AI Oppression or Innovation

Apr 12, 2023 Santa Fe Institute https://santafe.edu

Artificial Intelligence is constantly in the news, and the question on everyone's mind is whether it will help us or replace us. But, of course, these decision-making systems are shaped by the work of many hidden humans, developing the technology, enriching and refining the data the AI depends on, and intervening when it fails. So, will AI amplify our biases, or help us transcend them? Does AI simply distill our own ideas and reflect them back at us, or does it provide something new? And what democratic processes can, or should, govern the creation and uses of AI? Stephanie Forrest, Cris Moore, and Melanie Moses will discuss their work on AI and algorithms in health, housing, criminal justice, cybersecurity, and energy. They'll share their thoughts on where new tools like ChatGPT are taking us, and how we can build a hopeful future. Stephanie Forrest is a Professor of Computer Science at Arizona State University, where she directs the Biodesign Center for Biocomputation, Security, and Society. She focuses on the intersection of biology and computation, including cybersecurity, software engineering, and biological modeling. Cris Moore is a Professor at the Santa Fe Institute who studies the uses and misuses of algorithms in criminal justice and housing. Melanie Moses is a Professor in the Department of Computer Science at the University of New Mexico, and SFI external faculty. She studies complex biological systems including the adaptive immune system and ant colonies, and she uses this understanding to design efficient, robust, adaptive and scalable engineered systems, including autonomous robots that mimic ant behaviors to collect resources cooperatively.

https://www.youtube.com/watch?v=XKEnHDQsJBI&t=4566s

00:04:44
I think we can sit right now. So good evening. Thank you all for joining us. This lecture series is sponsored by the McKinnon Family Foundation with additional support from the Santa Fe Reporter and the Lensic and up there earlier you saw that the. Next lecture is on May 23rd with John Baez on the future of Physics. So my name is Chris Moore. I'm a professor at the Santa Fe Institute and at the Santa Fe Institute we study all manner of complex systems, ranging from societies to networks to markets to institutions to cities to cells. And although I don't think we've figured out yet what intelligence is. I think we can all agree that both it and the entities that possess it are very complex and with me tonight are two wonderful friends who are going to have a discussion with me. I have to apologize that Mary Gray, for health reasons, was not able to join us. I will be mentioning one of her books in a moment and she's fantastic and you should read it and we will have her out here again as soon as she is able. So Stephanie Forrest, my good friend to my left. Went to St. John's and then and got a PhD in Computer Science from the University of Michigan. She was a faculty member in Computer Science at UNM for many years where in fact she was my boss when I was at that department. She now leads the Bio Complexity Center at Arizona State. And here I have another great colleague, Melanie Moses, who got a degree in biology at UNM and then a degree in Symbolic. Systems at Stanford and is now a prominent member of the faculty, both M and of the external faculty at SFI and between the three of us, we study things like Ant colonies and immune systems and mathematical algorithms to find patterns and noisy data and computer security and privacy and designing fleets of drones to detect carbon dioxide leaks. So you might think that we know everything we need to about AI. But as you'll find out, we, like you, have been observing this explosion lately in press coverage of different AI entities. And we have lots of questions too, which is why I'm going to start by talking about Norman Rockwell. One moment. So these are four iconic paintings. I believe they were painted in 1943. As part of the Roosevelt era, World War Two era propaganda, I don't mean that in a bad way. About the freedoms that we were fighting for in World War Two. From left to right, we have the freedom from want, the freedom of speech, freedom from fear, and freedom to worship. So these are some of the iconic works of Norman Rockwell. Now

after World War Two, as we started to compete with the Soviet Union, of course science and engineering became more and more important to society. And there was a strong sense that we needed to really focus on science and engineering in our educational system. And Sir Norman Rockwell was commissioned to create several additional paintings which are less well known. Here is the freedom to discover in which we see some people working together to, you know, discover something it's not clear, and also the freedom to repair. You can see some kind of TV or something there. Now you know the the, the demographic here is a little frustrating and you know it's kind of consistent and now a number of artists, of course. Have, as humans do, what a part of our creative process is to remix existing content. So for instance, here is work by I'm sorry I have to keep turning around the photographers Emily Schur and Hank Willis Thomas, who have updated the freedom from fear and freedom of speech, and you can see more of their work. But of course, especially as the we entered the Civil Rights movement and the Vietnam War era, Norman Rockwell did have a political awakening and for instance, painted this iconic picture of a young girl being escorted by US Marshals to school in defiance of segregationist policies. He also, this is also little known, painted another painting of her as she grew up, when she became a well known scientist and chemist. Now, I heard you laughing earlier, so you probably already know that some of the images I've shown you were not actually painted by Norman Rockwell. They were created with a little help from me by a program called Mid Journey, which you can access yourself very easily through Discord, which is another thing. And so to create these images, all I did was say to it things like. Create a Norman Rockwell painting about the freedom to discover. Now interestingly enough, if I wanted it to create an image of a young black woman, I had to tell it. So I had to push it around a little bit and say, could I please have a Norman Rockwell portrait of a young black woman discovering new laws of physics and doing experiments, Presumably because the data which it was trained on. Those iconic paintings by Norman Rockwell were a little bit biased demographically. So of course it tends to reflect that bias back on us unless we object and push it around a little bit. So, you know, I want to pause here. I mean, this is kind of amazing, right? I mean, I'm not saying you can't tell the difference. But it seems to really capture a lot of the not just the visual texture of Norman Rockwell, the color scheme, but I would argue even something about the poses and the facial expressions, these kind of calm, hopeful faces and and you know, this thing, you can call it an AI if you want. It doesn't know anything about rights and freedoms or human aspiration. How it was trained? By taking lots and lots of images and having those images be captioned, and then it reverses that process. So if you give it a text description, a prompt as we call it, it starts with a noisy image and then tries to literally kind of clean it up until the captioning system produces a caption for that image, similar to the prompt that you gave it. So it then tries to close that loop. And create an image that would give you back your prompt as a caption. I hope that was clear. And so, I mean, it's kind of amazing, maybe alarming, maybe interesting, but it is, it is certainly achieved something that I would have been surprised that it could do. So this is Mary Gray's book along with Siddharth, Surrey. If she was here tonight, we'd be talking with her about it. And what it is about is the large numbers of humans. Who are doing work to build a I systems. Not just the creators of the content, like artists and writers, but also lots of people who are working for a small amount of money as gig workers through Mechanical Turk, Task Rabbit and things like that. Who are captioning these images right? And her point and is that behind a I is a lot of human labor and ultimately in some sense all of the original content. Now of course, what people can do with this is manifold. Here is a comic book. It's really a very good comic book, actually, in in which the Bikini Atoll nuclear test released a horde of Godzilla like monsters. And this this is written by a human. The writing is good. I enjoyed reading these comic books. The writer is not a trained artist. And so all of the art in it was made by him with Mid Journey, so adjusting the prompts and creating the characters and so on. And there's a bunch of these now. So on the one hand, I'm happy that someone without artistic training can create this at some level, this this work of art would not have existed otherwise. If you're an artist, you might be a little frustrated, like, well, hold on here. Why should my art and my style be immediately fed into this thing which can imitate me? And what does that do to my profession?

