THE BROOKINGS INSTITUTION

WEBINAR

CHATGPT AND THE FUTURE OF WORK

WASHINGTON, D.C.

Wednesday, March 15, 2023

WELCOME:

SANJAY PATNAIK
Fellow and Director, Center on Regulation and Markets
Bernard L. Schwartz Chair in Economic Policy Development
The Brookings Institution

INTRODUCTION AND TECHNOLOGICAL OVERVIEW:

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

KEYNOTE PANEL:

SUSAN ATHEY
The Economics of Technology Professor, Stanford Graduate School of Business

DAVID AUTOR
Professor of Economics and Margaret MacVicar Faculty Fellow, Massachusetts Institute of Technology

PRASANNA TAMBE
Associate Professor of Operations, Information and Decisions, The Wharton School,
The University of Pennsylvania

ALBERTO ROSSI (Moderator)
Professor and Director of the AI, Analytics, and Future of Work Initiative, Georgetown University

CLOSING REMARKS:

ALBERTO ROSSI
Professor and Director of the AI, Analytics, and Future of Work Initiative, Georgetown University

* * * * *

PATNAIK: Hello and welcome to this event on ChatGPT and the future of work. My name is Sanjay Patnaik, and I'm the director of the Center on Regulation of Markets at the Brookings Institution. This event today is done jointly with AI Analytics and Future of Work Initiative of Georgetown University, run by my colleague Alberto Rossi.

As you probably have seen in the last couple of months, ChatGPT large language models have been front and center in the media. And there is a lot of anxiety among economists and among the population of what this means for the labor market, for the future of work, for a lot of the jobs that we have been accustomed to doing. And what is interesting is, when we look back a few years it looked like that automation in blue-collar jobs, like self-driving car, self-driving trucks, would accelerate very rapidly, but to some degree, the large language models now have overtaken those advances in automation and pace and are really coming upon us very fast. And interestingly, a lot of those tasks that attribute in large language models can actually replace or can help augment tasks in the white-collar working space.

So today, we have some of the world-renowned experts on this topic here with us, thank you very much for joining us. We have, first, a brief introduction about the technological perspective from Anton Korinek, who is a Rubenstein fellow here at Brookings. And then we have a keynote panel with three outstanding experts, Susan Athey from Stanford, David Autor from MIT and Prasanna Tambe from the University of Pennsylvania, moderated by my colleague Alberto. I hope you enjoy the event, and I'll hand it over now to my colleague, Anton. Thank you.

KORINEK: Thank you very much, Sanjay and Alberto, for organizing this topical event. And welcome, everybody. So I will start by sharing my thoughts on large language models such as ChatGPT, and how they contribute to cognitive automation. And let me start just by saying, and this field is moving so incredibly fast. So if we go a little bit through the history of AI over the past 15 years, there was a paradigm during the 2010s that I want to call the deep learning paradigm, and that had a large impact on our world. That came up with lots of impressive applications. But during that paradigm, there was still a category difference between human and artificial intelligence. In some ways, what we are now seeing is a new paradigm, the paradigm of the 2020s, the paradigm of foundation models underlying to generate AI applications that we are seeing. Now, this new paradigm, of course, builds on the deep learning paradigm, but in some ways it's also qualitatively different. In some ways it feels eerily humanlike. We have these huge models with hundred billion parameters and more, and their size just keeps growing and it's already quite close to the complexity of the human brain, which has something like 85, 90 billion neurons. And the leading category of this foundation models are large language models such as ChatGPT, Cloud, the new Bing, Google's Bard, and since yesterday we have the latest addition, which is GPT-4.

So I felt I should probably start by just sharing a couple of thoughts on ChatGPT-4, since it has just been released yesterday. I'm showing you here one of the main charts in the paper that introduced the bot, you can't really see the labels, but let me describe. The blue bars reflect the capabilities of the previous version of Open AI's GPT-3, or 3.5, on a wide range of standardized tests such as the LSAT, GRE, and so on and so forth. The green bars reflect the new capabilities that the system that was just released yesterday, GPT-4, can achieve. And the scale on this goes from 0 to 100% and essentially reflects the percentile among human test takers, which this system would be able to achieve. So one thing that you can see here is, I will point out where my mouse point is right now, the LSAT, right here in the middle, GPT-4 ranks in the 85th percentile of all LSAT takers, which is quite impressive. If we look at the GRE Verbal test, which a lot of people use for graduate applications, it already ranks in the 99th percentile of all human test takers. So, this is really a pretty impressive performance. You may have seen on social media, there are a few people who say that even though GPT-4 is a significant advance, and is clearly more powerful than the old models, it's a little bit underwhelming. Well, my interpretation of this is that we need to distinguish between the form and the substance of what these models are producing. Last November, when the first version of ChatGPT based on GPT-3.5 came out, it really wowed people because of its form. It was the first publicly available AI system that could produce high quality and coherent content in really well-crafted sentences on almost any topic. Now GPT-4 has a pretty similar form, it also produces well-crafted sentences, but the level of depth in which it can go in terms of content is significantly faster, and I think that's really the difference. So we are already used to seeing this new form of ChatBots, but GPT-4 goes a level deeper and can produce more insightful content and ultimately that will presumably also have a larger impact on the world. So what are all these advances driven by? And here, in this slide, I'm showing you Moore's Law, which has been going on for more than a half century, and which, essentially, captures the observation that advances in computing have been making computer chips twice as powerful every two years. So what does that mean? If you have a chip that's twice as powerful every two years, you have exponential growth here. And after 20 years, you have capabilities that are scaled by a factor of 1000. However, what is powering these generative AI systems is actually proceeding according to a much faster rate, and I have pulled this graph from a paper, which shows that the amount of computational power, or compute, that the top AI applications over the past decade have been using, has in fact grown by a factor of two every six months. That means it quadruples every year. It grows by a factor of a thousand every five years. So that's significantly faster progress than what Moore's Law suggests. And that's behind these recent advances that we have been witnessing for the past few months, including yesterday, when the new version, GPT-4, came out.

So how are these systems trained and how do they work? They are trained through a process called self-supervised learning. The system is fed vast amounts of data, and during the training process, the objective that it is asked to accomplish is simply to predict the next word in any given sentence structure. The next word, which is kind of blinded out for the system. It sounds like a pretty simple, not particularly impressive task, right? But the impressive thing is that based on this training, really advanced capabilities have emerged over the past five years. So these large language models, they can suddenly write coherent sentences, not only predict the next word, but write full sentences. And since November, they have become really good at writing coherent paragraphs and even essays. And the latest version, GPT-4, can of course write significantly longer texts than the previous ones.

And there are a whole bunch of capabilities that suddenly emerged that actually surprised the creators of these systems. They didn't necessarily expect that, but they found out that, oh wow, when we train these systems on more data and more compute and we create larger deep learning networks, these systems can suddenly translate. You can suddenly perform logical reasoning, they can do math, they can be creative, and they can also write computer code. So all these capabilities have emerged once the systems had sufficient complexity, and I think the best way for understanding this is that during the training process, the systems essentially develop a world model, a model of how the world works, because if you better understand the world, you can better predict the next word in any given sentence. And that world model can be applied to a very vast range of different tasks. So in some ways, this realization, that next-word prediction can produce these, what seems really quite intelligent outputs, it really forces us to reevaluate how we think about the human brain. One other thing is that people believe that there is widespread capabilities overhang. So what does that mean? That means these systems actually may exhibit far greater capabilities than what we currently know, and nobody has tried it out yet, and therefore, we don't know what other things the systems can do as well.

