should tell me if I'm getting

43:17

43:17 part of your position wrong here. What I

43:19

43:19 hear you saying to me and you know

43:21

43:21 convincingly and [clears throat]

43:22

43:22 forcefully is that the AI job apocalypse

43:25

43:25 is a very real thing. that mass

43:28

43:28 displacement of workers with no real

43:30

43:30 answer to that is a very real thing.

43:31

43:31 When you talk about broad societal

43:33

43:33 engagement, I think most Americans, most

43:37

43:37 people in most places, if they heard

43:38

43:38 that and they were convinced of it,

43:40

43:40 right? And polls show most people think

43:43

43:43 AI is going to take jobs and not create

43:44

43:44 them. But they would say actually just

43:46

43:46 stop, right? If if what you are going to

43:49

43:49 do is make it unclear how me or my

43:51

43:51 family or my children or my friends will
43:53

43:53 have a job, say in America, like please
43:56

43:56 just don't like let's just stop for now.
43:59

43:59 And so you actually seem to me to have a
44:02

44:02 stronger negative perspective on the
44:04

44:04 jobs question than a lot of the people I
44:05

44:05 talked to even in the labs. You're
44:07

44:07 shaking your head at me. I do talk to
44:08

44:08 people. I'm not I'm not coming from
44:09

44:09 nowhere on this and definitely than a
44:11

44:11 lot of the economists I talked to. So,
44:14

44:14 for whom then is AI a good trade if you
44:17

44:17 believe the job effect is going to be so
44:19

44:19 ruinous for most people?

44:21

44:21 >> It's funny that when I was telling this

44:23

44:23 to the person who some people consider

44:26

44:26 the lead economist looking at AI, he

44:29

44:29 said, "How can Mechanical Turk still

44:31

44:31 exist? How Amazon run that if AI is good

44:33

44:33 as you said?" And I said, "It won't." A

44:36

44:36 week later, it was completely shut down.

