THE BROOKINGS INSTITUTION

WEBINAR

# THE TURING TRAP:
# A CONVERSATION WITH ERIK BRYNJOLFSSON
# ON THE PROMISE AND PERIL OF HUMAN-LIKE AI

Washington, D.C.

Wednesday, November 2, 2022

PARTICIPANTS:

**Fireside Chat:**

ANTON KORINEK, Moderator
David M. Rubenstein Fellow, Center on Regulation and Markets, The Brookings Institution

ERIK BRYNJOLFSSON
Director, Stanford Digital Economy Lab,
Jerry Yang and Akiko Yamazaki Professor and Senior Fellow,
Stanford Institute for Human-Centered AI

* * * * *

P R O C E E D I N G S

MR. KORINEK: Hello there and welcome. I'm Anton Korinek. I'm a Rubenstein Fellow at the Center on Regulation and Markets and Brookings and a professor of Economics at the University of Virginia and the Darden School of Business.

It's a pleasure to welcome everyone to our virtual fireside chat. Our topic today is the "Turing Trap", and the promise and peril of human-like artificial intelligence. I am honored to introduce Erik Brynjolfsson to you today, with whom I will speak about this topic.

Erik is the director of the Stanford Digital Economy Lab, as well as the Jerry Yang and Akiko Yamazaki professor at the Stanford Institute for Human-Centered AI. Erik conducts research on the effects of digital technologies, such as AI, on the economy focusing on questions of productivity, growth, and social welfare. He has written several best-selling books on the topic, including "The Second Machine Age", which was one of the books that first introduced me to the topic and inspired me to work on the economics of AI. So I am particularly excited to speak with you on our topic today, Erik.

Before starting our conversation, let me thank the team that has made today's event possible, especially Megan Waring at Brookings, and Asme Usef (phonetic) at Darden. And let me thank those of you who have submitted lots of interesting questions during the sign-up process for this event. I will do my best to incorporate as many of them as possible in our conversation.

Erik, you have really been a pioneer in research on digital technologies and the future of work. You thought about this topic long before it received the level of public attention that it receives today. I think you even wrote your dissertation on a related topic, right?

MR. BRYNJOLFSSON: That's right.

MR. KORINEK: What was it that originally got you interested in this space? What made you think it is important long before many others did? And how have your concerns evolved since when you wrote your dissertation?

MR. BRYNJOLFSSON: Well, first, let me thank you, Anton, for inviting me. I find your work very influential in my own thinking as well, and it's an honor to be able to chat with you about these

topics once again, this time in public and join the Brookings audience.

You're right, I've been thinking about this a long time. I do remember reading Isaac Asimov's "The Foundation Trilogy", like a lot of people, a lot of kids did, and that was very influential. It combined sort of economics — I guess he called it psycho historian with technology. And that's kind of been a little bit of an implicit guide for me. As an undergrad, I was very into both those topics. Right after I graduated, I started a company called Foundation Technologies -- actually in homage to Isaac Asimov -- with Todd Loofbourrow that built expert systems, which is an early kind of AI. And I taught a course at the Harvard Extension School, one on AI and one on expert systems. So I was into that pretty early. It was just the '80s. That's how old I am.

After a little while though I decided I needed a little more depth in it, so I went to MIT hoping to combine the two, but it was a little hard to do them simultaneously. So I mostly focused on economics and worked with Oliver Hart and Tom Malone and Ernie Berndt. I remember one assignment as a first-year grad student Ernie asked me to plot the use of computer power in the economy, first in nominal terms — and people are spending more and more on computers — and then I did it in real terms, where you multiplied by the actual processing power that was being purchased by each dollar. And there were all these like stunning exponential curves. And I had a real aha moment. I was like looking at different industries, different areas, and they all had this like incredible exponential growth because of Moore's law and the related technologies. And I was just like, wow, this is amazing. And whatever we see today, it's just — I could — you know, you didn't have to be a genius to extrapolate a little bit and see it's going to be much more important.

So that made me double down my interest in technology. And, as you say, I did write my dissertation on information technology and the reorganization of work and have been working on it ever since.

MR. KORINEK: Wow. I am really glad you did it because, as I mentioned, otherwise I wouldn't have had you as an inspiration.

MR. BRYNJOLFSSON:

MR. KORINEK: Very kind.

MR. KORINEK: So this brings up to our main topic today. You have recently written a piece in which you coined the term the "Turing Trap" to refer to the promise and peril of human-like artificial intelligence.

MR. BRYNJOLFSSON: Right.

MR. KORINEK: Can you tell us a little bit more about what motivated you to write about the Turing Trap? And can you describe the main concern captured by this?

MR. BRYNJOLFSSON: Sure.

Well, truth is for a long time I've been playing around with the ideas of how AI and related technologies can substitute for humans or augment them. One of my first popular books with Andy McAfee was called "Race Against the Machine". And in there we distinguish between technologies that we race against, one or the other wins, versus technologies that help us — we call it racing with the machine versus against the machine — to augment us. And I gave a Ted Talk on that as well. And it was the theme of course in the second machine age. But then it really crystalized sharply when I dove in a little deeper into some of the literature about augmenting or complementing versus substituting.

And Alan Turing, you know, obviously one of the most brilliant people of our era, thought a lot about this and we're not getting very close to being actually to deliver on Alan Turing's question, which is how can we have machines that are intelligent. And, you know, he defined it as basically a machine that imitates humans so well that you can't tell which is which. It was based on a parlor game called the "imitation game" where men would pretend to be women or women would pretend to be men and you had to guess which was which. So could a machine imitate humans that well. And I remember thinking as a kid, when I was a teenager, thinking, oh, that's a really good if you — you know, if a machine is indistinguishable, that means it's intelligent.