And the final thing about the workings of these systems that I want to emphasize is that there are fairly predictable scaling laws. So over the past five years, we have kind of come to understand if we throw more computational power at these systems, and if we increase the training data, if we increase the number of parameters and so on, how will this affect the outcomes? How will this affect the results that they produce? And progress is quite predictable. So if we continue to double the amount of compute that we are throwing at these systems for just the next few years, we can kind of already predict what progress will be taking place in the capabilities that they have. And those are pretty amazing capabilities. Now, how should we think about these systems more broadly? We talk to people. There are kind of two camps on the significance of these large language models. There is one camp that I want to call the camp stochastic parrot, or advanced autocomplete. And that camp essentially says, well, these systems, they are just dumb statistical engines, they recognize word frequencies and they can autocomplete things, but they have absolutely no inherent understanding. The second camp is, oh, wow, these systems are on the path towards human level artificial intelligence. And you can see fierce discussions between these two camps.

Well, I think it's easy to both over and underestimate these systems at the same time. The interesting thing is that their capabilities just work fundamentally different from ours, and in some applications, they can produce really amazing outputs, and in others, they can fail horribly, embarrassingly, and it's so difficult for us to relate to that because our minds function quite differently. So people say they have a different capabilities surface. So I want to continue my discussion of these two camps, and instead of saying, well, the truth lies somewhere in the middle, I actually want to suggest that both camps are right, at least up to a point. If we look at the camp stochastic parrots, they're absolutely right that these highly sophisticated statistical systems have significant limitations, the length of their prompt is limited, the drift in the outputs they produce, their training data is kind of outdated. Then they also tend to hallucinate, and sometimes they're really not grounded in our ethical values. On the other hand, the camp, this is the path towards human level AI, observes their capabilities. There are, for example, CEOs on record that state that their cognitive ability is significantly higher if they use these systems, and I'll read a quote here, "Anybody who doesn't use this will shortly be at a severe disadvantage," according to the CEO of Coursera. There is also a growing number of academic studies that point to the productivity gains that these systems deliver. It points to gains of 20%, 50% or more for certain categories of cognitive workers. So that's really quite impressive.

Now, what are the capabilities? And for that, I want to draw on a recent NBER working paper of mine, in which I have kind of summarized 25 use cases that I have grouped into six categories, and they range from ideation to writing assistance to background research, coding, data analysis and even math. And let me tell you, GPT-4 has really made significant advances in its mathematical capabilities compared to the previous versions. In the interest of time, I will jump over the list of 25, but let me refer you to this NBER paper of mine if you're interested. It's simply entitled “Language Models and Cognitive Automation.” I apply to economic research, but it really applies to any cognitive work that you may be performing. So I'll jump over this. Let me highlight, perhaps, the importance of prompt engineering. So this is actually a really important new task that we were all not aware of six months ago, but now we suddenly have a new job that's called Prompt Engineer. And it's in some ways a mix between programming and using natural language. It is in fact programing in natural language and the performance and capabilities that we get out of these systems depend a lot on how good we are at this prompt engineering. So one way of thinking about it is that it induces the model to shift into the kind of desired latent state of its world model to produce the outputs that we want to see.

Now, what are the short-terms lessons of this cognitive automation? So right now, and yesterday, we just had an increase in those capabilities. They are highly useful as assistants and as a [inaudible]. They can help us automate micro-tasks, like little things here and there, that we do throughout our workdays, if we are cognitive workers, and they can deliver significant productivity gains. So what's the economic advice in this kind of world? David Ricardo taught us already, more than a century ago, to focus on our comparative advantage, and that means we will really have to change our workflows to optimally take advantage of these new systems. These new generative AI systems are really good at generating content, so in some ways that may be devalued, whereas we humans still seem to be better at discriminating content, that's complimentary, and that will allow us to succeed. And we can also provide feedback for organized projects and so on.

What's the medium outlook, medium-term outlook? Well, these systems are becoming better and better, and we could just see that yesterday. And they will probably also be adapted to a lot more specific use cases. So for example, OpenAI announced yesterday that it already has partnerships with ten different companies on specific use cases for GPT-4, and they range from Morgan Stanley to the Khan Academy to the government of Iceland. And incorporating these new systems into the economy will take a lot of time, but it has the potential to significantly restructure how the economy is functioning. And then, the role of humans in many of these cognitive tasks will likely decline. And in a lot of things, we may increasingly turn into rubber stampers with a human veneer. So I think that's a development that we should expect in a lot of the cognitive tasks over the next five years, perhaps decade. So this really creates the potential for cognitive automation.

And perhaps one last question, in which I want to spend a slide, is how does this new cognitive automation differ from the traditional physical automation that has been basically all the rage for the past 200 years since the Industrial Revolution? So what's new? We have been automating tasks and jobs ever since the Industrial Revolution. Well, one significant difference is that it affects a new category of workers, cognitive workers. A second really important difference is that it produces outputs that are non-rivalrous, that means it can be rolled out very fast. This is something that can be copied and used by 100 million people after one month, as we observed after the original rollout of ChatGPT. And then finally, it also chips away at our last comparative advantage versus machines. So having said that, this will raise lots and lots of new interesting questions. This new area of cognitive automation will have big effects on labor markets, education, technological progress and ultimately, social welfare. And as of right now, our human brains, perhaps enhanced by these language models, are still the best technology available to answer these questions. And with that, I want to hand over to our panel, but to make the job for them a little more interesting, I actually want to briefly demonstrate the new capabilities of ChatGPT powered by GPT-4, and I want to ask the system to suggest a few questions for our panel, so let me write here to illustrate how these systems work. Please suggest a few questions that I could pose a panel of world-renowned experts at a Brookings event on ChatGPT and the future of work. And so when I press generate here, we will see what the system says and it immediately comes up with a whole bunch of questions. I will just let that run and hand the microphone back to Sanjay.

PATNAIK: Actually back to Alberto.

ROSSI: Okay. Well, thank you very much, Anton, for the fantastic introduction. I'm super excited about this wonderful panel. Before we get started with the panel, let me just introduce the panelists, starting from Susan Athey, Professor Athey is the economics of technology professor at Stanford, GSB. She's an elected member of the National Academy of Science and is the recipient of the John Bates Clark Medal, awarded by the American Economics Association to the economist under 40 who's made the greatest contributions to thought and knowledge. As one of the first tech economists, she served as consulting chief economist from Microsoft Corporation for six years and has served on the boards of multiple private and public technology firms. She was a founding associate director of the Stanford Institute for Human Centered Artificial Intelligence, and she's one of the founding directors of the goal of Capital Social Impact Lab at Stanford, GSB. Susan is currently the chief economist of the Antitrust Division of the Department of Justice, but she's here in her professor capacity. Thank you very much, Susan, for being here with us.

Second, we have David Autor. David is the full professor in the MIT Department of Economics. His scholarship explores the labor market impacts of technological change and globalization on job polarization, skill demands, earnings level and equality and electoral outcomes. David has received numerous awards for both his scholarship and his teaching. Most recently, David received Haynes 25th Special Recognition Award from the Haynes Family Foundation for his work, transforming our understanding of how globalization and technological change are impacting jobs and earning prospects for American workers. Thank you very much, David, for being with us.

And third, we have Prasanna Tambe. Prasanna Sonny is an associate professor of operations, information and decisions at the Wharton School at the University of Pennsylvania. His research focuses on the economics of technology and labor. Recent research projects focus on understanding how firms compete for software developers, how software engineers choose the technologies in which to specialize and how AI is transforming HR management. Much of his work analyzes data from online job sites, career platforms and other labor market intermediaries. Well, thank you so much Prasanna for being with us as well. Now, before we get to the questions, I would love for each one of you, and maybe kind of talk a little bit about your background and your research interests around ML/AI and the future of work. Maybe we can start with Susan.