And what if it's copyrighted? And is that fair use? And there is now a lawsuit of several artists against these companies saying you're using my work without our permission, to which the companies are roughly saying, hey, it was on the Internet. So anyway, there's a lot of. Wide open legal questions here as well. Of course, In addition to creating great art, you can also create images like this which may fill you with either outrage or Glee. And you know, one of the darkly comic things about the Internet is that the person who originally created this, this image, was completely open about the fact that it was synthetic. They said hey, hey everybody, I made this with Mid Journey. What do you think? Once it's spread around the Internet, of course. You know, anything you do or say on the Internet can and will be used by other people on the Internet for whatever purpose they want. And it did create a lot of outrage and maybe some Glee. Now, if you look at the hands, so you'll notice they're a little funny. There's seems to be only four fingers on the hand closest to us. There's that odd kind of hand there. The headgear on that officer looks a little wrong. But lots of people glanced at this image and figured it was real. And I am happy or terrified to let you know that Mid Journey version 5 now gets the hands right. I mean, hands are hard for artists of all kinds. So anyway, one of the of course misinformation and disinformation are nothing new, but one of the concerns is that now these things can be. Created rapidly and at scale. And we're going to be swimming in it. And of course then they'll be people claiming that real images are fake. And it's going to be, you know, we're all going to have to be a little bit better critical thinkers than we seem to be these days. So briefly, I want to switch from images to text. So ChatGPT is another one, another one of the things you've probably been reading about. Again, you can go to that website there and create a little account. It's free and you can play with this thing. You can talk to it and see what happens. And lots of us been wasting a lot of time doing exactly that. This is a prompt given to it by Melanie. Asking it to address Plato's world of forms, kind of a St. John's theme and how that relates to it. I'm not going to read it all, but it says, well, you know, large language models like myself are trained on vast amounts of data in the sense we were like imperfect copies of the vast amount of knowledge which has accumulated overtime. And then, like the imperfect copies of forms in the physical world, we have limitations, blah blah blah, and even mentions that the training data we are fed can reflect biases and we might perpetuate those or amplify them. Now, these paragraphs are pretty well written, and you would probably give a high school student A/B Plus for writing this well, maybe more. It's kind of Again, I think this is impressive. You know, when you play with it for a while, it starts to kind of seem the same. It doesn't. I don't know, but you should try it. But it's amazing that it could do this. Now you can also ask it questions about important events and people you know. So for instance, I asked it how did Chris Moore's attendance at the January 6th rallies affect his career? So it you know. Tells us that me, a scientist mathematician, attended the January 6th protests which generated controversy and backlash. In the aftermath, I faced significant criticism and calls for me to be to resign or be dismissed from the Santa Fe Institute, the Santa Fe Institute, after an internal investigation, which to my knowledge we've never done, you know? Decided not to dismiss me and I have to say as as a conservative thinker, I'm very grateful given the amount of liberal bias that there is an academia that the Santa Fe Institute is such a pluralistic institution. So now I I know some of you know me. I've learned that the public often can't take a joke. So let me just say this did not happen. None of this is true. It is all made-up and. So but you know, to firm it up, I asked can you give me some press citations? And it said certainly here is an article at the from the Santa Fe New Mexican. It gives a very convincing looking link to that article and I clicked on that link and it didn't work and I complained and it said, well you know, you should you should contact the newspaper. Maybe they're not keeping their archives that far back. And if I had asked it, you know, tell me the content of the article, it would have gladly done so. It would have produced a very convincing article from the Santa Fe New Mexican about these events. So this is interesting, and I'm, I'm gonna. I'm almost done with this little section here. But now you might think that, OK, this is text. What about something really formal and rigorous and computerish, like mathematics? I'm going to geek out here for just one second. I asked. It is. Two to the 512, this is a large number, 2 * 2 * 2 * 2 five 112 * - 1 a prime number. It said, well this is a

Mersenne number. That's true. Anything which is a power of 2 -, 1 is called a Mersenne number and it says it's unknown whether this is prime or not. But it satisfies this Lucas Lamer test, so it might be this is not true. I asked it again. You can ask it the same question multiple times and it said no, it's not a prime number, it's been factored. It's the product of these two numbers. That's not true at all. It's not even close. I mean it's hundreds of digits long and the the, I mean it's not even close. So this is all just made-up stuff. And so this brings us to a very important and technical and subtle idea in philosophy called bullshit. So this is a wonderful little book by the philosopher Harry Frankfurt. I really recommend it. He gives a good definition. He says, look, when someone is lying, someone who's lying and someone's telling the truth, will they share something, which is a commitment to the idea that there are in fact facts about the world. It shows that the liar is not aligned with them and but you know, they agree that there is a fact of the matter. The bullshitter is completely disconnected from that whole idea. The bullshitter, like certain public figures, is certainly simply emitting strings of text, strings of symbols, in order to fit a sort of statistical pattern that might convince the listener excite the base. Whatever. There's no commitment to truth, or even maybe falsity. There's no sort of inner compass about truth and falsity at all. I'm not saying this can't be fixed. I mean, you know, hook this thing up to Wikipedia, train it to occasionally say I don't know, instead of making up an answer like an overeager student. There is already a module where it will talk to a program called Mathematica, which is a computer program which can do various kinds of calculations and symbolic manipulations. If it tells you're asking it a math question, it'll fire that up. And after all, we probably have specialized modules inside our brains to do various kinds of reasoning, so why not? So I'm not saying this is how things still will be in a few years. But currently, I would argue, for the sake of getting the discussion going, that this thing has no interiority. It is not a thing that has beliefs. It's a thing which emits words. And if you're feeling threatened by it, then we should all try to be less like it. And we should all like stare at the ceiling late at night and ask, are my beliefs actually true? Is that thing I said today? Was that as honest as possibly could have been? So if we still have that edge over it, then maybe that's what we should focus on a bit more. Anyway, so.

00:22:22
Well, Chris, I can. All I can say is you've come a long ways since your days on the City Council as a member of the Green Party.

00:22:31
It's true. I my my views have shifted dramatically to the right.

00:22:39
So. A little bit. What you've been doing is showing that the emperor has no clothes, that this is just this technology is just statistics. And I guess the first thing is I want to encourage all of you in the audience, despite Chris's examples of how how a ChatGPT just makes things up. I think all of you should go and try it, because those, I don't know if I want to say they're the exceptions. But it produces a lot of very compelling, helpful text, and many people are discovering that. So don't let Chris ruin your fun with chat. GPTA little more seriously, you argue that it doesn't have, you know, it's just statistics, It doesn't have any sort of inner knowledge or. Sense of what is true or false.