I have since completely changed. I now think it's actually a bad test of intelligence. It's more like a magician levitating a woman in a magic show. Does that really mean that gravity has been solved or does it mean that people are gullible and can be tricked. And likewise, more importantly, it can

have some very negative economic effects. If you have a machine that closely imitates humans, then that makes human labor superfluous. Why do you need humans to do a job if a machine can do it. So it tends to drive down wages.

And a lot of people think that by definition, tech progress means you substitute for humans. The reality is that most tech progress is not substituting for humans, most tech progress is amplifying humans or complementing them. So a shovel or a bulldozer allows a person to do more work, it doesn't replace the worker.

But all through history, the past couple of hundred years really, the value of human labor has mostly gone up. Not for everybody, not for all groups, but mostly we are paid about 50 times more than we were a couple of hundred years ago because our labor is more valuable. Why is it more valuable? Because we can leverage ourselves with a lot of technology, both hard technologies, like machines, but also soft technologies, like businesses and software and innovations.

So tech progress that augments humans can increase wages. And the core insight from the Turing Trap is twofold. One is that by having machines that mostly imitate, we lower wages, having machines that augment, we raise wages. And most of us prefer a world where wealth is widely distributed. Both of those approaches can create wealth, but the first one, the imitation or substitution approach, tends to have most of that wealth concentrated in whoever owns the machine, capital owners, whereas the second one, augmenting, tends to have the wealth widely distributed. And I think most people prefer a society where everyone is participating.

The second insight is that if you are simply taking what's already being done and using a machine to replace what the human is doing, that puts an upper bound on how good you can get. I mean if you simply automate the process of, I don't know, making clay pots, so the clay pots can be done very, very cheaply, then you have a lot of clay pots, but you don't have new things. Whereas the bigger value comes from creating an entirely new thing that never existed before. You know, a supersonic jet or a nanoscale actuator or a new way of solving protein folding. so these new things are where most of the benefits come from. That's why we have iPhones now is because somebody invented something new,

didn't simply make a cheaper telegraph.

Anyway, so you put those things together and I came away convinced that we need to put more effort on thinking about how to augment humans and less on simply substituting for them.

MR. KORINEK: Yeah. So I think this concept of augmenting humans versus automating humans, it captures the main concern really well. But it still leaves out a lot of the details that need to be sorted out to basically distinguish which technology is which.

I think in the broadest sense we are concerned with what John Hicks called labor using versus labor saving technological progress.

MR. BRYNJOLFSSON: Yes.

MR. KORINEK: Now, you and I, we have both worked on how to classify specific technologies into these two categories and there are lots of different factors at work. The substitute ability of labor with technology, the price elasticity of the goods produced, and so on.

Can you give us a quick primer on how to sort out this complex challenge? What are the factors that matter to make a given technology labor augmenting rather than labor displacing — labor automating?

MR. BRYNJOLFSSON: Sure. Yeah. That's a great question.

Tom Mitchell and I wrote a little bit about this in our article in *Science* a couple of years ago called "What Can Machine Learning Do". And we laid out six factors. And it's worth noting that most people, including me, when we first think of technology, we think of substituting, we think of technology can substitute. But it's also possible, as I mentioned, that technology can complement. Those are two of the factors. Other factors depend on the price elasticity, that is that when you lower the price because of technology, does that lead to less total spending or more total spending. I'm drawing a demand curve here with my hand. If the demand curve is very steep or inelastic, then as you lower the price you end up spending less and less. That seems intuitive. But there are times when the demand curve is pretty flat and as you lower the price, the quantity grows even faster and then you end up spending more on the good. So, for instance, if jet engines make air travel cheaper, even though it's now possible to go from

one place to another more quickly and cheaply, that does not lead to a reduction, or it has not led to a reduction in the spending in air travel. We actually spend more on air travel than we did, and the Wright Brothers did or than people did in the '50s because the lower price makes it much more appealing. And a lot of technologies are like that. I mean computers have gotten cheaper and we spend more on computers than we did 20 or 40 years ago because they're so valuable and cheap.

So that's price elasticity.

I'll quickly mention the other ones. Income elasticity also matters, how much wealth we have. People buy different things when they're wealthier than when they're poorer. You may buy more luxury goods or eat out more and less of other things. There's the elasticity of the labor supply, which is that we respond to wages ourselves. Humans also are market participants. And, finally, there's business process redesign, which is just reinventing and changing all of these variables, so they're not necessarily fixed.

So those six factors, substitution, complementarity, price elasticity, income elasticity, the elasticity of labor supply and business process design, all affect whether or not a technology is going to lead to higher wages or lower wages. And those are to some extent choice variables. We can design technologies that go more in one direction versus the other, which therefore means we can design technologies that tend to raise wages, or we can design technologies that tend to push wages down. It's very much a choice.

MR. KORINEK: Thank you, Erik.

So you have laid out this challenge in all its complexity. Now, if we want to provide some

—

MR. BRYNJOLFSSON: I thought it was simplicity. (Laughter) Yeah, but you're right. It's moderately complex, yeah. But once you lay them out, it becomes really those six factors.

MR. KORINEK: So if we want to operationalize that if we want to provide really tangible advice to entrepreneurs, venture capitalists, policy makers, you know, our audience, what are the kind of one or two most tangible things that you would look at?

MR. BRYNJOLFSSON: Well, thanks for mentioning those groups, because another very important part of the Turing Trap article is I look at these different groups, technologists, entrepreneurs and business people, policy makers, and strikingly each one of them currently has misaligned incentives. Currently, technologists are — I've spent a lot of time with them — many of them, not all of them, are very focused on looking at humans and thinking how can a machine match that. It's kind of a very inspiring goal, is the Turing test. And so they try to make a robot hand that's like a human hand or play chess and checkers and other games like humans play them. And all these tasks the humans are doing, substituting when they should be thinking what an entirely new thing that we've never done before. But that second thing is much harder. It's hard to think of something totally novel compared to simply automating something.

And I spend a lot of time with business executives. I teach a business school and go in the field interactive with them, and same thing. You see them focused on looking at a task that they're already doing and then thinking how can awe replace the human with a machine as opposed to how can we do something new, which requires more creativity.