44:39

44:39 Uh and yes, the susion of knowledge from

44:45

44:45 you have to say, isn't it pretty

44:47

44:47 stunning that the more you know, the

44:49

44:49 more concerned you are? Uh you know,

44:52

44:52 take Daario. Daario's spoke out about

44:55

44:55 job impacts. Now, he's a bit more

44:58

44:58 guarded. Uh Hinton, I agree, went too
45:02

45:02 far and he said, "Okay, it's coming
45:04

45:04 tomorrow." That there'll be less
45:06

45:06 radiologists. And of course there are
45:07

45:07 more today. So we have some of that
45:10

45:10 taking place and we have the analogies
45:12

45:12 with the past where people are saying
45:14

45:14 well this is like the PC not like
45:17

45:17 evolutionary history. This is like
45:19

45:19 evolutionary history. This is like the
45:22

45:22 aliens really are here and you know have
45:25

45:25 come they didn't happen to spacecraft.
45:27

45:27 They were created in laboratories. But
45:30

45:30 uh that's the kind of thing we're
45:32

45:32 dealing with in if we retain control uh
45:38

45:38 then eventually you do get not to
45:40

45:40 overuse the word that you've used get to
45:43

45:43 this super abundance uh period and there
45:47

45:47 you have deep almost philosophical
45:50

45:50 religious issues of if you don't have
45:52

45:52 the shortages that we've organized
45:54

45:54 society around uh you know why you
45:57

45:57 should learn and and how you find
46:00

46:00 purpose then you have deep philosophical
46:03

46:03 religious problems but you do not have
46:06

46:06 uh dying of malaria don't have enough
46:09

46:09 food don't have access to a doctor

46:11

46:11 problems and so

46:14

46:14 if we get to that a younger generation

46:16

46:16 will figure out okay how do we live how

46:19

46:19 do we spend time that's a very different

46:21

46:21 world than the way have today I think

46:23

46:23 what this generation has to do is make

46:26

46:26 sure we get through that with humanity

46:30

46:30 still in control and with the

46:32

46:32 disruption, the number of bioterror

46:34

46:34 events or the number of people whose

46:37

46:37 lives are damaged by job loss, we need

46:39

46:39 to minimize that over what's probably a

46:43

46:43 20 year plus transition period. So what

46:46

46:46 when when you say because this is

46:48

46:48 something you say at the beginning of

46:48

46:48 that essay that this could be and I'm

46:50

46:50 paraphrasing the most powerful driver of

46:53

46:53 inequality or of equality where what are

46:58

46:58 the highest leverage for you good things

47:01

47:01 that can come out of this I can imagine

47:02

47:02 somebody listening to our conversation

47:03

47:03 you're thinking why the hell would we do

47:06

47:06 this with this set of risks yeah that's

47:09

47:09 where the timing thing is troubling

47:13

47:13 you know whichever definition of

47:15

47:15 abundance you're using. You want to say,

47:17

47:17 "Oh, my health bill was less than I
47:18

47:18 expected. Oh, I was buying a new house
47:21

47:21 and it was less than I expected or I was
47:23

47:23 renting and they seem to be lowering the
47:25

47:25 rent. Uh, you know, my my electricity
47:29

47:29 that's but AI
47:31

47:31 uh, you know, because of the efficiency
47:36

47:36 and invention that comes with it. That's
47:40

47:40 you know what over the 20-year period
47:42

47:42 you do get kind of mindblowing advances
47:46

47:46 in things that people can relate to the
47:49

47:49 the cost of their food, shelter,
47:51

47:51 education because those are highly
47:54

47:54 regulated areas. So the good stuff if
47:57

47:57 we're not careful arrives more slowly
48:02

48:02 than the biocyber risk, the psychosocial
48:06

48:06 risk and the jobs risk. those things in
48:09

48:09 the next 5 years are very significant
48:13

48:13 and so people will decide although you
48:16

48:16 know it's hard for a single country to
48:18

48:18 check out people will decide whether to
48:20

48:20 slow down stop get rid of AI uh and you
48:25

48:25 know yes it will have political
48:27

48:27 difficulties stopping data centers isn't
48:30

48:30 going to slow this thing one iota so
48:32

48:32 anybody who's against data centers
48:34

48:34 because they think that'll slow down AI

48:36

48:36 that's a waste of effort effort the data

48:38

48:38 centers are going to get built

48:40

48:40 somewhere. Uh so the broad view of the

48:42

48:42 the technology that has to be expressed

48:45

48:45 in in in some other form. I want to get

48:47

48:47 at something you're saying here because

48:48

48:48 something that that I've thought about

48:50

48:50 too which is that there are a lot of

48:52

48:52 rate limiters a lot of weak links in the

48:54

48:54 chain when you try to take even in an

48:57

48:57 optimistic view AI advances that then

48:60

48:60 have to be built in the material world

49:02

49:02 right you have an acceleration of you

49:04

49:04 know good candidates for pharmaceutical
49:05

49:05 development but you still need to find
49:07

49:07 you know rats to test on monkeys to test
49:09

49:09 on human you know trials you have slow
49:11

49:11 regulatory agencies etc and then you
49:14

49:14 have this tremendous acceleration of
49:16

49:16 intelligence
49:17

49:17 that is for lack of a better term native
49:20

49:20 to the digital world and is acting there
49:22

49:22 in a kind of constant way. And you I
49:26

49:26 mean right now you see this right you
49:28

49:28 can accelerate AI with very little
49:29

49:29 regulatory overhang but you cannot get
49:32

49:32 you know the open AI parking lot covered

49:34

49:34 in solar panels without permits and

49:36

49:36 hearings. And so it I think you have a

49:40

49:40 very high chance of getting into a very

49:42

49:42 weird both like economic and social

49:44

49:44 world where the digital world is

49:47

49:47 spinning into this other thing. But most

49:50

49:50 of the things human beings need

49:52

49:52 ultimately come out physically like we

49:54

49:54 need shelter and we need food and we

49:55

49:55 need you know all all this and they run

49:57

49:57 through institutions. And so a lot of

49:59

49:59 the ways it could actually make our

50:02

50:02 lives better are sort of limited by

50:04

50:04 normal human factors, but a lot of the
50:05

50:05 ways it could make them worse or weirder
50:08

50:08 are not.
50:10

50:10 >> Well, we definitely need to look at the
50:12

50:12 things that slow down the good stuff,
50:15

50:15 you know. So, for example, because the
50:17

50:17 Gates Foundation's nonprofit, we can
50:20

50:20 work with regulators on how they use AI
50:23

50:23 to do their job. And you know whether
50:25

50:25 it's organoids or uh biological models
50:31

50:31 we can speed up that regulatory piece
50:34

50:34 and there's some countries uh that are
50:37

50:37 very much in engaged in that type of
50:39

50:39 acceleration.

