**CRAIG:** Hi, I’m Craig Smith and this is Eye on AI. This week, I talk to Ilya Sutskever, a cofounder and chief scientist of OpenAI and one of the primary minds behind the large language model GPT-3 and it’s public progeny, ChatGPT, which I don’t think it’s an exaggeration to say is changing the world.

**CRAIG:** This isn’t the first time Ilya has changed the world. Geoff Hinton has said he was the main impetus for AlexNet, the convolutional neural network whose dramatic performance stunned the scientific community in 2012 and set off the deep learning revolution.

**CRAIG:** As is often the case in these conversations, they assume a lot of knowledge on the part of listeners, primarily because I don’t want to waste the limited time I have to speak to people like Ilya explaining concepts or people or events that can easily be Googled – or Binged, I guess I should say – or that ChatGPT can explain for you.

**CRAIG:** The conversation with Ilya follows a conversation with Yann LeCun in the previous episode and so if you haven’t listened to that episode, I encourage you to do so. Meanwhile, I hope you enjoy the conversation with Ilya as much as I did.

**CRAIG:** Ilya, it's terrific to meet you, to talk to you. I've watched many of your talks online and, read many of your papers. Can you start just by introducing yourself, a little bit of your background. I know you were born in Russia where you were educated, what got you interested in computer science, if that was the initial impulse, or brain science, neuroscience or whatever it was, and then I'll start asking questions.

**ILYA:** Yeah, I can talk about that a little bit. So yeah, indeed, I was born in Russia. I grew up in Israel, and then as a teenager, my family immigrated to Canada. My parents say I was interested in AI from a pretty early age. I also was very motivated by consciousness. I was very disturbed by it, and I was curious about things that could help me understand it better. And AI seemed like a very, like a good angle there. So, I think these were some of the ways that got me started.

**ILYA:** And I actually started working with Geoff Hinton very early when I was 17. Because we moved to Canada and I immediately was able to join the University of Toronto and I really wanted to do machine learning, because that seemed like the most important aspect of artificial intelligence that at the time was completely inaccessible.

**ILYA:** Like to give some context, the year was 2003. We take it for granted that computers can learn, but in 2003, we took it for granted that computers *can't* learn. The biggest achievement of AI back then was Deep Blue, the chess playing engine.

**CRAIG:** Yeah.

**ILYA:** But there it was like you have this game and you have this research, and you have this simple way of determining if one position is better than another.

**ILYA:** And it really did not feel like that could possibly be applicable to the real world because there is no learning. And learning was this big mystery. And so, I was really, really interested in learning and, to my great luck, Geoff Hinton was a professor in the university I was in, and so I was able to find him, and we began working together almost right away.

**CRAIG:** And was your impulse, as it was for Geoff, to understand how the brain worked or was it more that, you were simply interested in the idea of machines learning? And did, at that time, you have the intuition that computers could someday perform mental tasks similar to those of a human.

**ILYA:** AI is so big and so the motivations were just as many, like it is interesting like how does intelligence work at all?

**ILYA:** Like right now we have quite a bit of an idea. It's a big neural net and we know how it works to some degree. But back then, although neural nets were around, no one knew that neural nets were good for anything. So how does intelligence work at all? How can we make computers be even slightly intelligent?

**ILYA:** And I had a very explicit intention to make a very small, but real contribution to AI. Because there were lots of contributions to AI which weren't real, which were like, I could tell for various reasons that they weren't real, that nothing would come out of it. And I just thought, nothing works at all.

**ILYA:** AI is a hopeless field. So, the motivation was, could I understand how intelligence works? And also make a contribution towards it. So that was my initial early motivation. That's 2003. It's almost, exactly 20 years ago.

**CRAIG:** And then AlexNet, and I've spoken to Geoff, and he said that it was really your excitement about the breakthroughs in convolutional neural networks that led you to apply for the ImageNet competition and that that Alex had the coding skills to train the network.

**CRAIG:** Can you talk just a little bit about that? I don't want to get bogged down in history, but it's fascinating.

**ILYA:** So, in a nutshell, I had the realization that if you train, a large neural network on a large -- sorry, large, *and deep*, because back then the deep part was still new -- if you train a large and a deep neural network on a big enough dataset that specifies some complicated task that people do such as vision, but also others, and you just train that neural network then you will succeed necessarily. And the logic for it was very irreducible, where we know that the human brain can solve these tasks and can solve them quickly. And the human brain is just a neural network with slow neurons.