And, finally, policy makers, the current tax code and the current investment that credits and a lot of other decisions in policy, right now are very heavily skewed towards encouraging capital and discouraging labor. Tax rates on capital are much lower than labor. Back in 1986 they were the same, but since then they've changed in a way that favors capital investment and discourages investment in labor. It's not clear why that's a good idea to do that. We probably shouldn't be doing that. And similarly, for investments in education versus investment in capital and so forth.

So for all these reasons, our whole economy, rather than being level, is skewed towards substituting versus complementing.

Now, as I said earlier, it doesn't have to be that way. I work with a number of entrepreneurs who are doing something very different. Let me tell you a story about one of them. Cresta.AI, started by Sebastian Thrun and Zayd Enam, is a company that helps call center operators. But it's not one of those ones that has a robot operator answer your call or a robot text editor respond to

you. Instead they keep humans in the loop, and you talk to a human, but the human is given advice by a piece of software about which topics are going to be most useful to this particular caller, maybe reminding them about a complimentary product or a price decline or how to fix a particular thing. And by augmenting the human this way they've done fabulously well. They can handle a much broader range of questions, there's higher customer satisfaction, higher throughput. It turns out to be much more effective than the machine alone or than the human alone. Is combining the humans and the machines.

There are lots of other companies that are doing this kind of complementing and it's turned out to be very profitable ultimately for the ones who do it well. And also the part —- one of the things like is that —- and we did a study —- Lindsey Raymond at MIT is a Ph.D. student doing her thesis working on this with me, and we've found that the less skilled workers benefitted the most. The people with less experience actually benefitted the most from this technology. So it led to a more even distribution of income as well. Kind of a win on both dimensions.

MR. KORINEK: I think that is a really useful example because we can all relate to how annoying it is to deal with a call center system where you are just communicating with a computer.

MR. BRYNJOLFSSON: It is very —- we've all had that frustration and we want to get to a human. And I think this is one of the odd things I find, that so many people are focused on making technology that does everything humans do. And we humans can do some amazing things that turn out to be very hard for a robot. You know, picking a blueberry or comforting a baby or understanding some of the questions at a call center. Humans can kind of manage them, even if they've never heard that exact question before. And machines have a hard time with that. So why not let the humans do what's easy for humans and let the machines do what's easy for machines. And those are two different things. We don't need to try to get the machines to do what's hard for them, but easy for humans, or vice versa.

MR. KORINEK: Yeah. So let me feed in two questions from our audience now.

The first one, can you think of an example of an industry or an area that is kind of best taking advantage of the notion of complementing humans with AI rather than automating humans? So what would be an example of best practice for that?

And then let me add a second one, because I think it's closely related. What role do responsible AI frameworks and metrics, you know, such as fairness metric, interpretability, explainability metrics play in the labor market context? Do you know of any responsible AI frameworks or metrics that are in discussion today and that you would feel help mitigate the problems arising from the Turing Trap?

MR. BRYNJOLFSSON: Sure.

Yeah, let me just quickly touch on some examples of technologies that are augmenting. First, one that we can all kind of relate to because it's historical, is like the jet engines that allowed more people to fly. So it complements rather than substitute. But now in addition to the Cresta example, my colleague Fei-Fei Li is working on robots that help older people to do some of their household tasks. They still want to be able to be sort of independent and be able to manage their household, but the machines assist them in some of the tasks.

We also have seen some hiring systems that Danielle Li at MIT and others have studied, and Lindsey Raymond, which look at helping HR managers hire more effectively. I'm involved in a company called Workhelix that does exactly this. We scan through thousands of resumes. When people are looking to hire a new person, it helps highlight the ones that are most likely to have the relevant skills. But it's all done in support or cooperation with the human. Part of the reason that we found that that is important is that in a typical occupation, there are some tasks that come up very frequently and there are others that are much less frequent. So there's sort of like a Pareto Curve or a long tail of tasks. And the machines can learn how to do the more frequent tasks, but they have a lot of trouble on that long tail of the infrequent tasks. Maybe they've never seen them before. And so if you have the human and machine work together, the humans, for now, we are much better at handling those one-off kinds of tasks.

So if it's something that's unusual, the humans handle it, if it's something more common the machine can maybe figure out the optimal solution. I think the self-driving car people have discovered this, that it was very hard —- it has been very hard to make a machine that handles all the extreme unusual cases. And this is a pattern in so many industries. A way of addressing it is having humans and machines work together.

MR. KORINEK: And on metrics, specifically metrics.

MR. BRYNJOLFSSON: Oh, yes, metrics. Thank you.

So the metrics, this is something I think we need to do a lot more work on. One of the things — since I've been hanging around with these brilliant technologists, I am blown away at how good they are at hitting metrics that are effective. So if you put a target out there, they will figure out a way of achieving it. And many of the metrics that Jack Clark has been describing, technologies have very quickly hit them, like beating Go or some of the other ones. But the problem is many of the metrics are geared towards imitating humans. And I think there's actually a real lack of metrics that are focused more on augmenting humans.

And so one of my goals in the next few months actually is to get a group of people together to come up with an alternative set of metrics, unlike the Turing test, which is an imitation metric, and focus more on augmentation metrics. There are people who are trying to do some metrics that look at some of the ethical side, an alignment issue. But I think we need to focus more specifically on these metrics of augmentation or complements. And maybe within the next year or so we'll be able to publish a new set of metrics with that in mind.

MR. KORINEK: Yeah, I very much agree with the need for that. So please count me in on that effort.

MR. BRYNJOLFSSON: I would love to have your input on that, and anyone else who wants to contribute to the effort, contact me and we'll put together a team to work on it.

MR. KORINEK: So now I have a question from the audience that's a little bit contrarian. If it is perilous for technology to automate human labor, should we not have invented tractors and heavy construction machinery and so on.

