

AI-GR Pod 25 12.12.24 Larry Summers

[00:00:00] I think it's possible that this is going to be the most important technology that ever got invented. I think that it's an even money chance that when historians write about the second-fifth of the 21st century, Russia and Ukraine and Donald Trump and Xi Jinping will be secondary stories, and that the real story will be the dramatic discontinuity that was associated with human like intelligence being achieved by non-humans.

I think that in a nearer term thing, it may not be [00:01:00] unreasonable to think that this is to the Internet, as the computer was to the calculator. The calculator was a very important thing. The computer was a much, much more fundamental thing.

Welcome to *NEJM AI Grand Rounds*. I'm Raj Manrai and I'm here with my co-host Andy Beam. Today we are delighted to bring you our conversation with Larry Summers. Andy, I think it's fair to say that Larry Summers truly does not need an introduction. He's been so influential. He's been the Secretary of the Treasury of the United States, the President of Harvard University, Director of the National Economic Council, and so many other things.

And as we dig into the episode itself, he's been a member of the Board of Directors of OpenAI for just over a year. I've listened to him a lot, but I think this conversation still surprised me. Maybe it was his enthusiasm or his predictions for what would be [00:02:00] remembered by historians of this period of the 21st century, the way he described how these models are continuing to evolve and also our uncertainty about where we're going really was quite powerful.

Yeah. I'm going to break the fourth wall a little bit here, Raj. And how cool is it that we got to talk to Larry Summers about AI? Like how amazing was that? Again, I agree with you. I'm a big fan of his. I've listened to all the podcasts he's been on and still, there were several things that I learned in this conversation.

We got a little glimpse into what happened with some of the OpenAI drama with the board turnover and with the leadership turnover. And so that was news to me. You know, I think you did a great job at asking him about his appearance

in the movie, *The Social Network*. And so that was like super great to hear. So much fun. So much fun.

So many gems like that in this conversation that I'm excited for the listeners to hear. And again, just like what a treat to get to talk to Larry Summers about such a wide range of topics. Really, really, really fun conversation. Totally agreed.

The *NEJM AI* [00:03:00] *Grand Rounds* podcast is brought to you by Microsoft, Viz.ai, Lyric, and Elevance Health. We thank them for their support.

And with that, we bring you our conversation with Larry Summers on *AI Grand Rounds*. Larry Summers, thank you for joining us on *AI Grand Rounds* today. We're super excited to talk to you. Larry, welcome and thank you for being here. This is a podcast about AI, so please forgive our nerdy question framing, but we always start with the same question.

Could you please tell us about the training procedure for Larry Summers' neural network? What data and experiences led you to where you are today? And please take us back as early as you can. You know, we say that we like to start with the initial conditions. I'm the child of economists. My parents were economics professors.

I have uncles who were prominent economists. So, I found myself writing [00:04:00] on a graduate school essay that while some children were taught to believe in God. I came to believe in the power of systems analysis. So, I'm a person who's always believed very much in data, empirical analysis, logic, argument as the way to get closer to truth and to believing that you're likely to do much better in contexts where you have more understanding than where you have less.

And that's why I've devoted myself to a career as an economist and tried to focus on studying economics for its own sake, but very much also as a tool for leading to human betterment in [00:05:00] spheres ranging from the better allocation of resources in health care to the avoidance of financial catastrophes that lead to millions of people becoming unemployed.

And so, my training has been from a kind of analytical social science perspective. So, trading the invisible hand of God for the invisible hand of Adam Smith. Is that fair?

That would be a way of putting it. You know, I've always, as a sort of ideological matter, always thought of the task as being to find a well-handled middle ground where you recognize that the invisible hand can't do everything, but that there are enormous dangers to heavy hands, [00:06:00] and to try to think of ways in which governments and policy makers can provide helping hands.

I think that's a great point to transition. So, you know, we want to dig into some of your work and your recent roles, particularly around artificial intelligence. And of course, we want to start with large language models and their oversight. So large language models like ChatGPT. We had Peter Lee from Microsoft on the podcast last year.

And he gave us a preview of GPT-4 before it came out. And I can still remember to this day, losing sleep after talking to him, thinking about what this means. You know, we've seen the evolution of models at OpenAI, but also, I think this very competitive ecosystem of other proprietary and open-source models emerged since then.