ATHEY: Great. Thank you so much and thanks for that kind introduction. So several of the things you mentioned in my bio are really the origin story of my interest here. I was first exposed to AI and ML starting in 2007 when I was working on the search engine at Microsoft, which itself was in a very nascent form. And it was really a chance to kind of see the canary in the coal mine for what was coming, as the search engine was both one of the first really large-scale production and effective AI systems, but also, it was something that was having a big impact on society where we needed to think about how to guide it, measure it and make it better. You know, the search engine, at the time was kind of a composite of a lot of different, relatively simple prediction algorithms that were each accomplishing a relatively narrow task. And then they were all composed together to produce the search results that you get, and I'll talk a little more about that later in the panel. But what I learned from there is that one of the big weaknesses of AI and ML, especially at the time, was that it was really doing mostly pattern recognition. And as a social scientist, you know, most of the empirical work that I had done, and other social scientists did, was about counterfactuals and about doing cause and effect. What would happen if Affirm raised price? If we raise minimum wage? If we change a policy? If we give different information to our citizens? Iif we rank results in a different way? Those kind of counterfactual questions about alternative worlds. And so that really sparked a research agenda for me that was trying to both make machine learning and AI better and smarter and, behind the hood, really understanding some of the things we already knew in the causal inference literature about mistakes you can make when trying to accomplish tasks. If you don't think about biases in the data that was generated in the past and you can draw incorrect inferences about the future. And I also worked on trying to bring in some of the amazing computational tools and all of the amazing pattern recognition. It's, you can think of that as an incredibly important helper to drawing inferences about counterfactuals and cause and effect. So that's really been the core of my research agenda.

Then in terms of, like, preparing for this new ChatGPT phenomenon at Stanford, we have this Institute for Human-Centered Artificial Intelligence, and one of the projects led by my colleague Percy Lang, was to understand foundation models, and ChatGPT is a foundation model, it's a general purpose technology that is going to impact, and people are going to build on top of it, it's going to impact lots of things around it. So we've been trying at Stanford for a few years to understand both the basic science of it, but also think about how to guide it in an interdisciplinary perspective. So one of the things that I've been working on is, because we can't, it's hard to work with these really big models, I've got a little laboratory where I built the foundation model around worker transitions and job transitions and careers. So that was a way to study the future of work, but also study foundation models at the same time. So I built a foundation model out of 23 million resumes and then did what's called fine-tuning to a representative survey data set to make it less biased. And I've been working on methods to try to see how we can, in the training and in the application of these models, reduce bias. And it's a tiny model compared to the huge models that we're thinking about, but because it's a more manageable laboratory, it also allows us to explore some of these questions. And then finally, I mentioned my goal of Capital Social Impact Lab and really trying to translate some of these technologies for social impact applications. The social sector really doesn't have the capability and the people and the money to figure out how to adopt a lot of these technologies. So we've been trying to put them into the field, at the same time, trying to advance the science of measurement and safety and ethics and impact as we go along. So that's what I've been bringing to this so far.

ROSSI: Thank you so much for the fantastic introduction. So we move on to David.

AUTOR: Sure, I'll try to, you know. Anyway, let me not spend time saying "well I'll try to". I, so I've been, I actually have been a sort of computer geek my whole life. I started, I taught myself programming when I was in grade school, and I did work in that field. But then after college, I transitioned to the kind of NGO educational sector. I directed the educational program for several years at a nonprofit in San Francisco that did computer skills education for poor kids and adults.

And that was early in the computer era. In the computer, you know, sort of modern computer era in the late eighties and early nineties. And I was really interested in the impact of how the technology was changing, what skills people needed, who could, you know, advance, and what skills were amplified and which ones were made less relevant. And I did that also overseas and then I started studying that academically and so I've been very focused on the interaction between technology and work and how it affects opportunity and how it both complements and substitutes for the skills that people, you know, possess and bring and learn, and I think that's what's so much in flux right now. And I think, you know, for the last couple of decades, we've had a pretty clear roadmap of the way the technology worked and what things it could do and what it couldn't do and what it would take to get from here to there. And I just think that roadmap has been, you know, kind of blown off the map, and so it's a very exciting time. It's also, I think, a lot, there's you know, I think there's understandable anxiety, and my degree of confidence, any prediction I make is much less than it was a couple of years ago. So looking forward to this conversation.

ROSSI: Thank you, David, Sunny.

TAMBE: Yeah. Thank you so much for inviting me to participate in this panel. I'm Sonny of The Wharton School at Penn. Also, like David, have been a long time computer geek, I was a software developer for a few years coming out of school, and so, developed this interest in those particular labor markets, looking at how developers kind of surfed skills, how they chose employers and in the converse of that, of course, which many people are interested in today, which is how employers are attracting developers, how do they stay on the frontier, how do they get the, get top tier talent? So I've been interested in quite a while and understanding some of these ecosystems, understanding the matching process and the unusual supply characteristics of this market, and what that means for how well and efficiently, and quickly firms can adopt and implement solutions like AI solutions. I'm also the co-director of Wharton, we have a new center on AI analytics, analytics for business, and very, we're very much newly interested in understanding this collaboration piece, which of course, is a big question now that's going to, that is already emerging, that's going to continue to emerge around these new tools. Third thing I'll mention is I teach two related classes, one's on applied data science, and that class, we very actively have been using ChatGPT to generate code, and I've been asking students to understand, okay, what's taking longer, what's more efficient, trying to understand what's actually going on, that's been a fascinating process. And so I'm looking forward to this discussion quite a lot.

ROSSI: Perfect. Thank you very much. So just to kind of start the overall panel, what I would like to do is to focus on the ChatGPT, on the technology. So starting with Susan. So, Susan, you're clearly a huge expert in ML methods, can you give us a sense, I mean, Anton touched on it a little bit, but could give us a sense of what ML methods generally do and why, and how revolutionary ChatGPT is. So in other words, like, why is ChatGPT so much better than all the chatbots we've been interacting with on Expedia or on some of these airline websites that were clearly not working very well? Why is this working instead so well and kind of what are the difference in methodologically compared to the other bots that we've been interacting with?

ATHEY: Yeah, and I think, like, similar to David, this is like sort of the first time that I was really surprised in the last, you know, year or so. I mean, and I kind of, I was, I had a pretty front row seat to this stuff, but I was still just surprised by the performance. So just to step back a little bit, I mentioned earlier, like, the search engine and most of the, you know, productive large scale uses of machine learning and practice have really been very simple algorithms doing very simple tasks. And it's just pattern recognition. And I think one of the things, we've started to get hints of this in the last few years, that sometimes, just putting together pattern recognition in a more sophisticated way, or piecing together lots and lots of pattern recognition can be really surprising. So you can be very dismissive of the stochastic parrot or the prediction, which I've been more on the skeptical side, frankly. But now we're seeing that somehow combining things can give you something unexpected. But just to step back, like, so what was my, you know, ten years of speeches about, like, why you shouldn't be worried about this and why this isn't that transformative. You know, if you think about the most successful applications for many years, they were things like just prediction and classification. So, you know, you have a data set, you have some pixels of images, you have labels, it's a cat or dog. And you know, the fact that the computers could be really good at telling cats and dogs apart, it's not, like, transformative, it's just we got some better functional forms that were, that work better with things like pictures, and we got more compute, we got some better optimization algorithms and we were basically just doing the same conceptual thing we'd done before. And you're fundamentally limited by the size of your dataset. Like if you have, you know, and we got some bigger datasets, more cat and dog labels and we got some implicit labels from things like, you know, people clicking on stuff on the web that gave us some big datasets labeled datasets. But ultimately, you know, we're limited, and very few people have access to that data and it large scale and, you know, you're just limited by the data size, but it's fundamentally pattern recognition. And so, and then when you talk about, like, the chatbots that were used, actually whenever people tried to use AI chatbots, they failed. And so, almost all the chatbots that you used before, like, a year ago were completely preprogrammed. So just the big decision tree, if you do this, then it'll do that. And that's just because that was all that worked. It was safer and better and generally people didn't need to do that many things, and so you just had a decision tree. So, and the few attempts to put them in the wild were kind of spectacular failures.