MR. BRYNJOLFSSON: Yeah, right.

MR. KORINEK: And let me perhaps add a little bit more of an economic twist to that. So there is this insight in economics that if you have a production process that consists of multiple complementary steps and you automate some of the steps, then it is labor using or labor augmenting —

MR. BRYNJOLFSSON: Right.

MR. KORINEK: — because it makes the remaining steps more valuable.

MR. BRYNJOLFSSON: Yes.

MR. KORINEK: But once you automate all the steps, it is labor saving or labor automating. So there are at least some economic reasons to believe that technological progress within a given sector is naturally augmentation at first and then automation at later stages. And to the extent the capabilities of the human brain are finite, then this line of reasoning suggests that the Turing Trap may even be unavoidable in the long-term.

MR. BRYNJOLFSSON: Yeah.

MR. KORINEK: Unless you've already stopped progress.

So what are your thoughts on that?

MR. BRYNJOLFSSON: I definitely don't want to stop progress and one of my themes is I think we need more progress, more advances, but our technology is advancing very rapidly, which is wonderful. I'd love it go even faster. Our skills, organizations, institutions are not advancing as fast. So that gap is creating a lot of problems. The solution is not to slow down tech, but to speed up our adaptation.

And to both your questions — or the first one, of course I'm very happy we have tractors, et cetera. I gave as an example; a bulldozer is an example that I gave of something that augments. I also want to say, which I want to be very clear, it's in the paper but I haven't had a chance to say it yet here, is that many kinds of automation are very, very good, and there are many dirty, dangerous, unpleasant tasks we want to automate. I would love to see lots of automation of those tasks.

My point is simply that currently there's this skew where we are over automating and we could be doing more augmenting, that there's more upside in augmenting. So I don't want to stop automation. What I'd like to do is increase the amount of augmentation and restore a little bit of a balance so that we do them equally.

And in the case of many of these tasks, as you just pointed out, as you automate or

replace the human labor in one piece of a task, there may be some other complementary tasks that are essential that make humans more valuable. So let's just stick with the tractor, you still need the human to drive the tractor, at least for a while. And so that ended up creating demand for that job.

Now, over time, as you pointed out, you may automate that as well. Is there another piece of the chain that becomes valuable? And for that reason, it can be very tricky to draw a sharp line between what's automation and what's augmentation or what's substituting and what's complementing. It depends on the level of the system that you're looking at. Some substitute on one level may be a complement at a different level.

I think that's okay. It means we should be a little careful about being very prescriptive. I don't want to sit on some hill and say do this, don't do that, do this, don't do that. Instead what I would suggest is a good tactic for somebody at a policy level is to use taxes and other incentives to level the playing field. As I said earlier, right now we have, I think perhaps unintentionally, very strong incentives to substitute for labor and replace it with capital. Put more money in the hands of capital owners, less money in the hands of laborers. I'm not sure that was a conscious policy or unconscious policy, but that's what our current tax policy does.

And so then what that does is every time a technologist invents a new technology or entrepreneur wants to place something in, there's a gentle pressure on them right now to do more capital-intensive substitution and less labor augmenting technology. I mean the first rule of taxation I think is that you get less of whatever you tax. So we tax labor more than capital. And that's steering our technology in that direction. We could instead level the playing field and say let's have it treated equally, or perhaps we should even go one step further and encourage augmenting and not substitution. And the tax system that had these broad gentle incentives would leave it to thousands or millions of managers and technologists and local decision makers to each make their own local decisions. And whenever they made a decision that was a little bit more augmenting, they make a little bit more money than they otherwise would have. And whenever they're making a decision that was substituting, they wouldn't get the subsidies that they're right now getting for that. And over time, as with a carbon tax or other kinds of

policies, it would steer the economy into technological progress a little more towards augmenting and a little bit less towards substitution or imitation.

MR. KORINEK: I like that very much because incentives work.

MR. BRYNJOLFSSON: Yeah, incentives work. And our tax code — just one more thing on that. It doesn't have to be just thinking at labor and capital taxes, which we're very focused on. But one of the things I learned from the folks at Brookings and American Enterprise Institute is we have some other tools, like value added tax or X tax, that treat things much more evenly. So it doesn't have to be like a capital tax per se or a labor tax per se. If we went to a much more of a value added tax system, that implicitly treats things much more evenly than our current system.

MR. KORINEK: Right. Yes. Now, let me ask this, perhaps the most fundamental question.

MR. BRYNJOLFSSON: Mm-hmm.

MR. KORINEK: So in some ways the premise of the Turing Trap is that our current system in which the majority of people derives most of their income from labor is desirable. And there are some clear benefits for income distribution, for political stability, and so on, but a fundamental counter argument would be that our system force people to work because we are too stingy to engage in a fair distribution for resources and all our economic models say that work actually creates disutility, work is a burden. It's not something that inherently creates utility. Now some jobs are of course more desirable than others and while I am certainly enjoying our conversation (laughter), but —

MR. BRYNJOLFSSON: I'm not getting paid for this, but I still consider it part of my job and I enjoy it a lot. And I think that that underscores a key point, that there are some jobs that are very unpleasant, there are some that are enjoyable. It's more complicated than that.

And one of the things I learned from Tom Putnam that the usual economist assumption that we should try to eliminate work and just give what people really care about, money, is not the way most sociologists or a lot of the data suggests. In fact you look at places where jobs have disappeared and people get paychecks from disability and welfare instead, people are not happy. Those are some of

the places with the highest deaths from despair, opioid, alcoholism, suicide, depression. So a lot of people enjoy contributing to society and get meaning from doing something productive and useful. And we should bear that in mind.