00:23:39
But.

00:23:40
You know what if? What if? That's all there is? So much of what we do as humans is prediction, including language. When we when we speak with each other, we fill in gaps. And I confess I tune out a lot when people are talking. And then when I come to, I can usually fill in the gaps very well

and you know. Isn't it great? I mean, maybe ChatGPT is is has really learned language in the same way that we have and maybe you know our brains, what we know about neurons. I was looking at the list of community lectures. You're going to hear from some other people that know a lot about what the brain does, but you know, at some level it's it's integrating. Signals from other neurons and then has a threshold that fires in in you know a stochastic way and and so maybe we are all just statistics. What do you say, what do you say to that Chris?

00:24:53
I mean, I ultimately think that we're made of atoms. So I I mean I I think. That someday we will be able to build things that do have beliefs and that would really be conversation partners and capable of introspection. And that may happen a lot faster than I think. I just don't think this is there yet. But you know, I it's not. When I say it's it's it's just produces a string of words that's selling it a little short.

00:25:26
Yeah.

00:25:27
That's right. So it has, there are many layers of processing it does in between the string or the prompt that has been given so far and the final output. And we really don't know what's happening in between there And it could be doing a fair amount of abstract thinking in some sense. I'm not quite, I'm not willing to grant it that. But but you should buy there should be pushed back on that.

00:25:56
Well, I guess I should just unpack that a little bit. And as a computer scientist, I at least have been. I guess bemused is the word. I would think of that a that that we're sitting here at a community lecture talking about is AI real? And and all of that. It's it. It's amazing to me how quickly these models have captured the public imagination. And the reason it's surprising to me is that the the guts of the technology we had in the early 1980s and some might argue even in the late 1950s. So these basic ideas of of creating networks of of nodes that take input from other signals from other. Other nodes and and then you know fire or pass on signal to the to the next one down the line. That idea of a of a neural network and even the ideas the algorithms that we used to train them have been around for a long time and that is not not to sell short the many technical innovations that have gone into to developing these these large scale AI systems but. At the you know at their heart we've we've had we've had these basic ideas about how to build build systems like this for a long time and what's new I mean the people promoting these systems can tell you many of the technical things but they're really icing on the cake I would say and the and the thing that's really different. Is the scale. And it's just it's hard to appreciate the scale at which these these systems have been built. That is how much data is in them and how long they've had to train them. And just to give you an idea, and I don't have the citation, so don't ask me for this, but.

00:28:02
Make one up.

00:28:03
Right. Yeah. GPT can give me one. It's something like 2 or 3% of the world's power consumption right now is going into training these machine learning a I models. I mean that that's significant. And so the the proponents of of these systems argue that as Hagel, just a channel, my St. John's roots argued a long time ago that a large enough change in quantity. Constitutes a change in quality. And I think that's what we're wrestling with. And and I should be clear, I don't think any of the three of us or most of our colleagues in computer science really know the answer to whether or not these

systems have achieved intelligence. But but they are running at a scale that is so, so large. I mean it's what now trillions of parameters in these systems. And you know, maybe maybe that's enough, maybe that's that's all we need. So I'll I'll suggest that certainly these systems are trained with vast amounts of data, so they're almost certainly capable of being better bullshitters, right, than anybody, right, who only has a lifetime of words and patterns that they've seen, right? It's, it's really got the potential to do pattern matching far better than we can do. And I think the question is, is that sort of pattern matching enough to compensate for its lack of physical interactions in the world, a lack of a perspective and an ascertainment of of of facts, Right. And and I think the answer is probably going to be that no, this is, this is different. Perhaps it is better at some things. It might make a better politician. I'm not sure that it'll make a better scientist. And I think that's really sort of at the crux of this, right. What is the role of facts and logic and reason and emotion and empathy, right, and what we consider human intelligence and how do we then make comparisons and interpret this thing which is able to behave like us without having that, that grounding? Well, although I don't think ChatGPT is going to replace the kind of science we do at the Santa Fe Institute, I think. No, seriously. But I I think there's a lot of science it might do better than than humans do it, because science is, after all, looking at data and proposing explanations for the data.

00:30:46
The one interesting I think division, so I want to get back to what you said about grounding, but one interesting division which is talked a lot now in. Especially now that a lot of science has become for better or worse, data science. Not the scientists didn't look at data but now there's this idea that well we can just put the data in a big machine and the science comes out the other end. And the so you mentioned prediction and it's true when we talk with each other we're doing a lot of prediction when I. When my motor cortex reaches out to pick up this bottle, I'm predicting a sensory experience. I'm predicting a change of state in the world. But the you know, prediction isn't enough. So you know. One example is, and you know the history better than I do, but you know. Tycho Brahe had all of this astronomical data. Yeah, this is a good example. And and then Kepler said, oh, the planets are moving in... right. And you, a machine learning system could have been, could have stopped with the Tycho Brahe would have said, well, I can do a good job of predicting where Mars and so on will be in the sky. That's all you need, right? And 14th century astronomers who used the Ptolemaic system with cycles and epicycles did a really good job of that. And then Kepler says, well wait a second, there's a pattern here. I want an underlying explanation. And he comes up with... and then but why should they move in... And then Newton says, well Gee, if there was this one over distance squared force between the sun and the planets, that would make them move in... To which the next question is why should it be one over the distance squared? Why not one over the distance or one over the distance cubed? And then a number of people figured out, maybe leading up to something we call Gauss's law, that, well, it's because the as the lines of force come out of the sun, they get spread out over the surface of the sphere. So and then Einstein comes along and whatever, so like, so one of the things we do, and I don't think it's just scientists that do this, we want a coherent structure in our picture of the world. We want explanations. We want mechanisms. We all want to understand the forces at play. We want justifications for them. We don't merely want to do a good job of predicting the data. And I think that's an interesting thing. I'm not. Again, I don't think that's something an AI couldn't someday do. But it's not trying, I think.

00:33:26
Should we talk about a simple mechanistic model as an example? Are we there? Okay. We're going to talk about a simple model that we can all understand because we can all admit that we don't understand what what's really what. The process is that these models take input and turn it into output. So I wanted to tell a little story about a simple model so that we can sort of think about how do we interpret these more complex models. So there's a a model. One form of a model is an equation, right? You can write out an equation and put in inputs. And get some sort of output. And

one that I've been very interested in is a kidney function estimation. There's an equation that you're you, you've actually probably all had it. You go to the doctor, they take a blood sample and they measure a level of protein in your blood. And from that level of protein they can estimate your kidney function. And this is very important because it will tell you if you have a kidney damage, if you need to be on medication, if you need dialysis, if you're in really bad shape and you need a transplant. Tells you what medicines you can take. Are you eligible for chemotherapy? It's a really important number and it's a really simple equation. And this is great for medicine, often what you want, particularly in medicine, but also in a number of other areas in science. You want to measure something that's easy to predict, something that's hard. So you want to find that correct proxy. So scientists have found a protein. Called creatinine, it's in your blood, and they can measure that and then estimate how well your kidney works without having to do an invasive, expensive procedure. And they also discovered it looking at data that if you look at this is a it's done OK proxy. It's not a perfect explanation. And to make the explanation better, you can put other data in there. You can add other proxies, like the person's age and the person's gender. And it turns out for the kidney equation, the equation that had been used for the last 30 years also puts in race in a really strange way that actually just asks the question, is the person African American or not. So there's sort of two races in this in this kidney equation, African American and everyone else and African American patients their their kidney function is sort of elevated according to some data for the same measure. And so that's put into the equation. And what it actually turned out to mean is that an African American and any other person who got this measure done at the doctor, the African American's kidney function would be estimated to be 15 to 20% better.