So where, the breakthrough, really, is that it's actually gotten, it's gotten, the AI has gotten good enough that it's sort of past the acceptable threshold. And so part of that, I think that the intense overview helped understand the difference when I said, like, you can take think of like just one prediction problem, I have a set of images and a label, but even just predicting a paragraph of text, in one paragraph, you have lots of prediction problems. You can use the first word to predict all the rest of the words. You can use the second two words to predict the rest of the words. You know, you can take the first ten words and predict the next ten words. So you have lots of X's, like prediction, predictive variables and outcomes all in one paragraph. And then if you have all the text that's out there on the web, you have actually a whole lot of data, if your goal is to take an initial set of text and to predict what comes next. So that's pretty powerful. Still, people were trying to do that for a while and they were sort of limited by the functional forms they were using. And so some of the recent breakthroughs have just been we finally figured out some functional forms that were tractable and computable, but that really took advantage of how the words related to each other in a sentence. So fundamentally, all along, since, you know, the last 20 years, the basic thing in doing this text modeling has been to take text and find a lower dimensional representation of it. You think about all the pairs of words in the English language. That's a lot. If you think about all the paragraphs of words in the English language, that's a whole lot. And we try to, what we try to do is find lower dimensional representations of text that is useful. And this [inaudible] context aware representations have worked really well. But the second part of the breakthrough is really just been in the engineering that, you know, if you, the dimension, the true dimensionality is very high. And so you need a really, really big model with lots of parameters in order to do a good job. And we've just figured out the engineering and put the money in to buy all the compute, we've gotten, the hardware and everything else to be able to build a really big model. So conceptually, it's still the same thing. It's just better functional forms, better engineering, better compute and lots and just very, very, very, very, very big. And then being what is doing very, very big do? It basically makes things very contextual so it can factor in the style. Like write something in the style of an HBS case or of a news article. And so that context, you know, if you have a big enough model like that, that context can produce different output for those different scenarios. And I think that's kind of what's been so magical.

Now, I still, though, I mean, at some level, so it's like putting together pattern recognition, but it still is pattern recognition. Like it's still not smart, like you're not, the architecture of this is very, very simple and it's still just a big blob. You're not putting intelligence into it, you're not putting structure into it. Everything is just learning parameters of the big pattern recognition model. So the mistakes it makes also are kind of predictable. Like if it learns from Reddit chats, it's going to sound like a Reddit chat. If it you know, if it learns from romance novels and then you start talking like a romance novel, it's going to spit back what's in a romance novel. So, you know, it's only, it's still only pattern recognition. You know, if you ask it for papers written by me, it'll put things in the format of a reference, and the words sound like words on my CV, but they're completely made up. And that's because, you know the words. It understands this is a very likely set of words like you could fool somebody else, it looks like Susan Seavey, it's just not my kid. And and so, you know, but one of the things that's kind of interesting and we can maybe talk about that a little bit more later as well, is that, you know, when you go to, say, the search engine Bing and ask the same question, partly I think it's using a more advanced model. So we still have to test this out today. But the difference is but the, you know, the search engine actually knows what a article is because it's got a very complicated set of heuristics and decision trees. So it knows what references are and it knows where you find them. And it actually can give you factual information that it references. So one of the things I think that's going to be really interesting going forward is how people take the strength of ChatGPT, which is that it can like summarize information and get plausible scientific text in a style and put it together with other kinds of technology either on an either in kind of helping people with prompt engineering, very complicated prompt engineering that forces ChatGPT to answer the question in a better way or post-processing of the stuff that comes out to correct errors. Like it's not that hard to recognize a reference in a ChatGPT answer to have a classifier that says that looks like a reference and then to do a function call to look up, is that an actual article or not, and then not return it if it's not an actual article. So I think that's really part of the frontier, we have this big black box foundation model, which is doing pattern recognition incredibly well. But it's a it's not very easy to get inside and tweak some things. You know, this thing works at scale, exactly, because it's a general purpose thing. And so how we put stuff around it, on the inside, on the outside, how do we fine tune it to different circumstances will really, I think, be a part of the determinant of how pattern recognition, together with other technology, can lead to a really powerful result.

ROSSI: Hmm. Well, yeah, so this is great. But in terms of kind of the task, I think Anton didn't really have time to get into it. So what are the kind of practical examples that ChatGPT can currently do? I mean, we know that they can form sentences, it can, can it understand images? Can it, what are the sort of things that it can do, and what, instead, are the sort of things that it's not going to be able to do for quite some time, according to your prediction.

ATHEY: Yeah. So let me not just talk about ChatGPT, but just this like category of technologies. You know, we already see this in, like, little apps on our phone. We can generate images, you know, we can generate images in the style of something. We can generate text in the style of something and summarize, and those are really powerful tools. I think the one that's easiest and there's already some research papers out about this. Copilot, the GitHub Copilot has been there for a little bit and it's incredibly powerful, and anyone who hasn't tried, you know, you can, it's just such a, it's such an accelerant to be able to not have to stop and think about the syntax and, you know, be able to get code written out. Now, the interesting thing is it's still, you might imagine as a user or just if you read about it in the newspaper, that what this is doing is you've encoded the rules in the syntax of a programming languages and it's following those rules and you kind of somehow communicated what those rules were. Certainly the ChatGPT version is not doing that. It's just generating code that it predicts would be a likely answer to your question, like from things like "stack overflow" and other code that it's seen in the past. So it's going to make mistakes and it's really interesting when you use it because it'll give you code and then you paste in the error message and it says, oh, I'm sorry, you know that, you know you need to do this, this and that, and you think, Well, why did it give me code if there was, that it knew was wrong, but it, it didn't know it was wrong and it might have taken a long time for it to check and see if it was wrong. It gave me code that seemed likely. But then when I give it the error message, it's sort of doing pattern recognition for what people have said when people have asked questions about that error message. And so it actually gives a really good answer and it seems smart, but it's just a sequence of patterns. If it wasn't, it wouldn't give me buggy code. Now, you can imagine over time, on top of that, it's very expensive to actually test code, like, that's going to be a function call. It's going to take compute, it's going to take memory, it's going to take time. But you could piece these things together a little bit. And in the future, I can imagine that you will improve once you get the basics down from this, like, pattern recognition stuff that you can figure out some efficient ways to start making it have less mistakes and self-correct its mistakes or even like sometimes you see this now it'll write something and then it'll delete it and then write something again. And you sort of get the idea that it's got a layer of more expensive checks that it's doing that slow things down a little bit and then if something's triggered, it does some more compute and gives you a better answer. And again, that's sort of like what the search engine does. It kind of triages, you know, whether or not it's worth it to go and do an expensive computation. And so over time, that triage piece of it may get better. But I think the coding is really good. And we've seen estimates that it's really improved productivity. I mean, all my students are using it. I've used it. You know, it just really picks things up. And I think there's a lot of these research tasks that have been incredibly repetitive and incredibly frustrating, like, where you have to click on something and look at it and then you click on something else and it basically has the same information and you're trying to get unique information and you just have to keep digging through all this stuff, and the ability of ChatGPT to summarize information and not show you redundant information I think just supercharges any kind of research process. And then also, like, just this kind of generating rough drafts and getting better styles and so on, I think it's just kind of game changing. But I think a lot of this has been covered and written and Anton covered a lot of this as well, so maybe I'll stop there just because I think there's a lot been said on this topic already.

ROSSI: Okay. Well, thank you. So the second thing I wanted to touch on was the effect of ChatGPT on this labor markets. And I wanted to switch over too to, David. So from a labor market perspective, what, if anything, makes AI different from this AI, this different from prior generations of IT. Is it just more of the same but faster? Or does it change the nature of like human machine substitution or complementarity? And more importantly, if we don't know the answer to this, kind of, what kind of data would we need to, or evidence, would need to gather in order to answer this question?