50:41

50:41 The one application that's going full

50:43

50:43 speed ahead and it kind of amazed me is

50:45

50:45 is the agricultural one because they're

50:49

50:49 getting better weather data to these

50:51

50:51 farmers in Africa. You know, we have

50:53

50:53 over a million farmers in India already

50:55

50:55 using the system. So, it tells them what

50:57

50:57 crop disease they have, what fertilizer

50:59

50:59 to use, what varieties to plant. We've

51:02

51:02 even put a bunch of services they can uh

51:05

51:05 take advantage of on there. Uh that one

51:07

51:07 is going full speed ahead. The health

51:12

51:12 and education, you could say, you know,
51:14

51:14 will it be the rich countries, the
51:16

51:16 middle income countries, or the
51:18

51:18 low-income countries that get there
51:20

51:20 first? uh because some of the the
51:23

51:23 barriers are different for low-income
51:25

51:25 countries. It's not regulatory. It's,
51:27

51:27 you know, is anybody providing the
51:28

51:28 tokens? Do they have the smartphone and
51:30

51:30 the connectivity which are the basics
51:32

51:32 and the the language thing that we've
51:34

51:34 talked about. So, they'll proceed at
51:36

51:36 different paces. Ideally, we'll learn
51:38

51:38 from each other. I mean, China actually

51:40

51:40 banned young people uh having social
51:42

51:42 relationships with AI. You know, they do
51:45

51:45 they know something we don't know. You
51:47

51:47 know, can we look at outcomes of
51:49

51:49 different experiments? Um, so there's
51:52

51:52 going to have to be a lot of that
51:53

51:53 learning, but you're right, the if the
51:56

51:56 benefits aren't coming quickly, then the
51:59

51:59 permission to operate, particularly if
52:01

52:01 we don't put on the safeguards for the
52:04

52:04 SC cyber attacks or the bio attacks. Uh,
52:08

52:08 this is going to be a very [music] tough
52:09

52:09 issue to stick up for.
52:22

52:22 When I was preparing for this episode
52:23

52:23 and telling people I was interviewing
52:25

52:25 you, one of the big space of skepticism
52:28

52:28 people now have is around what they've
52:30

52:30 heard with you and Jeffrey Epstein. So,
52:32

52:32 I've looked at your house testimony, but
52:35

52:35 what do you say to people who have lost
52:36

52:36 some faith in you and just see you
52:39

52:39 differently just now linking sort of
52:41

52:41 your name and his?
52:44

52:44 you know, they should read what I said
52:47

52:47 to Congress. Uh,
52:49

52:49 I got to answer all the questions. Uh, I
52:55

52:55 in trying to raise money for global
52:57

52:57 health thought that Epstein could
52:59

52:59 connect me uh because of his
53:01

53:01 relationship with them and I had me
53:04

53:04 >> relationship with whom?
53:05

53:05 >> With billionaires.
53:07

53:07 >> But you're a billion I mean you're the
53:09

53:09 billionaire on some level. I'm giving
53:10

53:10 all my money away. And I I spend a lot
53:13

53:13 of time trying to raise money for for
53:15

53:15 global health.
53:17

53:17 >> And so you felt that he had connections
53:18

53:18 that would be of value to you in that
53:21

53:21 >> to raise money for global health. Yes.

53:23

53:23 That's that's what that I mean did you
53:25

53:25 read it? I did read it. That's all we
53:27

53:27 that's what we met about. I mean other
53:29

53:29 than okay, there was a dinner with Larry
53:30

53:30 Summers where we talked about the
53:32

53:32 economy. The whole discussion was is
53:35

53:35 there a chance here to help out?
53:38

53:38 spending time with him, you know, was
53:39

53:39 clearly a mistake. I've changed my bar
53:41

53:41 for even somebody in a non-paid
53:44

53:44 intermediate role. I'll never take any
53:46

53:46 risk on that again because our work is
53:48

53:48 is very reputation sensitive. The hole
53:53

53:53 at the center of that story for me is
53:55

53:55 how this guy was so compelling to so
53:60

53:60 many very smart, very wealthy people. I
54:03

54:03 mean, there are a lot of people who want
54:04

54:04 your attention and don't get it. people
54:06

54:06 want Larry Summer's attention and don't
54:07

54:07 get it. That there's something about
54:11

54:11 Epstein's
54:12

54:12 charisma or the way he presented himself
54:15

54:15 that drew people into at least thinking
54:17

54:17 he could be very useful to them.
54:19

54:19 What was that? What like I've read his
54:22

54:22 emails. He does not seem like a pleasant
54:23

54:23 person to email with. What was the kind
54:27

54:27 of factor about him that allowed him to
54:30

54:30 weave this web that you and so many
54:32

54:32 other people were in?
54:33

54:33 >> Well, I wasn't in a web. I was in
54:35

54:35 discussions about raising money for
54:37

54:37 global health. I didn't go to any island
54:39

54:39 or meet any women anyway. Don't call it
54:41

54:41 a web. Uh the
54:45

54:45 how he performed the bootstrap of, you
54:48

54:48 know, owning the fanciest house I've
54:50

54:50 ever seen in New York City and having
54:53

54:53 Larry Summers and the number two guy JP
54:55

54:55 Morgan there. Uh you know, so that was

54:59

54:59 an interesting dinner.