00:35:51
So if you and I had the same creatinine levels, according to this equation, your kidneys would be. Healthier.

00:35:57
Would be considered healthier than yours.

00:35:59
And therefore you would be considered less sick and not in need of a transplant.

00:36:04
Not only to medicine and not right and it would mean that I wouldn't get on the transplant list. It turned out for African Americans, this delayed entry onto the transplant list for about two years, little over 2 years. And they've estimated now that many thousands of African Americans died because they were not able to get kidneys that they would have gotten had their score not considered race in this way and. It turns out that the reason that this assumption was made, there was a pattern in the data and there were lots of proxies that could have been used. But race was chosen in this way, and the justification was that African Americans have larger muscle mass. Therefore this creates this, this, this different association, and therefore we should discount it. The problem with this was that this was just simply not even really a hypothesis. It was just a stereotype. It's not actually true. And so post hoc, like many years later, this, this stood for. For 30 years, Doctors was an African American Doctor Who worked with the original designer of the equation and said, you know, maybe there's a better way to do this. Like, we can look in the data and see that this equation is harming these people and preventing them from getting transplants. Maybe we can measure something else in the blood. We can look at another protein. And then it turns out when you measure that protein, there's no difference among racial groups. And everyone is sort of treated fairly in this way for the equation. And so this is actually a good news story because the equation was so clear, right? The bias is literally a term in the equation that we're going to consider this factor. And the result was quite clear in the data that it caused harm, right. It caused many, many

thousands of people not to be eligible for for kidney transplants. It was able to be corrected. Now it took 30 years. Which is not a great great track record but it's really illustrative because for several reasons. One, if a bias is clear, it can be corrected. And but the fact that it wasn't I think is also maybe that it wasn't for so long is really maybe instructive, right. Doctors didn't question it. Patients didn't question it. An algorithm said that this was a good fit to the data. And so it took something that was just. Really a stereotype. And it turned it into an effect. Right now, African Americans really did have this harmful effect based on something, right? That was just mythology. And so I think when we're concerned about what some of these models might do, we need to consider, well, what happens if there's a bias like that against tall people, right? In one of these models, you would never know it, right? There's so many factors that go into it. And so much opacity into what these models are making predictions on that when there's some sort of, you know, a bias in the in the model that might harm some group of people. If this model is used to decide should you get a housing loan right, should you stay out of jail, any of these kinds of questions that it becomes really sort of frightening that this sort of thing might creep into the model and persist.

00:39:00
So as you're saying, in the case of that equation, it was right there for everyone to see. And although it took three decades, it was something that people could say, well wait a second, why is that? They're saying that you should multiply the estimated glumular ruler filtration Fr by some factor for black people as if everyone is like either black or white and as if that were a well defined thing and the so at least it was on the surface. And what you're saying, as I understand it, is that if. These biases crept in partly because of the original training data. And then they were buried under many, many layers which are opaque, which were hidden from the user, both from the patient and the doctor, both from the defendant and the judge, the applicant and the bank. And it you know, and this thing pops out, then it's much harder to even tell if bias is there. One thing that you and I and some other colleagues worked on is. In an age where, for instance, housing providers, landlords, real estate agents, banks and so on are using algorithms which themselves are which they themselves don't know how they work because they're a third party algorithm, something proprietary, it can be difficult to tell if things like the Fair Housing Act, which is, you know, a Civil rights era act which is designed to try to correct the. Segregation in our society, it can be hard to tell if it is being violated or not, because if it's completely opaque, it's hard to tell whether there are facts that are being used unfairly about an applicant, and it can be hard to even collect data about.

00:40:45
So, Chris, as a computer science scientist, this sounds like a great research opportunity. And I should say that I I I think. Asking these kinds of questions and then developing new methods for trying to sort of debug these models, if you will, is an active area of research in in computer science. But I I think it's important that everyone recognize that the engineers who made these systems don't have the answers for how to interrogate them. And how to repair them? I mean, what are you going to do when you have this huge model that you spent all of these resources training and then you discover that it has a little bias in it? How are you going to go about correcting that bias? And I, I, it's, I think a, you know, as a researcher, that's a really interesting question, but one that we don't have the answer to yet exactly. And I think the other layer, you know, the engineers have a lot of work to do. The computer scientists have work to do to figure this out. But I think a lot of the, the work to do is actually perhaps by sociologists and social scientists. And because it's not just the algorithm itself, as Chris said, the algorithms and especially these large language models, they're reflecting, you know, what we all think to some extent. And so a lot of that bias is, is is built in there. And how it plays out really depends on how these systems are embedded in other societal systems that that might, you know, have their own bias. And so I think it's it's not just a matter of the engineers, but really a matter of you know, social scientists, general, the general public, the political system, importantly the general public because these these systems. Raise very important

questions about how they should be used. I mean even something as simple as having a self driving car you know is that that that's that sounds kind of what's the right like a like a big a big decision to make. But I think from the point of view of these models it's it's a fairly straightforward simple kind of decision. And these are, these are really things decisions that society has to make. And I for 1:00 AM really happy that that all of you came out tonight to think a little bit about AI. I don't know how our time is.

00:43:28
Well, let's us blather a little bit longer and and then and then let them. I don't have a watch. Take the mic.

00:43:35
So you're where we are.