And I'm not saying that our current system is like the one I want to freeze in place. I definitely don't think that. I think we should continue to evolve. But there's a subtler point, which is that I do think if people are contributors to society and have a stake in society, they're going to inherently have more bargaining power, whereas if people are useless — was it Harari called them the useless class, or Tyler Cowan had this phrase — I think he coined it — of zero marginal product workers. If there's a lot of people that are not needed for production, I mean perhaps benevolent people like you and others will push for giving them a universal basic income and payment and so forth, but they're kind of at the mercy of that kind of benevolence. They don't have the bargaining power to say hey, if you don't pay me, I'm going to stop working and then you're going to lose something. If they have no bargaining power, I think that's a perilous and precarious existence for them. I hope that those with the power will be generous and kind, but I — you know, history says that you can't always count on people in power being that way.

So I think a more stable equilibrium is one where people are needed for production, they're indispensable. And so that they are part of it. And that way if not getting a fair shake, you know, they can go on strike or they can go to a different job or they can do something and then the natural equilibrium will be for them to get a share of the spoils.

I know what you're going to say next, which is what about as technology progresses.

MR. KORINEK: Yeah, let me first a little bit back on what you've had to say so far.

MR. BRYNJOLFSSON: Yeah.

MR. KORINEK: So I think we agree that in our current system unemployment is something that the vast majority is terribly afraid of. Our systems of supporting the unemployed are very stingy and we would really need to radically reform our system of income distribution.

MR. BRYNJOLFSSON: Yes. Yes.

MR. KORINEK: But I think that there is a risk that we act out of a sort of status quo

system, that we are too bedded to our current economic systems —

MR. BRYNJOLFSSON: Right.

MR. KORINEK: — of paid labor and the main source of income. And it's a system that has prevailed for some time, for some 250 years, but it is not something that humanity has inherently evolved with. And, you know, as you have faced more and more powerful AI systems, they could actually free us from the need to work.

MR. BRYNJOLFSSON: Right.

MR. KORINEK: And that could lead to much greater human flourishing —

MR. BRYNJOLFSSON: Right.

MR. KORINEK: — than merely focusing on augmenting human labor.

So I guess in short, the question is should we free labor rather than augment labor?

MR. BRYNJOLFSSON: Mm-hmm. I think it's a little bit of a semantic argument. I agree that people should be free to — freer, have more choices for the kinds of things they work on. And many people, like you and I, we love our jobs. And I think there are a lot of other people who love their jobs and I think we should move towards a society where most people love their jobs and are doing things because they really like doing it, they can do it as a volunteer basis, or they can get paid for it. If they don't need the money, then that's awesome.

So but I really would make three points. The first one is for many people some sort of work is fulfilling. Now, we could potentially get more and more fulfillment from things that's not paid labor, and I'm okay with that. Maybe people enjoy writing stories for the fun of it or playing video games. Or maybe they'll enjoy working and that's okay. I can see that evolving a bit. So that's the point from Putnam and others.

The second point, which is really the key one, is that I think a stable equilibrium is one where people have some bargaining power, that the money — the support they get is not just from the generosity of other people who have all the power, but it's something that they actually have power over. A quote attributed to Louis Brandeis is that you can have a great concentration of wealth, or you can have

democracy, but you can't have both. That may be a little extreme, but I think there's something to be said that a concentration of economic power tends to lead to a concentration of political power. And so if we want widely distributed bargaining power, widely distributed political power, we may need to have the suitable economics.

That said, there's a third point that's implicit in what you're say, is that over time I think the machines will be able to do more and more of what humans do. And this equilibrium that I see lasting for a few decades, which is a pretty long time, of humans continuing to be indispensable is not necessarily the one that will always be here. And there will come a time when machines can do so much that humans are currently doing that it will be difficult to sustain something where humans are always indispensable.

And so we will over time need to develop some institutions that continue to give humans political and bargaining power, even when they don't have the same kind of economic indispensability. That's very tricky and that's a big challenge. I don't think it's the challenge that we face in 2022 or even 2030. It's something we should be working on. For now I think we can address it by keeping people involved in the production process.

When I look around, when I look out the window, when I look around Palo Alto, or any city, I see a lot of work that need to be done that can only be done by humans currently, with current technology and for the near future technology. So there's no shortage of childcare, elder care, cleaning the environment, innovation, arts, teaching, a lot of other things, a lot of dexterous things, you know, deliveries and whatever, that only humans could do right now. And so we're not running out of work for them, there's no shortage of work for them.

So let's go ahead and make sure that everybody has a way to contribute to society and a way to earn income that way. And over time, smart people like you can be thinking harder about what should we do in a few decades when machines are so good that we can't find work for people to do. But I don't think that's the world we face right now.

MR. KORINEK: I think we are converging on a solution here.

So I very much agree with you that in the present we want people to earn a decent living from their jobs, but at the same time in the longer-term we may have to be open to that system of income distribution.

MR. BRYNJOLFSSON: Yeah. I think that's right. And we don't exactly know when that longer-term is. I had a good dinner here the other night with — one of the things about being here in Silicon Valley, Greg Brockman came over from OpenAI. He has a very aggressive time scale. He thinks — I think I can share with you that he was talking about 2029 when machines would be able to do most of the things that humans can do. There are a few other people, Ray Kurzweil and others, that are that ambitious or optimistic about the technology. My guess is it's going to take somewhat longer, but we have to have a distribution. There may be some amazing breakthroughs that speed it up, there may be some barriers that push it further — you know, Rod Brooks has talked about centuries.

MR. KORINEK: Yeah, I thin Elon Musk also said 2029. And it's not my median estimate, but I certainly wouldn't completely rule it out, but yeah. So we have —

MR. BRYNJOLFSSON: Hopefully we'll both be around to see that.

MR. KORINEK: Yeah. So we have discussed one specific themes that are human-like AI so far, but there's also a bunch of other dangers. If we make our AI systems more and more like human agents, then there are some dangers that, for example, Nick Bostrom has written about extensively. For example, the danger that future AI systems have planning capabilities that are superior to us humans and —

MR. BRYNJOLFSSON: Right.