AUTOR: Yeah, I think it is qualitatively different, it's not just more, and let me just step back a little bit to sort of both complement and substitute for what Anton and Susan said. So I think it's a little bit dangerous to make analogies about the number of neurons and so on and sort of say if it has the same number of neurons as the brain, then it's the same, I don't think that's what Anton meant. But others may have heard that, and that's not correct. You know, it's different from our brains. And I think it has capacities we don't have, and we have capacities that it doesn't have. And that's been true for all of the tools we've made for a very long time. And what makes AI different from prior generations of software is that all software that, you know, historically was followed the following, you know, kind of cookbook, you decided tasks you want to automate, you figure out all the rules, you specified them completely, and then the machine executes that series of steps without judgment or without, you know, adaptation or learning and carries them out. That was great for calculation, that's great for playing games, that's great for storing data, that's great for many, many, many things. But sometimes the machine did not make inference, it did not learn from context and didn't generate new things. It was in some sense, you could think of it as, what's the order I want to use, it just it's just executing. Now, of course, AI is also just executing. It's all deterministic, right? It's all, you know, atoms and molecules. But that's not a useful way to think about it anymore, any more than it is useful to say, you know, when I go home for dinner at night, that's just molecules moving from MIT to my house in Newton. It's true, but that's not a useful analogy.

So what's different here is that AI is able to effectively learn, instead of saying to get from here to there, you follow this series of steps, you say you started here, you end up there, you figure out how to get from here to there, and you can make inferences on that. And that means that the range of things that are subject to this tool is much broader than the set of things that we had to hard code to accomplish. Those tasks were routine in the sense they were quantifiable. This can do many, many non-routine things, and so the word generative, I think is very, very useful because it really is evocative of what's going on, it's sort of creating stuff that we didn't, we don't think ourselves of having programmed into it. It's also very hard to predict for the same reason. And so I think what we want to think about then is that we make lots of tools. We all use tools all the time. Most of what we do is accomplished with tools, whether it's getting from place to place, the computer you have in front of you, the writing instruments, the clothes you're wearing. We are surrounded ourselves by tools and we are completely feeble without tools. And this is a tool as well, but the question is what type of tool is it? And is it, and in particular, is a tool that complements our expertise and makes our skills more valuable, as many tools do, right, you wouldn't take a, you know, a pneumatic hammer away from a roofer, you wouldn't take a scalpel away from a surgeon, you wouldn't take a computer away from an economist, right, those are all their tools. And so in many cases, tools make our skills and expertise and knowledge and creativity more valuable. But then there are some tools that do the opposite. They actually commoditize what we have that was valuable, right? So if you were a London taxi cab driver and you had exhaustively learned all the streets and byways of London, and all of a sudden GPS came around, well, it's great for consumers, but your scarce skills are no longer scarce. And that's what we need to understand, that relationship, and also direct, right, so this technology is incredibly malleable, right? It can do many things, right. You can use it to build the Great Firewall of China. You can run the world's largest surveillance state. You can do surveillance capitalism if you want to, or you could use it for education, you can use it for medicine, you could use it to make people without medical degrees effective as doctors. And so to say what AI will do kind of misses our agency in the entire operation. AI will do what we invest in to make it do that thing. And so we ought to be thinking hard about what we want to get out of it. Now, of course, lots of people are going to do lots of things and it's cheap and lots of bad actors and, you know, people with profit motives, they may be good or bad. Many, many things will happen. But we have a shared interest in directing that technology in a way that will be complimentary to us, and therefore more advancing societal goals, helping us solve some of our hardest problems, like, you know, climate change, you know, nutrition, health and spending and doing less of just replicating human capabilities. Because let's face it, we have human capabilities. Those aren't scarce. Let's figure out something, the best way we can use it. Huge challenge. Thank you.

ROSSI: Okay. Yeah, but always continue on this kind of labor market aspect. So one thing that we've seen is this kind of an acceleration of the trends we were observing over the last 20 years, right, in terms of the kind of exacerbation of wealth inequality or income inequality. So, David, like, given that you spent the last couple of decades trying to understand how technology affects labor market dynamics, what are the biggest trends that you've observed the past 20 years that have changed in the recent past before because of AI?

AUTOR: Yeah, absolutely. So the thing that's, you know, what we've been doing a lot of with automation of various forms over the last several decades is codifying routine tasks whether they're office clerical, administrative tasks, whether they're repetitive motion tasks on the factory floor. And a lot of those really have changed enormously. The office of today is completely different from 40 years ago when people had typewriters, filing cabinets and calculators and staples, and those were basically, staplers, those were the main tools. They didn't have spreadsheets, they didn't have search, they didn't have word processing, they didn't have many, many tools. And as a result, there are many fewer clerical [inaudible] workers than there used to be, per capita, they do a very different job. They don't do any more typing and filing and sorting and copying, right? They fix problems. They resolve travel issues. They handle those, you know, God awful receipts and so on. So it's a different problem. And this has led to a lot of bifurcation of labor market. Actually, a lot of those were middle skilled jobs, they've been hollowed out, as a consequence, we have a lot of people who are highly augmented, who are doing professional, technical and managerial tasks, whether it's research, whether it's medicine, whether it's law, whether it's creative tasks. And then we have a lot of people who are doing in-person services that are hard to automate, but also are hard to make more productive as a consequence, are pretty low paid and used relatively generic skills. Many people could be a good waiter or security guard or cleaner with, you know, a little bit of training. And that means those will not tend to pay well. And so the challenge this creates is, well, what is the new terrain that will be opened up? What will be complemented now and what will be substitute? Will it be the case that basically all thinking will be done machines and all people will basically be building stuff, assembling, you know, handling the physical tasks of the world or, how quickly will, in fact, AI make robots much, much more capable because a lot of the challenges in robotics is not the actual movement, it's the calculation around the movement. So, I think, you know, if I had to put this in terms of two questions that I think simplify it, although probably too simple. One is, does AI, what skills does it complement? Does it like -- for example, Susan mentioned coding; I have some students, Shakked Noy and Whitney Zhang, who did a really nice study about writing tasks and showed that basically professionals who do kind of, you know, industrial scale writing, they save a substantial amount of time using ChatGPT, and it kind of brings the bottom up. It makes the less good people somewhat better. And, I don't think that's a universal, I don't think if you give it to someone who's illiterate, it would make them better, right, but I think there's a complementarity. So who is complemented? Is it only the best? The people who are the top of the field, is it the people who are in the middle, can it enable people with a good skill set to be more effective in many things? That's a very good scenario. And then the second question you have to ask is, well, what is this sort of [inaudible] jargon? What is the kind of elasticity of demand for all those things? So if everybody gets 50% more productive everything, right, that's good, unless we saturate the market with that, then all of a sudden things that were valuable when only a few people could do them are now so cheap and commoditized that in some sense they don't command much value anymore. Now, again, it's good for productivity, it is not good for distribution. And so I think, you know, so the questions are, again, just let me rephrase them and then pause. One is, what human skills are complemented and given more leverage, made more effective, you know, having some fundamental skills, then this tool allows you to do a broader range of a certain task. And that, I think is the good scenario, right? Where we take people who have some expertise and make them more effective at a broader range of things. And the other question is, how fast we saturate the market with those things such that even if we get, you know, you can do a broader range of things, they're now just dirt cheap or water cheap, which is the way we use water, and suddenly that's not such a good living.

ROSSI: Okay, perfect yeah, and kind of moving on the other side of the labor market. Not so like thinking about those people that actually develop these tools. Prasanna, I know that a ton of your work centers around understanding, kind of, the labor market for AI skills. So the question I have is like, where do companies acquire the talent needed to deploy these AI tools within the firms? Which companies are more successful? Which ones struggle the most? Is there, like, is it very difficult to acquire talents to design this kind of large AI tools that we see becoming more and more popular?