And we like to talk about the scale hypothesis. And what I mean by that is this idea that with increasing compute in data, we are going to keep seeing improvements to the fundamental capabilities of these models. So, I think many of [00:07:00] us agree that we're on some type of, you know, S-shaped curve, but where exactly on the sigmoid we are, is debated.

And maybe I can ask you how you think about the continued growth of these models? Whether you have some intuition about whether we will keep seeing major improvements to capabilities with increased scale, as we've seen for the past few years. I think closely related to that are your perspectives on what the key bottlenecks are for continued exploration of larger and larger scale.

You know, energy, data, compute, they all come up. But how do you think about these problems and the sort of fundamental increase in capabilities at scale of these models? Well, let me first say that there are two groups of people. There are the people who know that they don't know, and the people who don't know that they don't know, and I'm in the first group.

I know that I don't know for sure. I think there's a general lesson that's helpful for us all to keep in mind with [00:08:00] respect to technologies. That things take longer to happen than you think they will. And then they happen faster than you thought they could. And there are all kinds of examples of that. I first met Jeff Bezos when he came to a CEO lunch at the treasury in 1998.

And I remember then-secretary Rubin and I saying to each other that the owners of malls had better watch out. And we were right. That was a legitimate insight, but that wouldn't have been a financially important, legitimate insight for another 15 or 20 years after we had it. I remember hearing now a decade ago about these automated vehicles that had driven up Route 101 from San Jose to San Francisco with no driver, and I would've been surprised at that point [00:09:00] if you had told me that no truck drivers have lost their jobs by the beginning of 2025.

On the other hand, I'm sure there will ultimately be pervasive impacts. It's been a long time since I read a paper newspaper, and that's another example of a change. But I remember not that many years ago, whenever I wrote an op-ed regarding it as an important part of negotiation, whether my op ed was going to appear in the printed paper or only online.

So, it's very hard to know what the exact timing is going to be. I would be very surprised if in the fullness of what these kinds of models are going to accomplish, we were past the fourth inning. [00:10:00] And I say that because it's relatively early days in terms of the amount of time that there has been since there were really serious models going.

I say that because the capacity of these models to self-improve is, I think, a qualitative difference from previous general purpose technologies. Electricity was great, but electricity didn't self-generate more electricity. Fire didn't generate more controlled fire. But AI is already going to, within a year or two, generate be likely to take on the tasks that many of those doing the software for AI are now doing.

And so that self-improving aspect [00:11:00] represents a very big, uh, change. Could I just maybe dig into that a little bit? Because like what you hear people talk about, especially, you know, like in the Bay Area, is that this is not a technology. It's the technology. Sam Altman has said things like this will capture the light cone of all future value.

And so like, in what sense should we take that literally, but not seriously, seriously, not literally, do we think that this kind of technology is qualitatively different, or it is yet another kind of technology that will increase productivity, enhance human wellbeing, and things like that. I think it's possible that this is going to be the most important technology that ever got invented. I think that it's an even money chance that when historians write about the second-fifth of the 21st [00:12:00] century, Russia and Ukraine and Donald Trump and Xi Jinping will be secondary stories.

And that the real story will be the dramatic discontinuity that was associated with human-like intelligence being achieved by non-humans. I think that in a nearer term thing, it may not be unreasonable to think that this is to the Internet, as the computer was to the calculator. The calculator was a very important thing.

The computer was a much, much more fundamental thing. So, it seems to me there are lots of prospect for thinking that this is going to be [00:13:00] very, very important. There are, I think, legitimate questions about what the constraints are going to be, and we don't really know. I think it is clear that there are substantial gains to be had simply from scaling without any innovation in design, and without anything more than mobilization of data that has not yet been mobilized.

Relative to what people thought a year ago, I would say that a reasonable best guess, as I understand it, is that diminishing returns is likely to come somewhat faster from scaling, but [00:14:00] that modifications of the technology to make possible longer chains of reasoning and more human-like reasoning processes I would say any positive surprise in that, I think has been far larger than any negative surprise in the efficacy of scaling.

And I think there has also been a substantial progress and substantial surprise in what one might think of as compression or distillation of these models in which it's possible to do more with less building on a very large model than one might previously have supposed. [00:15:00] I think it's also important to understand that the history of profound technologies is that the initial focus is always on how they can perform previously defined tasks better than they've been performed before.