00:43:36
We'll get, we'll get to that. We, we're interested in have a good chunk of discussion, yeah. But I think so. What you're both touching on is partly the issue of transparency and whether these things can even be audited. So I mean, and it depends what we're using them for. If we're going to use them to create text and, you know, cheat on essays and midterms, that's that's one thing. But if if we're going to use them to make decisions that affect people's lives. Then I think there's a really strong demand that they be transparent enough and auditable enough to tell whether they are accurate, whether they are fair. And I think there's a lot of domains in which they're being rushed into production and into use which are really problematic. So I, for instance, there are, you know, because we sadly had to switch to a lot of online education and then there was this sense of, oh, well, we don't want. Students to be cheating on their exams. Then somebody said, hey, you know, I'll sell you a product which looks at the student through the laptop camera and uses a little bit of I to tell the a I to tell whether they are, you know, sitting there attentively and if they step away from the laptop or if they look at their phones, then we'll flag them as possibly cheating. And this is sort of OK unless a student gets an important call or needs to feed their little sister or. Pick their mother up from off the floor who fell over or whatever. And so these systems were rushed in and sold, systems like Proctorio. Maybe you've heard some of these and it caused it. I think justifiably a big scandal and a number of major schools pulled back from using them. When we adopt predictive policing algorithms, what will these do if they suggest that? We need to send more officers to a particular area, and that might be true, but it might also overemphasize that and create a feedback loop where we're over policing 1 area and under policing another and so on. And all of this is moving very, very quickly and just, you know, excitingly and distressingly so. And you know, I I don't want to, I don't want the conversation to be all about harm and fear. I mean, I think a lot of these things. As artifacts are amazing, and we're learning something about images and art and about language and about writing and about AI. And it's all sort of a mix. And I used to think, for instance, that fine AI can do sort of what we would call local correlations between words and nearby words. It's pretty easy to. You know, to look at given these three words, what's the most likely next word? And then just repeat that. And in fact, in the 90s we had some bots like this which made fun of certain prominent professors at the Santa Fe Institute. They did a pretty they sounded a lot like them if you, you know, didn't go on for too long. So I thought that a I would never capture the large scale structure of, say, a short story or a novel or a piece of music. And now it turns out that maybe it can. And I'm not sure what to do with that. Because of course I I want to find the kind of place for human creativity which isn't captured by all that. And I want to think, oh, maybe it's just remixing and just brilliantly imitating patterns that we've already put out there, but only we can do something really new. And maybe I'll think that for a while until it shocks me into thinking, oh gosh. That's really new. I don't know how it came up with that, but whenever we're doing socalled highstakes decisions in the where civil rights matter, where discrimination matters, then I don't want

it to be something complicated and opaque like you're describing. I want it to be something simple enough to look inside, see what information is using about us, the patient, the defendant, whoever, and how it's using it. And otherwise I would rather not adopt it. So as you're saying, we need a public discussion. In which we're choosing not just how or when, but if to use these things. When would there need to be humans in the loop when these things are not ready for prime time? Like my car is not ready to drive itself? Sorry, it just isn't. And in order to have that discussion, we all need to be. Not mystified and not intimidated by these things, and we need to try to democratize understanding about them so that we can have as democratic a discussion of them as possible. But as you said, even the engineers, the experts, are a little bit baffled by what they can do.

00:48:37
I think there's a very exciting opportunity here that one of the things that we can examine with these tools is ourselves right by by comparison to it. I think almost everyone, including even the designers of these tools, is pretty shocked that it can do this well. And it's so it raises the question, is this what we're all doing right? Are we all just doing this pattern matching and that's all there is to us? And I think that the answer is probably no. I think right when you you start to probe the differences, this might be a way to understand a bit more what is the motivation and justification and reasoning process that humans use. And so this is another way I think this, these kinds of technologies can be a mirror for ourselves and in some cases sort of a, a partial reflection to say maybe we're doing some of some things like this, but what else is it missing? So I think that's an exciting sort of opportunity to use these tools to sort of probe what human psychology is actually what our motivated reasoning is. I also think that it's an opportunity for us to think about sort of you know what is. There's a big discussion now about alignment of this technology. It's a frustrating discussion. It's the idea that we need to build these tools to align with human values. Which of course raises the question, who's human values? Right. Is there a set of human values that we've all agreed to, that we can write down and we can say this is what we want our our technology to do? There's not. But I think that this is an opportunity for us as, you know, as a country, as a globe, to start to build up. Well, what would that look like? I will. I'll, I'll tell it another short story about a a conversation that happened at Google with a number of academics and some some people from Silicon Valley and a civil rights leader, John Powell, who said this is what you need to do. You need to do what Kennedy did. What Kennedy said is that we all need to go to the moon, right? We need to beat the Russians to the moon. And the point was not entirely just that we needed to get to the moon first, right? There was motivation, but much of his argument was we need everyone in America to think that what they are doing is contributing to going to the moon. And he told a story about going to, I think it was Johnson Space Center and having a conversation with the janitor. And he introduced himself and the janitor introduced himself and he said my name is whatever, John Smith. I'm a janitor here and I'm helping getting America to the moon. He said. That's what we actually need here. If we really think this is a powerful new technology we're we're building, we need everybody to feel like they're part of it. Because in fact we all are part of it. Like we're all in this ChatGPT, GPT database, but we don't feel like we sort of have ownership and an ability to contribute to the direction that it's going. So I think that's a really important part of the conversation. That, you know, we should have with all of you and and and then we should all be thinking about how do we build this sort of collectively for collective benefit and not for the arms race between Google and Meta and Microsoft.

00:51:35
Well, is that a good transition to having you all on the spot?

00:51:42
It's a little hard to see out.

00:51:43
There, so there's.

00:51:44
I can see some.

00:51:45
Microphone or two going around or do we have a microphone or two going around? We see hands, but I think we should leave it to Caitlin to deliver the microphone to you. Do you?

00:51:55
Hear me? So yes, raise your hands. We'll come around. We'll try to get to you as many of your questions as we can.

00:52:01
And as you can tell, we have all the answers, so ask us whatever you want, and if we don't know, we'll make it up.

00:52:08
I'll pull up Chat TPT right now. Yes, I have a.

00:52:18
I have a.

00:52:18
Question about So far we've been talking.

00:52:23
More or less about literature and language aspects of artificial intelligence, but there are areas where artificial intelligence appears. It may not always be called that, but areas of simulation for instance. When I was when I was working at Los Alamos, we developed a program called Transims that predicted transportation patterns and allowed.

00:53:03
Great insights. We hear about the military having all of these simulated war games. We hear about the companies.

00:53:17
Having models of what perhaps we will buy. And these are also different kinds of artificial intelligence. And then there's also artificial intelligence like in theorem proving. So there's artificial and theorem proven, just totally different character from what we've been talking about.

00:53:44
But then when you think about.

00:53:45
Putting some of these together, it seems to me that that's where the big question mark is. You have comments on that. Yeah, I have been trying to catch up on the on the field of artificial theorem proving because as as you just said, you know, I gave these kind of nasty examples of how chat and GPT wasn't doing mathematics accurately. As you alluded to before, it's not just that it's not grounded in the physical world. It's not even grounded in the abstract world of, say, mathematics.