TAMBE: Yeah. So thank you. So you certainly know that AI talent has been scarce for a while, top AI talent particularly, and these most recent round of tech layoffs notwithstanding, we don't expect that to change substantively in the short run. I mean, top AI talent still is incredibly hard to find. It always strikes me when I talk to technology managers, not just HR management, just managers generally, how many of their decisions are focused on really the hiring piece. You know, not just the obvious ones, but the decisions they make around, for instance, how much to participate in open source communities and things like that. A lot of it has to do, ultimately, with the way they want to be able to compete for some of these developers, really, and of course they want to get paid, but they also want to be at the forefront of the community that are developing these. We we've talked about how quickly they are evolving, and how rapidly they're evolving, so it is really critical for these developers to be at the forefront. At least in my mind, it raises kind of an interesting question, which is, and some of my colleagues, including some of the ones on this call have written about, you know, things like superstar firms. But you have this sort of Matthew effect where you have these developers who are attracted to the top companies. And while this talent is scarce, it's hard to understand how, if you're behind the frontier, how to attract and compete with these workers, so that's one thing. I'd also like to mention, this might be obvious to people on this call, but the broader audience perhaps that I've spoken to, it's perhaps not so much, the first step, one step, of course, is acquiring AI talent, but it seems like firms are settling into the reality more that there's really a long pathway that a lot of successful firms have been going through, right, in terms of complementary assets. I mean, labor of course matters, but data which has come up several times in these calls, business process transformation, all of these things are difficult. And I think we're now hitting a phase where firms are saying, okay, this is hard and this is expensive. So even if you can find a way to attract this talent, it's no easy thing to have the type of data you need to train these models. And some of the firms that we see be successful at it have been working at this for years and even decades really, you know, starting from collecting consumer data in the dawn of the Internet, so to speak. So this is really a long pathway that goes well beyond AI talent is one piece of the puzzle which you have to have in place to get things right. I think the other force, last force I'll mention here, which is an interesting one, maybe pushing against this, is a lot of resources are being expended by tech firms, big tech firms, particularly at lowering the barriers for using machine learning, right, so there's a lot of interesting action on maybe automated pipelines going back to the release of deep learning frameworks for public consumption and then automated pipelines and so on. So I still think we're in a world, you need to have AI engineers, you need to have data scientists. But it's going to be interesting, I think, to watch how this progresses and how much this lowers the costs for firms to enter into this if they happen to have the data on hand.

ROSSI: I see. And so in terms of, like, the international landscape, so how competitive is the US compared to other countries right now? Do you expect this to change in the future? And if you think that over time we may kind of lose our competitive edge, do you think, what policies can the US implement to remain competitive?

TAMBE: Yeah, that's an interesting question. You know, I think if we'd had this conversation 18 months ago, the general tone of that answer might have been different, in the public at least, I think there was some amount of public anxiety or angst, especially with respect to machine vision tasks and those sorts of things about America's position, particularly with respect to China and their tech ecosystem, and that with their large tech firms have been doing I think maybe two things have changed since then. One is large language models, of course. I don't think it's a controversial opinion to say that the U.S. is viewed as a leader in large language models, and there is certainly other language models in other parts of the world that have done well. But I think there is a conversation now happening about how the US gets so far ahead and what can we do to catch up. So I think from a large language model perspective, that's been changing the conversation. And then of course the relationship between the Chinese government and their tech firms has been changing in ways over the last year or two, which I think has been challenging the relative balance, I guess, in terms of rate of innovation in those two sectors. You know, in terms of the other part of your question about what we can do, I think most of the top AI talent is probably still produced here, partly because of the complementary resources you need to learn these skills, there's the evergreen question, of course, about our ability to retain that talent, right? Those don't go away. There's certainly investments in basic infrastructure, I think, that can go a long way for top AI talent. The last thing I'll mention is there's a sense in which it certainly works, you don't have to look much further than a place like Toronto, right, where you've seen industry, government, academics kind of come together to develop a powerhouse in terms of an AI ecosystem. And so it certainly seems like there is a role here for these these these different pieces fitting together.

ROSSI: I see, so we have a number of questions that are coming in from our audience, and I would like to throw them at all the panelists. One, I think it's kind of related to ethics. So the question goes as follows, how can organizations ensure that they are using ChatGPT in an ethical and responsible manner? And what steps can be taken to prevent biases or other unintended consequences from arising? So I don't know if any of you want to take it, I mean, this is really to, you know, kind of ethical consideration, aggravating biases.

ATHEY: Yeah, maybe I can start with that, and it was a while ago, but to pick up on one of the things I mentioned in the introduction, this is something I've been thinking about. There's actually a close relationship between thinking about causal inference and thinking about fairness and bias, because if you have a really high quality model of the world that will tell you what would happen if something changed, then that high quality model is less likely to be biased in a lot of the ways that we understand it. A lot of, you know, biased algorithms can come from sort of using shortcuts that are proxies for true underlying causes. And one of the reasons that you see these problems in, like, classical big prediction models is that you're kind of throwing in a soup. And that's the beauty of them. You can just throw in a soup of predictors and let the model work and then you get an outcome. But it and it's actually been hard to kind of muck with the interior of those black boxes without sacrificing performance, both computational and predictive performance. But then you get, you know, correlation and not causation and models that don't transfer well to different environments or it's not performing well with the world changes. So these are pretty closely connected. And I have some writing about this, for example, an overview paper in Nature Machine Intelligence that's just about prediction problem. Now it gets all that much harder when you come to these large language models because you know the state space is just so large. So at least like if I'm reasoning about bias in image recognition, I can sort of talk about the features. And yes, you're using a neural net to represent the features, but there's like a pretty close connection between what you can see and like whether there's an ear or whiskers or what's the color of your skin, if you're talking about humans. When you go to these large language models, the representations of the text are so complex and so high dimensional and we don't really, like, understand the ground truth well enough that it can be very, very difficult to even start to wrap your head around it. And then I also want to remind that the beauty, the reason you can take this architecture and apply it to just all the text in the human language and it's out there and just let it go is because it's not sort of carefully crafted in a way. So when you go to try to think about debiasing it, it's like, how do you even start? And that's also a problem, by the way, that, like, a government that wanted to censor or even if you wanted to, you know, try to make things safe for kids or safer schools or, you know, safe for a company to use in their chatbot, it's very hard to like, make it behave reliably in any way or make it withhold information, because it's not like there's like just a little button you can push. Like it's just a big representation of language, and nobody, including the people that make it understand it. The only way you understand it is to test it. And the and the space of the testing is so large. So it's very, very hard. So I've been playing with this in these labor models and I've been looking at things like the gender wage gap. And so that's like a very simple problem because I want to make sure, if I want to say that there's a gap between men and women's wages, I want to make sure that that gap is real and not just an artifact of the fact that I used a low dimensional representation of people's histories. But that low dimensional representation was, say, more accurate for men than women or, you know, gave a different answer for men than women in some way, or didn't fully account for salient aspects of of the women's histories. So I'm trying to make sure that what my model tells me is a wage gap is really a wage gap. So we've been actually tinkering with the architecture of it's the same architecture that's behind the LLM models to try to de-bais them in just one dimension at a time. And we've made some progress and we have some theorems and we have some algorithms that kind of make sure that it's not creating artificial wage gaps in one dimension, but that's a single binary variable gender. Imagine now, like, you're trying to like de-bias an entire language model in, I mean, in just ways that you just can't really comprehend. And the performance, I get, like, only small performance gaps if I just want to be de-biased in one single dichotomous variable, but trying to like do it globally, you know, it's just very, very hard. I'll mention one other element of that. So part of it is like changing the architecture and changing the objective function, but that's going to be hard to scale. A second thing is this fine tuning, and I alluded to that, but let me talk a little bit more about it now because it really affects some of the applications you would do. So there's hundreds of papers. This is relatively easy to do. What you do is you take a trained model, somebody spent a whole bunch of money and a whole bunch of compute and a whole bunch of engineering time and trained a model that produces a low dimensional representation of the English language. Once that's trained, you can put some words in and it'll spit some stuff out. What you can do next is take that to a much smaller dataset. Maybe it's a proprietary dataset of data with confidential information, you know, personal information, something confidential for your firm. But you can take that big model and fine tune it for your smaller dataset, and that's going to do a couple of things. It's going to make it give answers that are much more salient to your environment. In my case, I took a large resumé dataset that's biased towards rich people and high educated people, and then I fine tuned it on a representative survey of American workers, and that's going to bias it in another way. It's going to make it fit better for the whole set of individuals we have in the workforce. And so that's another way to de-bias. It's a different approach and it has different properties. But you can you can sort of find a corpus that you think is okay in the dimensions that you care about and then fine tune the model in that direction. Then a final thing that people have done, and this is happening in ChatGPT, is that you you incorporate human feedback. So you get people to rate like, is this a good or bad answer? So you could have people rate is this biased or unbiased? And then you add that in to the objective function. So it's that human feedback component. So those are those are basically like, you know, three different ways you change the architecture in a way that you anticipated debiasing, and that's one. Two is you fine tune to a more representative dataset. And three is you incorporate human labels, but all of those are inherently not scalable fully, because you're limited by human feedback, you're limited in the number of dimensions you can try to de-bias, you're limited by the size of the corpus of unbiased data and in all the ways you might have conceived of whether that data is biased. So this is going to be an ongoing research area and an ongoing corporate innovation area. I would say this is like, we've got 20 years of work ahead of us to get this right.