55:02

55:02 uh even if if Jeffrey never said a word.

55:04

55:04 I don't know how that bootstrap take

55:06

55:06 took place. But there were

55:11

55:11 lots of billionaires that he was

55:13

55:13 involved in the moment where you kind of

55:17

55:17 decide okay uh what am I doing with my

55:21

55:21 wealth? My how do I minimize taxes? What

55:24

55:24 do I do with my family? That type of

55:27

55:27 thing. And so, you know, that's I don't

55:30

55:30 think anybody else was drawn in for that

55:33

55:33 reason. My, you know, case is kind of a

55:35

55:35 unique case in that he said, "Oh, you'll

55:38

55:38 have more money than you'll know what to
55:39

55:39 do with." And eventually, I insisted
55:42

55:42 that he take me around to see
55:44

55:44 billionaires. He had me meet with five.
55:46

55:46 Uh, turned out, uh, none of them had a
55:49

55:49 near-term intent, and I ended the
55:52

55:52 relationship with them within a month of
55:54

55:54 that. And and so the the thing he was
55:57

55:57 able to do because that's interesting to
55:58

55:58 me is like even for you he'd walk into
55:60

55:60 the house it was that fancy that people
56:02

56:02 around it gave him credibility and so
56:05

56:05 that he was able to put on a show such
56:07

56:07 that what he was able to offer was of
56:10

56:10 value. There's all this kind of social
56:12

56:12 capital and wealth flowing around.
56:13

56:13 >> Some people socialized with him. I did
56:15

56:15 not other than to my surprise after
56:18

56:18 dinner uh magician coming in David
56:20

56:20 Blaine. I didn't spend one minute
56:23

56:23 socializing with them. He offered, you
56:25

56:25 know, come to the island or, you know,
56:27

56:27 come to some show in Paris. And I said,
56:29

56:29 I would not choose to do that. That
56:31

56:31 would not be a good idea. So then I want
56:34

56:34 to widen back out to the Gates
56:35

56:35 Foundation. You said you're going to

56:37

56:37 spend down its money, \$200 billion, by
56:39

56:39 2045. Given everything we've talked
56:41

56:41 about here with AI, how is your vision
56:44

56:44 for how that spending plays out changed?
56:46

56:46 >> I mean, there's two dramatic things that
56:50

56:50 have happened.
56:52

56:52 one is the incredible AI capabilities
56:55

56:55 which continues to be an exponential and
56:59

56:59 the second is the reduction in overall
57:02

57:02 generosity towards the poorest in the
57:05

57:05 world. So I have one piece of of
57:09

57:09 enablement
57:11

57:11 which is we'll discover drugs faster,
57:14

57:14 we'll [clears throat] discover new seeds
57:15

57:15 faster, and that we can talk to those
57:18

57:18 farmers and get them to buy the right
57:20

57:20 seed or talk to the person living with
57:22

57:22 HIV and help them seek the care they
57:26

57:26 need. So AI can be a big enabler for our
57:29

57:29 goals, including malaria eradication,
57:32

57:32 polio eradication, cutting childhood
57:34

57:34 death in half again.
57:36

57:36 But now the amount of resources
57:39

57:39 available is way less than you know sort
57:44

57:44 of a golden rule view of the world would
57:46

57:46 suggest it should be. There is from the
57:49

57:49 foundations connected to anthropic and
57:50

57:50 open AAI this sense of a an amount of
57:53

57:53 nonprofit money coming online in the
57:55

57:55 next couple of years that will dwarf
57:57

57:57 sort of any moment in philanthropy
57:59

57:59 before it. I I know people in that world
58:01

58:01 that the sense of this being something
58:04

58:04 very unusual is is is very present. Does
58:07

58:07 that change what's possible here?
58:08

58:08 >> Be numeric. Uh take the cuts in
58:11

58:11 overseas.
58:12

58:12 >> I'm not saying it I'm not saying it
58:14

58:14 answers the cuts in over we save lives
58:16

58:16 for \$1,000 per year. And so getting less

58:19

58:19 money from philanthropist net uh doesn't
58:23

58:23 save those lives. uh you know the amount
58:26

58:26 of money to save lives at \$1,000 per
58:31

58:31 life saved will be dramatically less. So
58:33

58:33 when I've looked at the way your
58:35

58:35 foundation is spending on AI and tell me
58:37

58:37 if I have this wrong. It's about 40%
58:38

58:38 education, 40% health, 10ish%