But ultimately, the real impact comes from the definition of new tasks that advance progress. The first cameras, moving picture cameras, were put at the back of theaters. To record plays. And then people realized there was much better things to do with moving picture cameras. It was originally envisioned that there'd be a market for only five mainframe computers.

Originally envisioned that there would be a [00:16:00] market for less than a million phones that people carried around with them. And so, I think that we surely have not exhausted what is going to be possible with software that is able to act as people's agents. Can I follow-up on something there? I actually am surprised that in your estimation that essentially everything will be footnotes to AI.

I think, Raj and I tend to agree with that. But I wanted to follow-up, like play that forward and get you to help us think about the consequences of that. So, if we have human-level intelligence, one immediate implication, and you can definitely correct me if this is wrong, is that the marginal cost of everything essentially goes to zero.

And I've heard folks like Demis Hassabis talking about the idea of radical abundance in the age of AI. That because we have these super intelligent machines who can do so many things, we will all kind of live in this utopia. But like, one thing that I go to is like [00:17:00] radical abundance in and of itself is not a purely good thing.

Like we have radical abundance of calories and that has been a public health disaster by sub measure. So is there something about the radical abundance idea of AI outside of the safety concerns that we should be thinking about proactively, that having a radical abundance of intelligence may actually have downsides that we aren't really appreciating yet.

Sure, look, I think you can overdo this idea of radical abundance. There's still only so much beachfront property. There's still only so much copper in the ground. There's still only so much of a whole range of goods that have some inherent scarcity to them. Some substantial part of what people value is goods that derive some of their value from their exclusivity. [00:18:00]

Not everyone can go to what's regarded as the top school. Not everyone can eat at what is regarded as the coolest restaurant. And so, the human instinct to compare and the human instinct to want things whose value is determined importantly in relative ways, both mean that this idea that we're all gonna abound in ecstasy with every need met and nothing to strive for does not strike me as being a plausible rendering of a place that the world is likely to get.

I'm not sure that the right, I don't think the right paradigm for [00:19:00] thinking about obesity actually has anything much to do with the universal abundance of calories. I think if you look at the upper 75% of the American population, there was no substantial difficulty in affording an adequate caloric intake 50 years ago and yet levels of obesity have risen. So, I don't think the right way to think about that phenomenon any has anything much to do with the abundance of calories. I think it has to do with changes in lifestyle.

It has to do with marketing practices and the design of products that are in various ways addicting, but I think linking and those raise all kinds of issues of consumer protection and [00:20:00] paternalism and how society should be

responding, uh, to all of that. But I think thinking of that as a problem of abundance is not actually a helpful way of thinking about obesity.

So, I tend to be a person who believes that there should be a strong presumption in favor of things that give people things that they want to have, unless there is compelling kind of downside. I think there is potentially, my guess is that it is a long way down the road, but I think there are important issues raised by questions of human satisfaction [00:21:00] and what the role of work is in thinking about human wellbeing.

And on the one hand, I very much think that we should probably be thinking about what it is people are going to do with all that time that is available. You know, I have never seen a systematic study of the experience of those who inherit great wealth, and therefore don't need to work to support themselves, and indeed can't influence their command over the ability to purchase things very much with any work that they're able to do.

And I'm not sure that those lives are on average more satisfying lives than those who [00:22:00] are less apparently fortunate in their inheritance. So, I think the question of purpose amidst abundance is a potentially large question. I have a bit more reservation about the UBI concepts that some in the AGI community are enthusiastic about for those kinds of reasons.

I think that's great. I want to switch gears just a little bit from the capabilities and the implications of the models. To their oversight. So, I want to zoom in on November 2023. This is just about a year ago. I think I got an alert on my phone that Sam Altman had been removed as CEO of OpenAI. So of course, over the weekend, I was glued to my phone.

We are using these models every day as researchers. We're using them in clinical studies. There's pilots, there's big studies that are underway. And of course, patients and doctors are using these [00:23:00] models every day. I think at a scale that, that is still vastly underappreciated. So many of us have come to rely on the technology and they've become central in our lives already.

And so particularly in a sensitive domain like medicine and health care, this really got me thinking about what this temporary, we know that Sam Altman came back pretty soon to OpenAI and you were appointed to the board of directors, but this really got me thinking about the sort of stability of these, uh, corporations of these companies, as we think about how we can apply these models in medicine and as they're being used.