MR. KORINEK: — that they may in some sort take over either progressively or maybe abruptly and really reduce the space for human agency.

MR. BRYNJOLFSSON: Yes.

MR. KORINEK: Now, in a worst-case scenario, this could be something like Nick's example of a paper clip maximizer that eradicates humanity, but we don't even need to go that far. It seems to me that there is really a close connection between Nick's concerns and the concerns in the

economic sphere that you have articulated.

So if we followed your advice and we make AI systems less human-like and less (inaudible), I think it could actually also mitigate the risks of an IA takeover. So I wanted to ask what are your through on that, do you agree with that observation, or do you have other concerns?

MR. BRYNJOLFSSON: I think that's true. And I share some of the concerns about AI alignment. I've spent some time talking to people like Nick — or to Nick and Max Tegmark and Eliezer Yudowsky. And Stuart Russell has written very eloquently on this. And they're smart people and I think it's a serious question. It turns out that it's harder to align humans and machine goals than I would have thought in advance. And it's not so much — it's not the terminator problem that they're going to be evil necessarily, it's more just that just defining what our goals are turns out to be surprisingly hard. And if we slightly misalign them and the machines are super intelligent, then you can get some really weird, unexpected outcomes and it can be hard to undo them.

So I am glad there are people working on that. Again, I don't see that as the most near-term challenge, but it's very important that we continue to work on that and think about it. And as you say, by having machines be very different and interdependent with humans, in some ways it can mitigate that a little bit and leave humans in charge. We have a conference coming up here at Stanford called "AI in the Loop, Humans in Charge". That's the slogan for it. And so that's a philosophy I can see helping with both kinds of problems going forward.

MR. KORINEK: Right, yeah. So let's stay with the theme of ever more advanced AI systems but bring it back to the present.

There's this new set of models out there that has gained prominence in recent years, and I think was actually your institute at Stanford that coined the term "foundation models" for this set of AI systems, right?

MR. BRYNJOLFSSON: Yes. No, these are amazing. I'm so happy to be here with my colleagues who are making these breakthroughs and I get to learn about them very quickly. Deep learning was the big breakthrough that people like Jeff Hinton have been working on for years and in

2012, when it won — it was the best system at ImageNet. Everybody realized, hey, this is a powerful new paradigm. And that's really what's been sweeping things.

I think these foundation models may be as big or as important as the deep learning revolution. That's saying quite a bit. These are generative models that are pre-trained on a large data set and then you can use that to solve new problems. So the ones that many people may be familiar with are like DALL E that does the artwork or GTP 3 that can do text generation stories or SAs or Tweets or poems, songs. Even these tools can be used to generate code. And Greg was saying that at OpenAI — Greg Brockman was saying at OpenAI about 40 percent of the code now generated by those very good coders is actually automated by these systems where you just say in basically plain English what you're looking for and it will generate the python that runs.

You still want to have a human in the loop to make sure it does it correctly, et cetera, but it's just astonishing how good these systems are. They tend to be very large systems that may cost millions or even hundreds of millions to train, with billions — I think PaLM Google system has about 540 billion parameters they said. It's just staggering the size of them. They're trained on huge corpuses of a lot of data. And as a result, they're able to solve problems in an almost spooky way. And even some people find that — there was the engineer at Google who thought that they were sentient. I don't agree, but they have plausible answers to questions about — that would seem to pass the Turing test.

MR. KORINEK: Mm-hmm. So I think, Erik, you were actually on the paper —

MR. BRYNJOLFSSON: Oh, and one more I should say. We were calling them foundation models because calling them simply large language models, whatever, is a little too narrow. So we see that they lay a — they build a foundation that you can build things on top of them and extend them. And even like very small companies can use the core foundation model to do something that may be more specific to their needs.

MR. KORINEK: Yeah. Now —

MR. BRYNJOLFSSON: Percy Liang really has been leading that effort on — at the Center for Research on Foundation Models.

MR. KORINEK: Cool. You were actually on the paper that coined the term, right?

MR. BRYNJOLFSSON: Right. Yeah. Yeah, exactly. A bunch of us run that paper and I like working with him. We're doing some projects right now to understand the economic impact of these foundation models.

MR. KORINEK: Right. That was my next question, what do you view as the main economic effects. Like, for example, you spoke about how these models are ever larger and cost ever more money —

MR. BRYNJOLFSSON: Yeah.

MR. KORINEK: — to train and run, which would suggest it's going to be more and more concentration.

MR. BRYNJOLFSSON: Could be.

MR. KORINEK: How do you view the opportunities and the risks?

MR. BRYNJOLFSSON: Well, the first thing, let's not miss the big story, which is it can — to be a huge increase in productivity and performance. So writers who work with them, coders who work with them, whether it's writing ad copy or stories or essays, people are massively more productive, and we're finding more creative interestingly. Because they're very fun, they generate creativity. You and I are in the National Bureau of Economic Research, and I had to give a discussant of a paper at an AI conference — actually it's coming up on two years ago now — and I used one of these foundation models to write my discussant comments. And I thought it was just — it was a lot of fun. I mean I still took responsibility for it, but the foundation model did a lot of work.

And then it's a little kick, since I was the last speaker, I had to do a version in the style of Taylor Swift, and I thought it was a pretty beautiful song if you're into economics and music at the same time, that this model had written. In just 30 seconds it generates it. So it was a creative use of it.

But what's the second part of — what was your question? It was —

MR. KORINEK: So what do you see as the risks and what do you see —

MR. BRYNJOLFSSON: Oh, yeah, the risk part. Yeah, that's the good part.

MR. KORINEK: — in terms of concentration?

MR. BRYNJOLFSSON: I think it could affect a large portion of the economy in terms of better productivity. But as you say, these models tend to be very expensive to train and there's only a few companies that can do it. And, by the way, academia can't really afford it. It used to be that most AI research was led by academia and now it's harder and harder to pay the table stakes to do some of these systems. So it ends up being Microsoft, OpenAI, Google, and others that make these large expenditures. And that business doesn't always have the same incentives as academia or as society. And so you have to think about that a little bit. Could to lead to more concentration of control and power.