58:41

58:41 agriculture and 10ish% institutions. Do
58:44

58:44 I have that?
58:45

58:45 >> I mean it's a it's a little more in
58:48

58:48 [clears throat] agriculture. Almost
58:49

58:49 nothing we do because we're like a
58:51

58:51 pharmaceutical company inventing new
58:53

58:53 drugs. And do you call that AI or not
58:55

58:55 calling that AI? We didn't put any of
58:57

58:57 that in our billiondoll
59:00

59:00 uh commitment we announced. We just put
59:02

59:02 in AI specific things like understanding
59:07

59:07 African languages. So tell me a bit
59:09

59:09 about your views on AI and education.
59:12

59:12 This is a place where I've seen a lot of
59:13

59:13 studies now and they have very very
59:15

59:15 mixed results depending on how they're
59:17

59:17 used. So where do you think it is
59:19

59:19 valuable and where is it at risk?
59:22

59:22 Well, the contrast between the ambitious
59:25

59:25 goals that the Gates Foundation set in
59:28

59:28 the year 2000, our goals for global
59:30

59:30 health and our goals for education. In
59:33

59:33 global health, we thought, wow, African
59:35

59:35 governments,
59:36

59:36 uh, traditional beliefs, uh, you know,
59:39

59:39 are we going to have any impact at all?
59:43

59:43 Much to my surprise, our work on global
59:46

59:46 health together with, you know, we
59:48

59:48 create with others the global alliance
59:50

59:50 for vaccines, we create the global fund,
59:53

59:53 President Bush does PEPAR, you know, a
59:55

59:55 ton of things happen including, uh,

59:58

59:58 primarily vaccines, uh,
60:01

60:01 roirus, numacus, and we cut child to
60:04

60:04 death more than in half from over 10
60:07

60:07 million a year to under 5 million year.
60:09

60:09 So we more than exceed any goal we would
60:12

60:12 have set for the field in global health
60:14

60:14 in education. We thought we could just
60:17

60:17 go see what really great teachers do,
60:20

60:20 videotape that, understand it, create a
60:23

60:23 feedback system for teachers to hear
60:26

60:26 that, uh improve their practice, you
60:29

60:29 know, constant learning on that and that
60:32

60:32 we could make education a lot better. As
60:35

60:35 you've just seen uh from the latest
60:37

60:37 numbers, uh kids are learning less today
60:42

60:42 in most rich countries including in the
60:45

60:45 United States than they learned 10 years
60:47

60:47 ago, 20 years ago. Uh so people start
60:51

60:51 appropriately with a very high degree of
60:53

60:53 >> Before I go to AI, tell me why you think
60:55

60:55 that those numbers have been very very
60:56

60:56 striking. Um why do you think that is?
60:58

60:58 Why are we seeing learning loss in rich
60:60

60:60 countries that now I think it's you got
61:02

61:02 to have a not you but one has to have a
61:04

61:04 broader answer than the pandemic.
61:07

61:07 >> Well, it's definitely we are post

61:09

61:09 pandemic and people debate is it the use

61:12

61:12 of cell phones? Uh is it you know

61:15

61:15 something to do with affluence? Uh

61:17

61:17 overall there are some outliers. uh

61:22

61:22 England uh actually is counter trend and

61:25

61:25 they went back to basics in terms of how

61:27

61:27 they teach reading as well. And so you

61:30

61:30 know high expectations is winning out

61:32

61:32 over just you know uh free form uh

61:37

61:37 creativity but you know there's a

61:40

61:40 there's something clearly large in those

61:43

61:43 statistics. It's it's very concerning

61:46

61:46 and so you can almost say why do we

61:49

61:49 continue to do education? Well,
61:51

61:51 education is key. I'm stubborn. Uh we've
61:54

61:54 got AI now. Uh and we have people where
61:60

61:60 the AI is playing a very specific role
62:03

62:03 of each student uh at the end of the
62:07

62:07 day. Uh this is here in New York City
62:09

62:09 classrooms where they're using a
62:11

62:11 curriculum called KDUM spends less than
62:14

62:14 10 minutes answering a few questions and
62:17

62:17 then the teacher's given right away a
62:20

62:20 sense of okay which concepts are the
62:22

62:22 kids struggling with which ones are
62:24

62:24 struggling with them? How might you you
62:26

62:26 organize the classroom? The data on that
62:29