ROSSI: Yeah, perfect. And also kind of very related to what you were mentioning. So we know that a lot of these AI/ML tools are trained on customer data, employee data. And so a lot of people are concerned about data privacy and security issues. So in the context of kind of new technologies in the workplace, what steps can organization take to protect their employees or customers personal information? Does anybody want to take this?

ATHEY: So I can start and then maybe, you know, others can layer in. But what I expect to happen is that people will produce a service like, look, I can make the fine tuning programs, you know, with a Ph.D. student in MyCompute. So basically that's not a big impediment. And there have been people already fine tuning these new models on smaller compute. So I think that problem is almost already solved, technically, and it just needs to be productionized. So one service I expect to come up imminently is services where companies can use this on their internal data, their email, the documents they create and everything else. But part of the service will be that it's not shared back and people will buy the one that doesn't share the data back. Now there's ways like there's Federated Learning and so on that in privacy preserving stuff so you could share the data back. But there's going to be, I think, some organizations that just don't want to take a risk. And so they're just not going to want to share. But that's going to be perfectly possible. They going to be able to basically fine tune on their own data and then it won't escape. Now, there are still some problems, like suppose you do this in an e-commerce site and you train this in all of your data, and then somehow people start doing hacking of your algorithm to try to figure out somebody else's purchase history. And we really don't have the full protections on that yet. I think, like, Bing has kind of hacked some of those by limiting the number of queries in a row just because you can you can test five queries in a row, you can't test 100 queries in a row. The dimensionality is too large. So I think we're going to find some hacks and over time, again, that's going to be part of the research and part of the innovation is to make it more safe. But generally I'm expecting that this is something that companies will do using their internal data without having an escape. Now, in some cases where the data is less confidential, you might find the prices are lower if you share your data back or at least a few share, you know, parameters estimated or gradients estimated from your data back with a common model. And companies will try to do that because their software will work better. If they could share data across multiple firms that have similar data. So that's something else you might expect, but it might, you know, depends on the nature of the data, the attitude towards privacy and so on which which way that will go.

ROSSI: Okay. I think we've touched upon a little bit of this, but I want to get deeper into it because we are receiving so many questions related to this. So one one of the key kind of theme in the questions that are coming into the chat is how will ChatGPT change education? Like, the question is everybody's thinking that there is going to be some upskilling that is needed, and is kind of, are universities going to be the ones that should provide this upskilling? Are the private corporations, are they going to do it by themselves? And also, in terms of the actual classroom dynamics, how can we teach where a lot of the, kind of, our tests and trying to understand how individuals have learned from our classes is by kind of asking them questions and maybe writing essays, But now they are kind of people are getting access to these tools that are becoming more and more powerful. So how do you think the education system is going to change and what are universities doing in this context? And how do you see this kind of evolving in the future?

AUTOR: I'll be happy to chime in. So, I mean, I think it presents both challenge and opportunity. The challenge is pretty clear, which is, you know, how do you determine if people are learning, if machines can do a lot of their work for them? I mean, you know, it used to be very controversial, by the way, to use calculators in the classroom and so we've gotten used to that, we've gotten used to the idea that, you know, testing, making people memorize timetables is not necessarily the best use of their time. I do think this makes it a little harder because the the range of tasks that you could sort of consult the machine for are so broad. And I think people are adapting to this in real time. You know, you may say, well, maybe it's like timetables. You don't need to know. You just have to have an idea and, you know, feed it into ChatGPT, or, you know, it's successor, and see where that goes. So I think, you know, this is a major challenge and I think an issue, even to the degree that we think we know how to deal with this, is educating educators on how to deal with that. They don't have a lot of time to, you know, to spend on this. So I think it's a responsibility to kind of figure out practices and then work to enable people to learn them and used them well. And let me say, the other side of this, which is not on sort of the testing, but on the production of education. And there I think there's great potential. You know, education is an incredibly expensive, slow, customized activity. It consumes a larger and larger share of societal resources, not just in terms of teachers and buildings and equipment, but in terms of the time, the fraction of people's lives that they spend in classrooms. People used to go to work at age nine and then at age 18, and now if they do a Ph.D. at age 40 or so, and of course, a lot of adults, we have an incredibly abysmal record, actually, of skills retraining for older adults, not for for young people, but, you know, people who are further in mid-career and further. And the question is, can we use the technology, not ChatGPT exclusively, but many of these technologies, to make education more engaging, more immersive, more accessible, cheaper, faster? You could, you know, one thing we know about adults is they learn really well in situ rather than, you know, from watching people on blackboards. Can't we create simulated environments where people certainly, you know, learn in some sense in the field from doing what they're doing, where the stakes are lower, but the feedbacks are very similar. So I don't think this eliminates the need for teachers, and people seem to be uniquely moody, motivating for one another. That's part of the reason teachers are effective. They just seem to have more power to motivate and videos, more power motivate than books, and so on. But it certainly can do a lot more customization and a lot more support and then a lot more simulation. And simulation is something we can't do well in the classroom without technology. So I think this is a fabulous application when I say, you know, we should guide the technology towards useful things. That's a very useful thing that we could we could be investing in.

TAMBE: Just a couple more things. So I've also, you know, even this morning, I was hearing students say how it's like having the best TA ever, right, in terms of education, the way they're learning encoding tasks, how to code using this tool is amazing. And just a couple of meta comments. One is that the diffusion path for this, we're all AI targets as employees or customers, we have been for a while, but I think with the reach of these companies like Microsoft, Google, Slack and Notion, we know that they're going to be able to roll these tools into their productivity software so quickly. It's going to be really interesting, I think, to watch people become AI users really rapidly. The scaling process, we're talking about the time when I might be a little unpredictable relative to what we've seen before. And the last thing is more of a question, you know, I love to [inaudible] I'm interested in, Anton mentioned before this notion of being rubber stamps for discriminating. Now there's this, it sounds like everyone here has used this for either Python or some other tasks like that, and, you know, I think it's fundamentally true that at least at this point in time, you have to know Python to use it for Python. And so what this means for that complement substitute calculus is something that I find really interesting and a question mark I'd love to know more about.