Percy Liang, my colleague, has an optimistic hope that the base foundation models maybe are trained by very large companies, but most of the applications built on top of it could actually help smaller companies. And so it may become more of like a utility, like cloud service or like electricity, and then other people do things on top of them and that wouldn't necessarily lead to concentration in those other parts of the economy.

So it could be a little more complicated than that.

MR. KORINEK: Mm-hmm. So in terms of funding these large expenditures, it looks like there are currently two models out there. One is kind of the American British model, let's call it, which is that most of the research in these foundation models is going on in private corporations. And then there's the Chinese model where our government is really kind of actively trying to concentrate research on foundation models institution that's kind overseen by the government.

Do you think we may want to move more in that direction? What do you see as the benefits, disadvantages of the two models?

MR. BRYNJOLFSSON: Yeah, I'm not — yeah, you're right that the Americans and the Chinese — or the West and the Chinese are pursuing different models. I don't want to call the one where the government helps support it the Chinese model because America has long had a tradition of supporting R&D quite a bit. And one of the reasons the United States became such a technological leader in the '50s, '60s, '70s, was because of massive R&D investment in people, like Vannevar Bush and

others that pushed for this endless frontier of research, government supported research, whether it's the space program or the internet or a lot of biomedical research.

We desperately need to step up at Western societies, and the United States in particular, to make bigger investments in these areas to support a national research cloud that universities can participate in and, frankly, to be — and partly to keep up with Chinese development because there is also a geopolitical aspect to these. They can be used, as I said, for a lot of productive uses. These technologies can also be used for military purposes. And it's critically important, as Eric Schmidt and others have pointed out, that if you want democracy and Western values to be able to defend themselves, we have to have technologies that are equal to or better than rivals who have different sets of values and may not support the kinds of liberal democracy that most of us value.

So this is an urgent economic, political, geopolitical need and I hope the government steps up and makes those investments. And it should be a partnership between government, business, academia, and civil society to make sure that we're addressing these in ways that are consistent with our broad values.

MR. KORINEK: Mm-hmm, yeah. So we have only ten minutes left and there is one last topic that I would like to touch upon in our conversation. We have talked about all this potential productivity, gains, but I wanted to talk a bit about the connection between our measures of economic output and welfare.

MR. BRYNJOLFSSON: Yes.

MR. KORINEK: Because I know you've done some interesting work in this field. And Joe Stiglitz once wrote, what we measure affects what we do —

MR. BRYNJOLFSSON: Absolutely.

MR. KORINEK: — and if our measurements are flawed, decisions may be distorted. So against this backdrop, have you some thoughts on how we should measure the benefits of digital technologies better, how we should measure how they really benefit consumers, and can you tell us about it?

MR. BRYNJOLFSSON: Yeah. Well, Joe's absolutely right. And one of the dangers is that we currently increasing —– the metrics we're looking at are more and more disconnected from what really matters in our society. I think it's one of the great ironies of the information age, this data rich age that in many ways we actually have worse data about how the economy is creating value. Traditional GDP is very good at measuring tons of steel, bushels of wheat, cars coming off the assembly line, it's very bad at measuring software, digital goods and services, and many of the other things that —– intangible assets that are increasingly valuable in our economy.

The good news is that we have platforms where we can reinvent our measurement tools using these digital technologies so that we can measure the digital economy much better.

And let me get a little specific, hopefully not too wonkish. GDP is our main measure of the economy right now, gross domestic product. And it's very good at measuring all the things that are bought and sold in the economy. If something has a price it will likely show up in GDP. However, if the price is zero, the weight in GDP, with some exceptions, is zero. So we know there are a lot of valuable goods that have zero price. Right now, what are we on, Zoom. You know, zero price, that I'm paying for that at least. And many people —– or Wikipedia or search or email or there's just a bunch of things that we get for free. Or some of them are advertising supported, some of them are just donated. They don't show up in GDP or they're not well accounted for in GDP. Advertising is an intermediate good, not a final good, so it's not part of the final production. That means that as the digital economy grows bigger, we're missing a bigger share of where value is created, not to mention health and environment and household production, all other things that also are not priced or not priced well.

So we have developed a new measure. Avi Collis, Kevin Fox, Erwin Diewert, Felix Eggers, and I have developed a new measure we call GDP-B. And the B stands for the benefits. Instead of measuring what we pay for the goods, which may be zero, we're trying to measure the value we get from it. Now, that's harder. Economists call that consumer surplus, but with a series of online choice experiments using these digital platforms, we can measure how much you would have to pay somebody to stop using a good. So we offer some —– you know, we looked at hundreds of thousands of users and

we've offered some of the $5 to stop using Wikipedia for a month, others we offer $20, some we offer money to stop using email or Facebook or other digital goods. And some people say, yes, I'll take the money and then we block them from using it, and other say no, no, no, I would rather keep using it. And then we know how much value — or we can guess how much value they have. It's not perfect, but we start getting these downward sloping demand curves where some people value the goods quite a bit, others are willing to give them up for little or even nothing. Obviously, some people don't even use them.

MR. KORINEK: What kinds of values did you come up with in these examples?

MR. BRYNJOLFSSON: Yeah. So like Facebook. You know, we were getting values like $30 or $40 a month for the median user. It was quite a bit that you have to pay people to stop using it. I mean they want to stay connected with their grand kids or their friends or whatever. And it varied by country. You know, it turned out in the Netherlands WhatsApp was fabulously valuable. I thought we had made a mistake, did we get the digits wrong, but the group of people we surveyed over there, they were just very dependent on WhatsApp to stay in touch with their boss, their babysitter, their friends when they're going out. In America WhatsApp had a much lower value. So you can see — and in different groups — I got on Facebook women value Facebook more than men. I didn't realize that. And older people value it more than younger people. And obviously you could see that people with more friends or who post more frequently value Facebook a lot more than people who don't do those things.