62:29 which is small scale are you know one of
62:33

62:33 these stunning results that make you
62:35

62:35 want to really scale it up. When you
62:38

62:38 scale up an education you go from the
62:41

62:41 teachers who willingly engage in
62:42

62:42 experiments and are probably self-
62:45

62:45 selected for flexibility uh to the broad
62:48

62:48 population. And so the number of things
62:50

62:50 in education that look good in the small
62:53

62:53 that either don't affect when you scale
62:57

62:57 up or the quality of implementation
62:60

62:60 degrades so much that the effect is
63:02

63:02 basically washed washed out. You know
63:05

63:05 that's the history of education uh
63:08

63:08 innovation. I do think as AI helps you
63:11

63:11 with motivation.
63:13

63:13 uh you know many of these software tools
63:15

63:15 have helped motivated students but
63:18

63:18 actually created more of a differential.
63:21

63:21 You know, it's a kid like myself who
63:23

63:23 went home and used Khan Academy for two
63:25

63:25 hours at night and it's the, you know,
63:28

63:28 the median kid who comes in and sees,
63:30

63:30 wow, he's just, you know, so much
63:32

63:32 better. That's discouraging. That's, you
63:35

63:35 know, probably hurts. So unless you

63:37

63:37 believe in trickle down or something,
63:39

63:39 this is not equity at work. Can the AI,
63:43

63:43 not only through personalized learning,
63:45

63:45 but personalized motivation,
63:47

63:47 uh, overcome that? And I I I really
63:50

63:50 believe that we can we can get that
63:53

63:53 right.
63:53

63:53 >> Well, there's a question here of AI
63:55

63:55 tools that are specifically designed for
63:56

63:56 this and then a question in we were just
63:58

63:58 talking about with cell phones and
64:00

64:00 smartphones and social media and of how
64:03

64:03 AI will diffuse through society. And I I
64:05

64:05 think the big concern here and one I
64:07

64:07 have as well is this what now gets
64:10

64:10 called cognitive offloading. That what
64:12

64:12 AI is being used for in mass numbers by
64:15

64:15 students and being really quickly
64:17

64:17 adopted for is
64:19

64:19 I didn't do the reading, summarize it
64:20

64:20 for me or I have to write this essay,
64:22

64:22 draft it for me or like help me on these
64:25

64:25 math problems. And even I think where we
64:27

64:27 saw a big study out of China that that
64:29

64:29 was striking to me where you saw kids
64:32

64:32 using AI had increases on their homework
64:34

64:34 scores and then when they had to test
64:36

64:36 outside of that context, their test
64:38

64:38 scores began falling pretty sharply. And
64:41

64:41 so certainly one can imagine AI programs
64:43

64:43 that'll be good here. But in terms of a
64:46

64:46 technology and a way of interacting with
64:49

64:49 the world, diffusing through society.
64:51

64:51 Well, of course it could help people
64:53

64:53 learn, you know, linear algebra or
64:54

64:54 whatever it might be. Then in practice
64:56

64:56 what it's going to do is e is people
64:58

64:58 will be using it to help them and you
65:01

65:01 know both from the student level and up
65:03

65:03 up through you know the professional

65:05

65:05 level

65:07

65:07 beginning to degrade the learning the

65:09

65:09 creativity that happens through the hard

65:10

65:10 work of drafting and struggling and

65:13

65:13 challenging yourself. Yeah. Well, you

65:16

65:16 clearly would get by modality where

65:20

65:20 a student can use AI to be lazy and

65:24

65:24 corrupt the measurement system. Uh or a

65:28

65:28 student can use AI to help them learn

65:31

65:31 more. I have to say in terms of my

65:34

65:34 learning about subjects, uh this is the

65:37

65:37 best time of my life. I mean, I take

65:40

65:40 YouTube videos and put them into the

65:42

65:42 chat. I have Als uh talk with each
65:45

65:45 other. What a nirvana. You know, I
65:48

65:48 always had the ability to send mail to
65:50

65:50 somebody like a Nathan Mirville who
65:52

65:52 knows physics. Well, now I discuss the
65:54

65:54 physics thing with the AI. You know, I
65:56

65:56 get to it and then if it's super
65:58

65:58 important decision like I'm going to
65:59

65:59 invest a billion dollars, I'll send it