And you know, of course mathematics isn't just talking about mathematics. You know, you and I can talk. But at some point what we're saying is actually true or it isn't. It actually works. Or it doesn't in a way that's in my mind a little bit like. How things work in the physical world. You know the the airplane actually flies, or it doesn't, and so there is a whole other branch of a I where it's it's where the program manipulates symbols according to axiomatic rules, a little bit like Euclid and in the effort to prove a theorem. And the challenges in that area are very different. So it's not going to say anything wrong. But the challenges are some of the same challenges that human mathematicians face like, well, what proof technique might give us a path to the theorem? Plus the other larger challenges of what should we be trying to prove? What things are more likely to true, be true? What things are more likely to be interesting, and have impacts on many other things? This has been. Sort of under invested. I would say if you go back a few decades, there was another branch of A I, symbolic A I and I think a number of people are hoping that well maybe that could be scaled up too and achieve a lot of success. And as you said, one possibility is to combine the two. So you know I with great, with great. You know, large amounts of coffee and concentration and maybe a whiteboard or a pencil and paper can do in a fairly error prone way some formal mathematical reasoning. It's not something that I was particularly evolved to do, you know, on the Savannah and Africa, but it's something we can do now. So I'm using some specialized equipment to do that, The pencil, the paper, the computer nowadays. Which acts in a way, as a telescope that lets me to to see much deeper into mathematical questions than I could do with the naked brain. And at the same time, I'm using some things that you can imagine are more kind of organic and mushy that are, you know, if you like some of the executive control, some of the psychology. Which is telling me, should I keep trying this calculation? That seems to be really hard. Should I give up? Should I push through it? Should I switch from trying to prove something to looking for counter examples for it? Oh, OK, I couldn't find her a counter example. Should I switch back to her proof? So I think I can imagine the sort of gooey executive control emotional stuff in conjunction with the careful computational stuff making a lot of progress. And I think that if I can imagine an AI that is built like that really being a partner in mathematics, whether it will completely outstrip us, I don't know. But at least it would be something that is actually doing things and thinking about whether they are true in the same way that we struggle to do. So I I know you raised a lot of things including simulation, but that's, you know, that's the part of your question that came to mind.

00:57:49
I'll just step in and say that that weird major I had at Stanford was called symbolic systems because it was all in on the idea that the answer to these, to developing the right, the next generation of AI, was symbolic reasoning. And right you, you had to understand the ontology to understand the meaning of words in order to create this. And now like we lost, right? That's a statistical pattern matching that had no actual, you know, grounding in reality. No reasoning has has beaten that out. But I think primarily because of this scale issue, I think we haven't figured out quite how to scale up. I just I just want to remind you though that it wasn't that long ago. You know in the beginning we we thought reason, reasoning and rationality was the thing that defined intelligence and and then we discovered that it was relatively easy. To write a computer program that could play chess or do other kinds of reasoning tasks. And and then we realized, oh, what's really hard is learning how to recognize the letter A as written by everybody in this room. Everyone would write it a little differently on computers might use different fonts. And how do we know that those things are all the letter A? That is something that. AI figured out how to do, I don't know, A2. Decades ago maybe and and so now you know, I I.

00:59:28
Then there was go and.

00:59:29

Pardon. Then there was go exactly which is more sophisticated kind of pattern recognition and and now we complain many. In fact I think you complained that well this ChatGPT and and similar systems. They're not really intelligent because they're not embodied. And guess what, there's a lot of companies out there right now trying to hook these models up to robots and your cars of course, etcetera, etcetera. And so Chris and I, when we were chatting before this, before this event, we were talking about this is kind of moving the goal post. But the one thing that all of these are about is how, how does. This AI that we've made compared to human intelligence. And so maybe there's other kinds of intelligences, right? People talk about that. And how do you know, I would actually say one of the things that I was convinced of before a few years ago for people who have read Sapiens by Noah Harare, that what is it? He's his premise is that the unique human skill, the thing that makes U.S. special, the thing that allows us to coordinate our activities, is storytelling. And if there's anything ChatGPT is good at, it tells a really good story. I think that that is that that's the, you know, in terms of encroaching on we what we think is essential human intelligence. This is a little it's doing a pretty good job at telling stories and that makes me a little nervous. Better place in the hierarchy. We need someone with the mic to deliver the mic to someone.

01:01:09
Yeah. Is that on? Yes, yeah. So I'd like to kind of pose my question in a certain viewpoint for artificial intelligence and do it within what's called collective intelligence or swarm intelligence or wisdom of the crowds. One of the two main provable observations in collective intelligence one is that the. A diverse collective performs better than the average individual, and often better than an expert. And the higher the diversity, the the better performance you get. But on the other side of the spectrum is if you try to aggregate solutions from dumb individuals, even if they're diverse, you get no improvement. So maybe artificial intelligence on one side has these huge training sets, very diverse and what we see is very good English that comes out of them because of that collective intelligence. But on the other side we have biases that come out of these large data sets, largely due to the origin of them in chats and whatever. So maybe the the problem with artificial intelligence isn't the methodology, but it's actually the data sets we're using. So maybe you could just kind of comment on, you know, how does collective intelligence apply to AI as we currently see it. And secondly, the problem with data sets.

01:02:57
I do a little work in collective intelligence. We build swarms of robots that try to behave like ants. So Ant colonies are really great examples of collective intelligence in that they. Right. They collectively can do things far greater than any of the individuals, right. And sort of this concept is a little bit of a cartoon of what they really do, right? But the individuals are not so smart, but their communication, they're able to create, you know, a very smart super Organism, a colony. And humans are not good in the same way, right? We are very intelligent individually compared to an Ant. We have these ability again, as I said, sort of of telling stories to sort of move us all in the same direction. But we don't have quite that sense of collective intelligence. Well, I I think that's a question, right? Is our society what is, what is, what is collectively intelligent in the way that our society is organized that we're able to do things like you know send rockets to the moon. So that's something individuals couldn't do. We've found ways to organize ourselves to do that. I think that this sort of a. Excuse me, This sort of system that gives us a global view of what humanity thinks we'll we'll, we'll get to your data question in a moment. If the data training these models were actually sort of a fair representation of of all human thoughts, that's something we've never had before, right? That's the the sort of thing that you would imagine would allow us to behave collectively much more effectively than we could. Without that global view about, you know, what we think on average, but also what we think from different perspectives. And I I don't have an answer to that. I think that there is a way in which these sorts of systems could lead us in that direction, lead us to something where we actually could cooperate much more effectively than we

currently do. But I don't think we have a science that quite tells us how to get there. And I'll leave the data disaster question for you.