So we got some pretty sensible, and sometimes interesting results. We're now doing it for ten countries around the world with a whole set of different digital goods. And over time my ambition is just as Simon Kuznets invented GDP back in the 1930s — you know, GDP didn't always exist. A team invented how you measure it. We are trying to invent a set of metrics for the 21st century that will measure what people really value, not simply what people are paying. And then we can hopefully with Joe Stiglitz's question of measuring the things that really matter and guiding policy makers and business people and just ordinary citizens to understand where is value being created. It's not just in steel and oil, a lot of it's being valued — created in digital goods, household production, health, and other things that currently are not property priced in GDP.

MR. KORINEK: Yeah. So one of the latest sets of technologies around which there's been a lot of buzz over the last year has been the Metaverse. And that's again one of these digital consumer goods, I guess largely speaking. How big do you estimate is the gap between the consumer surplus and the measured contribution to GDP by something like the Metaverse?

MR. BRYNJOLFSSON: Yeah. Well, right now it's still pretty small because the Metaverse is pretty tiny. Actually last week I had Herman Narula come and speak at my seminar. Here's his book — it happens to be on my desk right now — "Virtual Society". And terrific book. And he's developing a lot of the infrastructure. His company, Improbable, is developing the infrastructure for that. I'm one of the people who thinks it's going to be quite big actually eventually. The average American spends seven or eight hours a day looking at screens, whether it's a computer screen or a phone screen or a TV screen. So maybe half your waking hours of the average person. That's a lot of time interacting with a kind of a virtual world. And I could only see that getting more over time and becoming richer, three dimensional perhaps, as with virtual reality. So that's very important.

And that means we're going to get more and more of our value from these virtual goods over time. Many of them have zero price, but we still get very positive value from them. And that gap between what we're measuring in GDP, the price that's paid, versus the value that we're getting, is going to get bigger and bigger. I mean in an extreme though experiment, if we went to a matrix world where we spent almost all of our waking hours interacting in a virtual world, and many of us would be kings with palaces and lots of people we're talking to and we interact with all sorts of interesting people, some of the would be human, some of them might be machines. In a world like that, most goods and services would have close to zero marginal cost, right. I mean they'd just be made of bits, but the value might be much larger than what people are getting today.

So we need to come up with these new metrics to understand an increasingly digital economy. And I think that’s big project for the 21st century. I'm glad I'm working with some people at the Bureau of Economic Analysis in Washington, with the Office of National Statistics in the UK, and other groups around the world to try to develop these new metrics that are needed for the digital economy.

MR. KORINEK: Yeah. Well, so we have only two minutes left, so let me ask you one final question after we have discussed the dangers falling into the Turing Trap and lots of related interesting questions for the past hour. Can you give our listeners a one-minute summary to take away, something that is actionable and basically responding to the question of what can members of academia, the business community, or the policy community here in Washington, DC contribute to avoid the perils of the Turing Trap and to ensure that this technological progress that we are seeing now really leads to broadly shared prosperity?

MR. BRYNJOLFSSON: Well, first off, thanks for having me here. But I would say that this is to me the biggest challenge for our society right now, how to harness the power of AI and related digital technologies, which I think are not only the most important technology of our era, but arguably of any era of human times. I think you share that view. Not everyone does, but I think it's really big. But at the same time, there's this very big risk that we're going to lead into a trap where this enormous power ends up being highly concentrated in a small group of people.

But the key thing to understand is that I'm not a technological determinist and I don't think that there's any inevitability of any particular outcome in terms of how we play this out. It's very much a choice. In fact, the tools we have now are more powerful than any we've had before, which almost by definition I mean we have more power to change the world, to shape the world in different ways. That's what a powerful tool does. Which means our values matter more. And the biggest thing I think people are lacking is that far too many people are — they ask me, what's going to happen, what's the technology going to do. And what they should be asking is what do we want to do with the technology, what are our values. So we need to think more about our values than we ever have in history.

It's one thing to have a set of value when you have a rock and you can't really make a big dent in the universe, but now, as Nick Bostrom and others have pointed out, we can make a huge change in the universe. And so we need to think very carefully what kind of world we want to live in. do we want a world with widely share prosperity, do we want a world where everybody has a stake, where everybody has some bargaining power, where everybody has a role to play. If we do, I believe we can create that.

And the job of the Digital Economy Lab at Stanford is to do the research to understand what are the levers that matter, what are the policies that are going to make a difference, how can we measure things more carefully, so we do a better job.

And I welcome anyone who wants to join in that mission to understand the digital economy and understand the levers that matter and start describing the steps forward to build a wildly prosperous society that benefits the many, not just the few.

MR. KORINEK: That was a great summary, Erik.

And thank you so much for joining us today, for sharing the insights from the important work that you are doing. And also thank you to everybody in our audience. I hope that you found the conversation with Erik as insightful and as inspiring as I did. And I hope that you will join us again for future events in this series.

So now let me just conclude with some virtual applause to Erik. Thank you again.

MR. BRYNJOLFSSON: Thank you. And please go to my website if you want to read about any of the things that we've been talking about.

Thanks, Anton, and thanks to Brookings.

* * * * *

CERTIFICATE OF NOTARY PUBLIC

I, Carleton J. Anderson, III do hereby certify that the forgoing electronic file when originally transmitted was reduced to text at my direction; that said transcript is a true record of the proceedings therein referenced; that I am neither counsel for, related to, nor employed by any of the parties to the action in which these proceedings were taken; and, furthermore, that I am neither a relative or employee of any attorney or counsel employed by the parties hereto, nor financially or otherwise interested in the outcome of this action.

Carleton J. Anderson, III

(Signature and Seal on File)

Notary Public in and for the Commonwealth of Virginia

Commission No. 351998

Expires: November 30, 2024