66:01

66:01 to Nathan and say, "Hey, could you could
66:03

66:03 you look this over?" But, you know, 24
66:05

66:05 hours a day, you know, keeping up to
66:07

66:07 date on the latest malaria thing or
66:10

66:10 vaccine thing. So that biodiversity is
66:14

66:14 going to be there and it really does go
66:16

66:16 back to this motivational
66:19

66:19 uh part of it.
66:20

66:20 >> Is that a place where there should be
66:21

66:21 more limits on the systems? I mean you
66:24

66:24 mentioned a few minutes ago China which
66:26

66:26 you know for all we talk about you know
66:27

66:27 it' be impossible for us to strike a
66:29

66:29 deal with them. China currently has
66:30

66:30 stronger national level regulations on
66:32

66:32 AI than we do including things like on
66:34

66:34 social companionship for kids. One of
66:37

66:37 the places that you talk about in the

66:38

66:38 essay where maybe we want to be more
66:40

66:40 aggressive is around children. And
66:44

66:44 that's a place where I think it is
66:45

66:45 reasonable for the government to be
66:46

66:46 paternalistic. Uh we should be
66:48

66:48 paternalistic.
66:49

66:49 And
66:51

66:51 social media always feels like an
66:53

66:53 experiment we ran on kids. A lot of the
66:55

66:55 internet does. And I don't so much mind
66:57

66:57 us running the experiment on adults, but
67:00

67:00 I wonder if we want to be more
67:03

67:03 aggressive
67:05

67:05 in what we don't let kids do with AI
67:08

67:08 until we have a better sense of how it
67:09

67:09 affects them. Everything from learning
67:11

67:11 relationships to uh you know trying to
67:16

67:16 make it harder to have AI help you cheat
67:18

67:18 on your homework.
67:21

67:21 >> Yeah. I don't think a black and white
67:23

67:23 ban uh is necessary because I do think
67:28

67:28 with monitoring, which I'm kind of
67:30

67:30 broken record on that, uh that making
67:35

67:35 sure it's staying within the bounds of
67:37

67:37 what's appropriate, what the parent
67:38

67:38 wants, uh our capabilities there are
67:42

67:42 very strong. There are many systems that
67:46

67:46 try to help instead instead of telling
67:48

67:48 you the answer uh they engage with you
67:52

67:52 on how do you reason through the
67:54

67:54 problem. And so then the question is can
67:56

67:56 you make sure the student isn't going to
67:58

67:58 those outside systems and just getting
68:01

68:01 the answer. Instead they're engaged in
68:03

68:03 the system that brings them along step
68:05

68:05 by step in that reasoning process so
68:08

68:08 that they'll be more capable the next
68:11

68:11 time. Um and and so yes, the seeing what
68:14

68:14 that student is doing and deciding,
68:17

68:17 okay, do you call in a psychiatrist? Do

68:20

68:20 you call in the parents? Uh a lot of
68:23

68:23 these systems, the idea of monitoring
68:26

68:26 will be very important. And at different
68:29

68:29 ages, the visibility of the parent about
68:33

68:33 what's going on and how time is being
68:36

68:36 spent. At say some very young age, a
68:39

68:39 [snorts] parent probably should have
68:41

68:41 complete visibility into the exact
68:45

68:45 dialogue uh that the students enga
68:48

68:48 engaged in. as the child gets older, you
68:52

68:52 know, are there things about sexual
68:55

68:55 identity or uh things that
68:60

68:60 you know that really it's beneficial for
69:04

69:04 that to at some level be uh private or
69:07

69:07 not. I think there's a lot of tuning to
69:09

69:09 go on with this. You know, one of the
69:11

69:11 partners the foundation's working with
69:13

69:13 is Common Sense Media who did a very
69:16

69:16 good job on, you know, this TV show or
69:19

69:19 this movie, what it exposes you to. This
69:21

69:21 is a much tougher area. Most of these
69:24

69:24 systems that have been set up with
69:25

69:25 parental control, the parents lack of
69:27

69:27 technical sophistication is such that
69:30

69:30 doesn't really matter uh that you tried
69:33

69:33 to engage the parent, nothing happens.
69:35

69:35 You know, in my case, I didn't know that
69:37

69:37 my daughter had a second cell phone. So
69:39

69:39 that you know was a fairly
69:41

69:41 straightforward way uh that she at in
69:45