ATHEY: So, let me maybe pick up on some of those points. I thought Sunny made multiple excellent points. Actually, one of them maybe is a bit of a side point, but the fact about this adoption, I think one thing that makes me whenever I give like the scary case for work for jobs is in the current environment, is that over the last 10 or 15 years, firms have been adopting more modern software architecture. They've been using software as a service, they've been moving as a cloud. They've been doing things much more modularly, and they've also gotten pieces of automation into things, like customer service where there's like , you know, AI assisted answers, for example, that makes people more productive. In that environment where you're already, like, partially automated and you've laid a lot of groundwork and you have figured out your data and you've figured out how to pass it around and figure out how to give everybody access to software as a service. If you get something in that makes, it like, eliminates the need for a certain aspect of customer service, the adoption at this point can be very, very fast because people are already using providers. So I think that, you know, that goes across a bunch of services that you mentioned, Slack and email and so on, so that a bunch of people have adopted a cloud software product where you can update that product instantly for everybody and then everybody adopts overnight. So it's like we've like been building factories to be ready for electrification and like now the lights are turning on, you know? And so even though it's been really slow to see some of the impacts, they may happen faster. Then circling back to another couple of comments that Sunny made, I totally agree with the best TA ever. You know, I was trying to make graphs and I like them to be really pretty, but I just don't have the patience to do all the syntax to make them pretty and, you know, I just copied the top line of an Excel spreadsheet to export of my data. So I got the top line of Excel, I get the list of variable names, you know, you copy it in and you say, like, write me an R script that's going to load in the dataset, you know, aggregate it this way, do it that way, make me a graph. You go back and forth and suddenly I'm getting this beautiful graph and you could really do that in a very natural language way. And I think there's just a few steps to get that to be really usable. So I needed ARStudio installed or Python installed, like, you needed to know something to be able to get started. But you can make a user interface that simplifies that, you can be running ARStudio in the cloud. You know, you can set up a server that extracts that so nobody has to install any software and then you can make it easier for them, you can ask it what is your variable names you know and then you can ask it what kind of chart to make, and then you can just iterate. I mean, it's just kind of amazing. Of course, it has to have the data to actually make your chart. So there's a little bit of work to be done there. But, you know, it's so close, right? Like, it's so close to just not needing any syntax at all that I just can't imagine it's not going to be there in a year or two. It's like, I feel, you know, if I dropped everything, I think I could write it, you know? So I think I think this is just going to be close. So this is going to be really interesting going forward. And then a final thing, this is very unrelated, but I actually did a research project very recently where we were trying to get women into tech jobs. And so we come up with a program that only cost $15 a person that increase the probability of women getting a job in tech from like point 2 to point 3, and, you know, is very successful. And what we did there is we got people who had the, kind of, aptitude and then we figured out, we interviewed employers and figured out what they could, the workers do to demonstrate their skills. And then we put them into groups and they work together in groups and used a technology platform to work together to create a portfolio, which then they could show to employers. And I think that kind of model really has a lot of potential here because you've got people with general aptitude. Part of the thing is that, you know, people you can, sure, you can do stuff on Coursera. Sure you can Google it and find it. But people don't really have confidence that if they make those investments, it's going to translate into a job. So you need some curation, you need some teamwork, you need to make sure that the things people are learning are things the employers actually want. But if you just put a few of those pieces together and, you know, we're doing this for Polish women and Ukrainian refugees and you know, it's working and it wasn't that hard. Like basically we built this with like a couple of people and very little money. So I feel like we can solve these problems with a little bit of thought in a scalable way, but you do have to have multiple ingredients to make it actually work.

ROSSI: I see. So one question that just came in, which I thought was relatively interesting is what, like, the question is asking what are the foreseen and unforeseen regulation on assisted AI tools? Like do you think that the regulator is going to step in and kind of determine what is, kind of how this tool should be designed or this is something that instead is that the government is going to be staying out of?

AUTOR: I mean, I don't think the government will be so involved in design. But the question is when they're used in high stakes settings, especially for decision making, then there's a role for policy. This comes to, you know, credit issuance, job screening, all kinds of things like that. And this is where this issue, as Susan was speaking of earlier, of kind of the, you know, the biases or the unknown properties of these models. Now, I mean, we have lots of decision makers in the world and they have lots of unknown properties, it's not that people do this well and without bias, there's lots and lots of room for improvement on human decision making. That difficulty, and I used to be actually very optimistic that, look, using machines that just you can sort of, you can tune out the bias, right, it's much more educable than the average decision maker. But is the case that the machines, they're so opaque to us now that it's like, that makes it harder to do. You know, I used to, you know, refer to Polanyi's paradox, which is that we know many things that we don't know how to explain what we do, right? We do them, but we don't understand them. And now we have the opposite problem in which we have machines that understand many things and can't tell us what they're doing. This sort of Polanyi's revenge, and so it makes us, it actually makes it hard, neither we cannot communicate our sort of tacit knowledge of machines, and now they can't communicate their tacit knowledge to us. So it does actually create this real challenge for, you know, kind of regulation or even testing. What is testing mean, like, you know, the notion that, well, I could drive a Tesla and on average it's safe, but every once in a while it's going to drive off the San Francisco Bay Bridge, you know, into the bay. Like, well, we just have to learn to live with that because on average, it's fine. That's not so exciting. So I do think the the regulatory challenge is extremely difficult here. Regulators don't understand this. And no one, you know, really has a complete handle on it. And so figuring out what are standards, right? What is a standard that something has to pass, I suspect a lot of things will look like that. What do you have to prove to certify that this thing is acceptable for making these types of decisions? And another really fruitful area of research, something [inaudible] has worked on with [inaudible], is how do machines and people interact in decision making? Because we're going to have a lot more machine supported decision making. But, you know, whose authority should, you know, prevail? When should you accept the machine? When should you accept your own intuition? How do you virtuously combine them? Will we just be, you know, rubber stampers for machine, will we be like those pilots that have forgotten how to fly? And so when the autopilot fails, we crash into the sea. Hopefully not. So this is going to be an important area for training and research.

ATHEY: Yeah. I should mention I have a little research paper that models formally the trade offs that David was just talking about, about kind of falling asleep at the wheel and the incentive problem that's created by having a high-quality AI. So, you know, trying to figure out, it's really like an almost, an organizational design problem where, you know, people were motivated to pay attention because they were making decisions. And if you take that away, then it changes the returns to your investments. And you need to reshape your incentives and structure to get good outcomes. And it's very hard to anticipate. Just to circle back on something, you know, something I saw in the search engine as well, like, you know, a typical engineer, even a manager of an algorithm, doesn't necessarily understand how or why the algorithm works. And it certainly is not trained to understand the unintended consequences. You know, how do you know your algorithm works? Well, you maybe test it, you tweak it a little bit, and you look at, you know, ten numbers in the AB test and they look good, then you go forward that doesn't necessarily lead to understanding. And some of my research actually was about, you know, trying to take the results of AB test and help enlighten people about what the mechanisms are, because it was I was observing that people weren't really understanding what they were building. And that's certainly the case for the people building these new large language models. I mean, certainly the people working with them have a little more intuition than the rest of us. But roughly everybody's learning about, I mean, you can write down the math. We can all stare at the math, it's not enlightening. So then, you know, you're looking at the outputs and that's all we know are the outputs. And it's not like any particular individual has had that much time to play with the outputs so far. So, you know, I think sometimes like, oh, we need to talk to the person who built this and they'll explain it to us and, no, like, they can they know, they've run a few more tests and they've gotten some intuition from this test they've run, but the space of tests is very large relative to the tests that have been run today. And mostly people have just been focusing on getting the things to work better, making them bigger, engineering them to work bigger, and just seeing like generalized performance outputs. So that's kind of where we are right now, and it's not like somebody has this magic insight that if we could just find the sage, they'll be able to tell it to us.

ROSSI: Yeah, this is, I think this is, we are out of time. I would like to take a moment to thank Susan, David, Sunny and also Anton for the fantastic event. I think that this was incredibly insightful and, you know, ChatGPT-4 was just released yesterday. You know being and all the other tech companies are jumping in, so I think that in another three months what we said today is going to be old and probably not relevant anymore, we are going to have to have another event and we're going to keep on going and keep on learning. But thank you so much for the insights. It was absolutely phenomenal and everyone enjoyed the rest of your day.

AUTOR: Thank you so much. It's a pleasure to be on the panel with all of you.

TAMBE: Thank you.