And so, you, as part of that period in November 2023, you joined the board of OpenAI, and my question for you is what made you say yes to that job? Raj, in the wake of the corporate governance transition, I'll call it, at OpenAI, I was approached by the people who were involved in forming new arrangements to [00:24:00] ask whether I would take on a position on the board. I think they came to me precisely because I was outside the situation and didn't have the extensive prior loyalties in any particular direction and was thought to be someone who could grasp the various issues, both technological and social, and who'd had a certain amount of experience between academic life, the private sector, and government with complex situations.

I asked myself, really, two questions. Did I think that I could make a contribution in this way? And did I think it would be intellectually fulfilling? And since I answered both those questions [00:25:00] affirmatively, I decided to take on that responsibility. And I've been very glad that I did and feel proud of the little bit that I've been able to contribute to OpenAI over the last year.

I should say to your listeners that the first thing that the new board did, or that the new members of the new board did, was review in a very extensive way, the circumstances surrounding that transition, millions of dollars were spent, tens of thousands of documents were reviewed, witnesses were questioned for many hours by a major law firm that had substantial experience with this kind of thing. And I can tell you with a [00:26:00] extremely high degree of confidence that none of the issues that were involved in the request for Sam Altman's resignation went to anything about the safety or the safeguards surrounding OpenAI products.

There were complicated issues of personality between him and members of the prior board that led to that decision. And there was a kind of collective judgment about how the institution could be best taken forward that led to the reversal of that decision. But you needn't worry or needn't have, with the benefit of hindsight, worried during that weekend about the [00:27:00] legitimacy or the quality of the products that you were using in your medical work.

Got it. Thank you. So, I think we want to transition now to the lightning round.

We always do this on each of these episodes. Larry, the rules are simple. We're going to ask a series of kind of rapid-fire questions. Some are very trivial. Some are less trivial. You can decide which ones are trivial, which ones are not. And our only guidance is that we try to have you respond in just one to two sentences, quick, rapid-fire reactions.

Does that sound alright? Are you ready for this? Sounds good! Alright. Okay, so the first question is: Is the portrayal of your meeting with the Winklevoss twins in Andrew Sorkin's *The Social Network* at all accurate? It is not [00:28:00] literally accurate, but it conveys the nature of that meeting. And I would only say this, you learn certain things as a university president.

One of them is that if an undergraduate is wearing a coat and tie on a Wednesday afternoon, there are two possibilities. One is that they have a job interview. The other is that they are an asshole. I don't think those guys had job interviews that Wednesday afternoon. Ah ha ha ha. Fantastic. Okay, maybe that might be the most memorable lightning round answer we've ever gotten.

Alright, I'm gonna hand it over to Andy. So that was great okay. So now we're gonna do a riff on Tyler Cowen's overrated underrated for you. Overrated or underrated: Arrow's impossibility theorem. Underrated. Underrated. So, it's interesting because I feel like most folks in my circle, if you don't know anything about economics, you know Arrow's impossibility theorem and you use it to shut down any attempt at consensus [00:29:00] making.

So, in what way is it underrated? I think it speaks to the difficulty in a, it speaks very, very powerfully to the difficulty of any kind of collective decision making, particularly in an increasingly complex and multidimensional world. And I think it is understood by a limited number of social scientists who understand it, but it is not as universally part of the canon of human knowledge as it should be.

And I should say, I also have a bias. Kenneth was my uncle. I didn't know if that was family allegiance, but I think that that was a very well stated case for the underrated nature of Arrow's impossibility theorem. Alright, the next question. Which is a harder job, President of Harvard University or Treasurer Secretary of the United States?

President of Harvard University, it's got a lot more [00:30:00] politics in it than working at the treasury in Washington given the, um, extreme decentralization of Harvard and of universities in general. Alright. If you could have dinner with one person dead or alive, who would it be? Probably John Maynard Keynes because he embodied the practical, intellectual, economics-oriented life that I have tried to lead, and he could express himself so cogently, powerfully, and eloquently.

Alright, this is our last lightning round question. Which is the best Harvard undergraduate house, and why is it Leverett House? I was a tutor in Lowell

House, so I'm gonna stick [00:31:00] with Lowell House. Excellent. Thank you, Larry. You have survived the lightning round. You've passed it with flying colors. So, we want to ask just one or two questions that kind of zoom out a little bit to wrap up.