69:45 late hours was able to stay on social
69:48

69:48 networking when I uh hoped that she
69:50

69:50 would be sleeping. Uh and so we have to
69:53

69:53 be realistic about how these things are
69:56

69:56 designed. People will worry about the
69:58

69:58 privacy issues there. That's fine. Uh
70:00

70:00 you know I believe we can make them it's
70:03

70:03 not an argument for being unmonitored in
70:06

70:06 terms of what your your kid is doing. So
70:09

70:09 next two years end politically, you have

70:11

70:11 the midterms coming up, then you'll have
70:13

70:13 a presidential election in 2028. So if
70:15

70:15 this were to go, not just in terms of
70:17

70:17 passing laws the way you wanted to, but
70:20

70:20 because you sort of center your call on
70:23

70:23 this
70:24

70:24 coming together of society to discuss,
70:26

70:26 debate, respond. What are you actually
70:28

70:28 envisioning here?
70:29

70:29 >> So I' I'd have two metrics. One is the
70:33

70:33 obvious safeguards to minimize the cyber
70:36

70:36 and biorisk and then the second would be
70:39

70:39 the nature of the dialogue. If the
70:41

70:41 dialogue ends up being one party is for
70:44

70:44 AI and the other party's against you
70:46

70:46 know then we'll have like climate change
70:49

70:49 uh which is not a great place to be. uh
70:54

70:54 you don't have room for debate when you
70:56

70:56 you only have the two extremes and the
70:58

70:58 default case at least right at the
70:60

70:60 moment seems to be very little middle
71:03

71:03 ground uh about uh how we should manage
71:08

71:08 AI. Let's say that
71:11

71:11 a very different administration and a
71:14

71:14 very different composition of political
71:15

71:15 power exists in 2029 and there's
71:19

71:19 interest in rebuilding American foreign

71:21

71:21 aid. What would your advice be on how to

71:26

71:26 build it back so it's effective, so the

71:29

71:29 American people understand what they're

71:30

71:30 getting for that spending and so it's

71:33

71:33 actually doing good?

71:35

71:35 >> Yeah. So, if you ask the American people

71:37

71:37 what portion of the budget's going to

71:40

71:40 help people in poor countries, uh, it's

71:43

71:43 generally more than 5%. You know, people

71:45

71:45 say 10 15%. So, the fact that it was a

71:50

71:50 half a percent and now it's going down

71:52

71:52 towards a quarter of a percent, that

71:54

71:54 would be pretty shocking to people, uh,

71:57

71:57 understanding that the money is well

71:59

71:59 spent, uh, that it really does save

72:01

72:01 millions of lives, that it made every

72:03

72:03 bit of difference in terms of HIV and

72:06

72:06 malaria and TB and, uh, maternal

72:10

72:10 survival. They should hear what they

72:12

72:12 should be proud of that even at that

72:15

72:15 half a percent level, you know, a moral

72:17

72:17 argument will be made. It'll be made in

72:20

72:20 the face of an overall debt level. Uh

72:24

72:24 that'll be very tough and whatever uh

72:27

72:27 safety network needs to be done because

72:29

72:29 of AI impacts. Uh that that will be

72:34

72:34 competing for those dollars.

72:36

72:36 >> Then always our final question, what are

72:38

72:38 three books you'd recommend to the

72:39

72:39 audience? Well, one I literally binged

72:41

72:41 this weekend was The Correspondent by

72:44

72:44 Virginia Evans. That's, you know,

72:46

72:46 fiction, very touching, uh, uh, very

72:49

72:49 upbeat. Appropo what we just discussed.

72:52

72:52 Uh, there's a one called Into the Wood

72:54

72:54 chipper by Nicholas Enrich, uh, which

72:57

72:57 talks about how,

72:60

72:60 uh, somebody didn't go to a party one

73:02

73:02 weekend and they decided instead to put

73:04

73:04 USA ID, uh, into the wood chipper. Uh
73:08

73:08 about a month ago, I read the infinity
73:10

73:10 machine by Sebastian Malib uh
73:13

73:13 >> AI book.
73:14

73:14 >> Uh yeah, history uh of uh Dean Deusibus
73:19

73:19 and Deep Mind, but you get a sense of
73:21

73:21 the whole founding of the AI [music]
73:23

73:23 industry does a uh an incredible job.
73:26

73:26 >> Bill Gates, thank you very much.
73:28

73:28 >> Thank you.
73:30

73:30 [music]
73:35

73:35 >> [music]
73:42

73:42 >> Hey.