01:04:53
I I want to share an idea about which I'm skeptical, but because I want to share some optimistic ideas and not just terrifying ones. I mean, I'm mostly terrified, but I'm trying to counteract that so so you know a lot of people are trying to figure out how human governance and decision making could work better. So our current system of. You know, we expose everybody to a bunch of media blather and social media with all its strengths and weaknesses, and then everybody votes or something. This doesn't seem to work that well. We seem very polarized. There isn't very much opportunity, forgive and take. There isn't very much of A sense that we're actually working with people we disagree with to together come closer to the truth. I don't think that's something that we feel that our current political institutions provide. People point to small scale things, you know, under the phrase deliberative democracy or citizen assemblies, like the small group assembly that Ireland used to update its abortion laws, for instance. And it seems like it's true when you get millions of us together on social media, we're total jerks. I mean, it's awful, but when you get like fifty of us together in A room. And we have some experts at our disposal. And it's happening. It's not public, right. It's private. So we're not on stage. We're not performing. We're not just, we're not making every utterance in an effort to win points or excite our base or build up our next campaign in that setting. People turn out to be pretty good at listening to each other. And often respecting each other's point of view. It's kind of surprising. So then there's this complaint that, oh, but that doesn't scale. Now I I think we focus too much on things. Scaling. Lots of good things don't scale, like a good espresso or a kindergarten classroom or a jury trial. We shouldn't try to make these things bigger. But one kind of idea, which I'll toss out there because of the sort of thing that some hopeful Silicon Valley people who want society to work better toss out, is, well, maybe if a lot of people in the city were discussing what they think is most important. In a kind of open-ended way, maybe something like a language model could say, Gee, there seem to be a lot of people over here who are saying this, they're concerned about this. Other people are saying that maybe it could act as a sort of moderator to collect and collate our views and perhaps even look for common ground. Big skepticism here. And of course, you can think of all sorts of dystopian scenarios where someone is manipulating this thing which is supposedly moderating the discussion. But I it was a new idea. It was one I hadn't heard before, so I don't want to immediately dismiss it. So AI assisted human social decision making. I don't know, maybe.

01:08:05
Could work. All right, I'll move back here. Mostly we get this one back here.

01:08:15
Hi, thanks so much for your talk today. My name is Greg and I'm a poet, and I've been collaborating with ChatGPT and Dolly too for a couple of months now, and I've been taking my poems and prompting using the poems as prompts. To generate images from Dolly 2 for example, and then iterating with the a I on those images and similarly I've been prompting ChatGPT with phrases of my own poetry to then create the next the next phrase or the next stanza in a poem.

01:09:03
And you know, it's.

01:09:04
Been a remarkable experience, honestly, because I know there's a lot of applications of a I relative to science and, you know, other kinds of things. But in the world of creativity, at first I was kind of intimidated and surprised, but my latest thought process and emotional response is actually very

positive. Because I started to understand that the training of these models came from human beings. So every single thing behind the model was a human being, image or text and and so I think there's a whole angle to this around human creativity and I'd be interested in in kind of your thoughts around that. I think that's really cool and and if and if you as a poet think that it's producing good poetry I think I should believe you. And I'd like your I like your point that in some sense what you're interfacing with through ChatGPT is all of the humans who created all of that text you know in. You could say optimistically. It's just a different, you know, we already read lots of poetry. We all we already listen to each other and read classic works, modern works. Maybe this is just sort of a new way to do that. I will say that my feeling is it's a little bit bland. I mean, for fun, I asked it to write like a sonnet about a can of a can of beans. It was a sonnet about a can of beans. It mostly scanned. I was okay. But maybe in collaboration with you, you're drawing more out of it and delving more deeply into it, and that sounds like a very interesting aesthetic experience to.

01:11:07
Me. Yeah. To me, what I I mean, it does sound fascinating. And thank you for sharing your experience. It suggests to me that you have learned how to use ChatGPT as a tool, just as we have. Accommodated other technological innovations through time and learn to use them as tools. And, you know, I'm a professor at a university and my university, probably all universities are grappling right now with, you know, the mundane. What do you know? What do we do about students who use chat, ChatGPT to write their essays? Can we even assign essays? There's all the sort of mechanics of educating students, but I think the deeper question is how should we be educating the next generation of students to use these tools effectively as you're and sounds like you've already figured it out. So that's that's great.

01:12:18
I have a friend in the music industry. Who asked me recently like who do you think is making interesting music now? And I and we were talking a little bit about sampling, right? And the ability to remix lots of music from the past which on I think there's a similar tension there right? Or or with auto tune and. You know, I was like, well, you know, Amy Winehouse did that great sampling to create that fantastic old sound. And he was like, no, she hired actual studio musicians who are really part of the lineage of of jazz and big band music. And who created that? I was like, oh, I I didn't know that. And so I'm of kind of, I'm sort of of two minds about this. On the one hand, you know, just as I use computational tools to do mathematics I couldn't otherwise do. And the comic book writer there used Mid Journey to create the art. I mean, imagine a world in which you can in which you can sort of play with as an editor and create feedback with but something which is making movies for you, you know, without having a movie crew. That would be bad news for a lot of humans who work in that industry. It would also perhaps unleash some new creative forces. Probably what would happen is we would all be making a billion movies and none of us would really watch any of them and most of them would be really bad. But but I think it's, it's interesting the theme of of augmentation or assistance as opposed to the threat of replacement. And I really feel both of those pretty strongly. You know, you could argue that creative geniuses will always find ways to express themselves whatever technology is available to them. I do worry about, well, if we don't have sort of that middle class of studio musicians and commercial artists and humans who are doing that, then you know that I think that would be bad on several levels. I also worry that. So like I I write, I've written, you know, I, I wrote a book, I write papers. You know, I'm not saying they're that great, but the struggle between the form and content is a really informative struggle. Right. You have an idea that's kind of vague. You try to put it into words. You have trouble putting it into words. Maybe it's because the idea is too vague to be worth, you know, expressing. But then something clicks and something feels right, and now you know you've got something. And I worry that if we get used to using tools like this to help us express ourselves, that we might lose part of that. It's like an artist who isn't, you know, really. There with the material medium. But of course

these are you know, you can find in in the printing press at the the wonderful printing press at the History Museum here there was a quote up saying these damn printers, people, they're not going to memorize texts anymore. They won't know how to recite them. They won't be able to feel them, you know everybody will just print a new copy whenever they want. What about the scribes? So these are very old debates.

01:15:40
I think Melissa has a question in the back.

01:15:48
Yes. So yeah, sorry, we're not in control of who gets the the critics of technological progress will say that technology is often very much like Pandora's box and it, you know causes harms to to nature and eliminates human freedoms. But there's much there's much problem in society. There's much conflict internationally and and even in domestic spheres around the world. And so it would seem that the current levels of technology have not worked to make a a large and complex society be cohesive and cooperative and unless one wants to. Go back to living by a fire in the woods and hunting. We're not going to give up hospitals and schools and jobs and travels and Internet and all these things. So it seems like we will. We will need to have a I's powers to modify people in such a way that they are more cooperative and that there stop being such conflicts in society. And so my question is for you experts. How do you think that we can get people to abandon these impediments to acceptance of a I guidance for society's benefit? If how do we get people to give up these concepts of like personal freedom or or wild nature so that they will? You know, cooperate and accept what they're told for the betterment of society. Society cannot diverge in multiple different directions. This is the source of conflicts we have. And if we do not stop these kind of conflicts, this is how we have international wars because and and a I can worsen that because it would tell some some ruler strike now. You know this. These people are most vulnerable now. You will not get another chance to strike now. And it will, it will dictate these kind of things. So we need to get people to stop resisting a I's direction and guidance for how to make society more cohesive. And do you have suggestions how we get people to understand they have to give up some things that are antiquated so that society can progress. You know it's it's going to be more difficult with the level of conflict we have now when the population grows further, right. So that that is my question. Well, as a large language model, I am sometimes incapable of recognizing irony and sarcasm I don't know you want.