So, you're probably aware of this essay called "Situational Awareness." Again, like Tyler Cowen and some other folks have really been big on this. It's written by this 21-year-old named Leopold Aschenbrenner, but it's an economic analysis of the next five years of AI. And in it, he makes the conclusions that power will be limiting that essentially scale is going to keep working and take us to AGI.

So, I don't have a question about necessarily the article specifically, but more, how useful are the tools of economics going to be for understanding the next decade? Are we, as ML researchers, we would say, are we in distribution or out of distribution? Can we make reliable predictions about what comes next?

Or is your sense that the next 10 years is going to be beyond our predictive [00:32:00] capacity from an economic perspective? I don't think of the test of economic analysis as being the ability to make predictions. For example, one of the great ideas of economics, the efficient market hypothesis, essentially has as its central idea that you can't predict the evolution of a future speculative price, because if you could predict it, it already would have moved.

And therefore, in a properly defined sense speculative prices are random walks. So, I'd reject Martingales. So, I'd reject the way in which you framed that question. But God, I think the principles of scarce resources, thinking at the margin, recognizing the importance of opportunity cost, understanding that there's no such thing as [00:33:00] a free lunch.

That incentives shape behavior. I think those principles are going to be as important as they've ever been. The contexts are likely to change in a variety of ways. AI and digital technology more generally promote economies of scale. They promote what economists call non convexities. And that's going to make the nature of the mathematical analysis different, and in some ways more difficult than it has been in the past.

It may make the pure invisible hand less effective in getting to the best possible outcomes. But there's nothing that I see that would suggest that economic analysis is not going to have impact. And I would rather expect [00:34:00] that with more actors, more perspectives, some of more competition, some of the forces on which economics tends to focus are likely to become more important.

While everybody likes to criticize economics, I am struck by how much the other social sciences increasingly emulate economics and emulate methodological approaches that economists take, I think to their substantial benefit. I think just one last question here that builds directly off of that. I think Andy brought up a few moments ago, Arrow's impossibility theorem.

So, maybe this is a good one to close on. So, you know, I think a lot about how these models are trained. And there's this kind of heavy compute phase where they're doing, they're trained essentially to do next token prediction with massive corpora of data, but then there's this kind of lighter compute, but very [00:35:00] important phase where humans are brought in to label examples of good or helpful outputs and then to rank outputs.

And so, this sort of transmutes the question of what values are embedded in these models to whose values are embedded in these models. And I think a lot about what those human values are that are embedded in these models. And I think a lot of work has come out of the United States, but we're seeing increased activity outside of the U.S. as well in building some of these models.

And so again, thinking about what Arrow's impossibility theorem teaches us generally we have a few different parties in medicine and health care, right? We have the payer, we have the patient, we have physicians, and the question that I really, I think about a lot, and maybe you can give us some sort of parting words about this is, what are lessons about studying values and preferences?

Again, I think in economics, this is very, very old and has been done for many, many decades. And my guess is that we're going to start borrowing some of those methods for evaluating, for thinking about how we influence what comes out of AI models. But what [00:36:00] are the lessons from thinking about values and preferences from economics, including of course, Arrow's impossibility theorem,

to help us think about what human values are embedded in these AI models. It's a very deep question and I'm not sure I can give you a good answer. My instinct, Raj, is that for a very large number of the questions that one's likely in medicine to be looking to AI models for the values are issues are likely to be somewhat secondary.

I've said of AI that House as portrayed in the TV show was an enormously powerful figure who was able to contribute a great deal. That in some sense, what he [00:37:00] did is the kind of thing that an AI system will be able to do

before an AI system is able to hold a patient's hand as they're fearfully going into surgery.

And the ability to make more accurate diagnoses more quickly, and more accurately, and more precisely, suggests steps forward in the treatment of a patient is not something that I think is fundamentally more difficult around values. When there are judgments that are reached, then there are human choices that will have to be made.

But I think that is likely to be something that remains in a human domain for quite some time to come. So, [00:38:00] I rather suspect that, um, we're going to have some substantial time before we're going to have to face the issues that you are describing. Amazing. I think that's a great note to end on. Larry Summers, thank you so much for being on *AI Grand Rounds*.

Thank you so much. Thank you.

This copyrighted podcast from the Massachusetts Medical Society may not be reproduced, distributed, or used for commercial purposes without prior written permission of the Massachusetts Medical Society. For information on reusing NEJM group podcasts, please visit the permissions and licensing page at the *NEJM* website.