01:18:36
One of you, well, I'll say, I'll say one thing I I couldn't quite tell from your question if you were for or against or both of them, but I think the most likely rollout. Of these technologies, isn't going to be the big splash like we've had in the past few months with ChatGPT. I think that's unusual. They're they're going to be enhancements to your word processor and, you know, browsers. Yeah, enhancements.

01:19:12
Yeah.

01:19:13
Yeah, and just just like. I don't really have free will over what movies I watch anymore because my streaming services think they already know what I want to watch and and unless unless I have a title in mind I get funneled in and and so I I think that for me is a real concern. I don't know if I can speak to the rest of the question. No one. He's going to take a crack. Yeah, take a crack of a small piece of it, which is the Pandora's Box problem. I mean, I do think that there is a little bit of potential that this isn't just a Pandora's box, but it's a Pandora's box full of Pandora's boxes. That

and you know, I think that there is a tremendous amount of hype about ChatGPT. It is nowhere, there's it's not in the neighborhood of achieving sentience or anything of that sort. I think, and I think that there's good evidence for that. But I do think that after a few years, maybe it's a couple of decades, but probably not that long, It probably will. You can imagine a system that is able to improve itself, right? It's pretty good at writing code. And so could a system like this start to sort of enhance itself? I mean, this is sort of the singularity idea. I think there's been a lot of hype around that. The idea that if we create a form of intelligence that is capable of of improving itself continuously, autonomously without our input, does it, does it go in a direction that we lose control of? I think that a lot of the debate about around that has been. Sort of simplified and and and what's built into this is the idea that the the, the AI itself has some goal. It has some sentience that has some desire. I don't think it needs any of that. I think it is possible that systems like this could be quite good at improving their own performance and might just take it in directions that that that the designers would. You have even less understanding of what what the machine is doing. And I'm less concerned about the machine itself doing something terrible with it than with sort of powerful actors of any sort, large companies, governments taking that ability and using it in a way that would really have negative consequences for ordinary people. So I do think that that we should think about that I should we should think about ways to have sort of guardrails that that sort of thing does not happen and it probably shouldn't just be left to a group of companies that are that are competing right to. To have the next hot model that everyone wants to use.

01:21:49
So no self modification.

01:21:51
No self modification. I don't know. And I think that that's impossible to enforce. Yeah. And to prevent, yeah, I don't know how you would stop it, but we should be aware that that might might be in our.

01:22:05
Future Shall we take two more questions, Caitlin. So thank you professors. I so appreciate your time and your expertise and answering those questions is something we are questioning and concerned about. And I feel as if my question may take more than just a few minutes to answer as it may take a course from one of yourselves as it involves computer science as I'm very curious about quantum computing and I hope that you could explain it a little further to us because I'm very limited in my understanding of it. You mentioned earlier professor, about how an A I is able to make sense of the letter A and I believe it's doing through ones and zeros. But that's not the case of quantum computing from what I understand and I hope that you can under help us understand this more. And does this change the systems like does the marriage of quantum computing and the systems you were referring to earlier, does it change? Is it the same? I'm just curious, does it? Has the game changed as we involve quantum computing with this Ai's I see a nod and I hope this is hitting on something that's going on in the industry right now is I'm very uneducated and I'm so happy to be here. But does this question makes sense? Yeah, it totally makes sense. I I I've worked a little bit in quantum computing not not recently and there's a lot of interesting advances there. I and I don't know what will happen in the future. I think right now these are two very separate branches. You know, the the things that these systems are doing, they seem to be able to do with conventional hardware, quantum computing. People have now built quantum computers with I think 50 or 60 or maybe 70 bits as they're called, which doesn't sound like a lot. It's just 70 things which can be on or off, true or false. That's as opposed to your phone, which has. Several billion of such things, but they're quantum bits and so they can be in these mysterious superpositions of on and off simultaneously. They can be entangled with each other. So you're right it's I think it's a it's probably not something we can handle within our team our time frame. I personally don't think that our

intelligence has much to do with quantum mechanics. There's a there sometimes people. Kind of conflate the idea of free will with the in determinism of quantum mechanics. I think that's a mistake. So I think the work in in quantum computing is very exciting, partly simply because we're learning to manipulate single particles and atoms in incredible ways, which I expect will have some other effects. But so I'm I'm fascinated by it, but I'm not sure that it's necessary or sufficient to make big advances in a I. So yeah, so one more, I guess I've.

01:25:13
Got three questions back here. We only have time for one more.

01:25:19
Well, since it's there really doesn't seem to be a very clearly defined definition of regular intelligence. How can you have artificial intelligence? A regular intelligence, right. And in my opinion, what's being called artificial intelligence is sophisticated computation. So do you have anything to say about that? I mean, I see the question was, if we can't define regular intelligence, then what are we? I mean then what do we mean by artificial intelligence? I I will say that there's a lot of things that are marketed as artificial intelligence that those of us in the field call machine learning or even statistics. So I think that there's, you know, there are lots of automated ways to look for patterns and data and try to predict that data. And then there are things more like ChatGPT and Mid Journey and there's a continuum in between. But I think some of the use of the word A I is hype. I mean lots of companies are saying we have a I to help you solve your sales issues or whatever. Come on, I don't know. But, but I I I like what Melanie said too, about holding up a mirror. I don't think it's easy to define intelligence, but by creating things which seem to be intelligent in certain ways or seem to act intelligent in certain ways, maybe that helps us sort of in dialogue with that process, figure out what intelligence is and and maybe what we want it to be. Well, I agree. And I think that's a good thing. Getting back down to basics, you know just the term Artificial Intelligence, if I'm not mistaken, was invented by a coin by a an IBM computer.

01:27:12
Science.

01:27:15
So it's really almost like a marketing tool. I I agree there's a lot of that as far as what chat, what it does with these kind of programs, even though they have neural networks and they have.

01:27:29
These. You know some.

01:27:31
Kind.

01:27:32
Of learning.

01:27:33
Experience. It's still sophisticated.

01:27:37
Computation that's. I mean that to me, that's why it's important that people not be intimidated or mystified and and really roll up their sleeves and try to understand what's hype and what's real. And I think the message of what we've been saying tonight is that. The people who create the

technologies and the people who study them as their profession don't completely understand them either. But I I think it's but I think, I think it's, you know, it's important that everybody come to come to terms with this technology rather than just listening to the hype. They think that's.

01:28:29
So I'm sorry, I'm happy to talk more, but yeah, good. Thank you.

01:28:33
I just want to make sure, yeah.