**ILYA:** So, we know that some neural network can do it really well. So, then we just need to take a smaller but related neural network and just train it on the data. And the best neural network inside the computer will be related to the neural network that we have that performs this task.

**ILYA:** So, it was an argument that the neural network, the large and deep neural network can solve the task. And furthermore, we have the tools to train it. That was the result of the technical work that was done in Geoff's lab. So, you combine the two, we can train those neural networks. It needs to be big enough so that if you trained it, it would work well, and you need the data, which can specify the solution.

**ILYA:** And with ImageNet, all the ingredients were there. Alex had these very fast convolutional kernels. ImageNet had large enough data, and there was a real opportunity to do something totally unprecedented, and it totally worked out.

**CRAIG:** Yeah. That, that was supervised learning and convolutional neural nets. In 2017, the "Attention Is All You Need" paper came out introducing self-attention and transformers. At what point did the GPT project start? Was it, was there some intuition about transformers?

**ILYA:** Yeah.

**CRAIG:** And self-supervised learning? Can you talk about that?

**ILYA:** So, for context, at OpenAI from the earliest days, we were exploring the idea that predicting the next thing is all you need. We were exploring it with the much more limited neural networks of the time, but the hope was that if you have a neural network that can predict the next word, the next pixel really, it's about compression.

**ILYA:** Prediction is compression and predicting the next word is not … Let's see, let me think about the best way to explain it, because there were many things going on and they were all related.

**ILYA:** maybe I'll take a different direction. We were indeed interested in trying to understand how far predicting the next word is going to go and whether it'll solve unsupervised learning. So back before the GPTs, unsupervised learning was considered to be the Holy Grail of machine learning.

**ILYA:** Now it's just been fully solved, and no one even talks about it, but it was a Holy Grail. It was very mysterious, and so we were exploring the idea. I was really excited about it, that predicting the next world well enough is going to give you unsupervised learning. If it'll learn everything about the data set, that's going to be great.

**ILYA:** But our neural networks were not up for the task. We were using recurrent neural networks. When the transformer came out, it was literally as soon as the paper came out, literally the next day, it was clear to me, to us that transformers addressed the limitations of recurrent neural networks, of learning long-term dependencies.

**ILYA:** It's a technical thing. But it was like we switched to transformers right away. And so, the very nascent GPT effort continued then, and then like with the transformer, it started to work better, and you make it bigger, and then you realize, to keep making it bigger.

**ILYA:** And we did. And that's what led to eventually GPT-3 and essentially where we are today.

**CRAIG:** Yeah. And I, just wanted to ask, actually, I'm getting caught up in this history, but I'm so interested in it. I want to get to the problems or the shortcomings of large language models or large models generally.

**CRAIG:** But Rich Sutton had been writing about scaling and how that's all we need to do. We don't need new algorithms; we just need to scale. Did he have an influence on you or was that a parallel track of thinking?

**ILYA:** No, I would say, that when he posted his article, then we were very pleased to see some external people thinking in similar lines, and we thought it was very eloquently articulated, but I actually think that the "Bitter Lesson" as articulated overstates its case. Or, at least I think the takeaway that people have taken from it, overstates case. The takeaway that people have is, ‘doesn't matter what you do, just scale.’

**ILYA:** But that's not exactly true. You’ve got to scale something specific. You’ve got to have something that you'll be able to benefit from the scale. The great breakthrough of deep learning is that it provides us with the first ever way of productively using scale and getting something out of it in return. Like before that, like what would people use large computer clusters for?

**ILYA:** I guess they would do it for weather simulations or physics simulations or something, but that's about it. Maybe movie making. But no one had any real need for computer clusters because what do you do with them? The fact that deep neural networks, when you make them larger and you train them on more data, work better, provided us with the first thing that is interesting to scale.

**ILYA:** But perhaps one day we will discover that there is some little twist on the thing that we scale that's going to be even better to scale. Now how big of a twist? And then of course with the benefit of hindsight, you'll say does it even count? It's such a simple change. But I think the true statement is that it matters what you scale.

**ILYA:** Right now, we just found like a thing to scale that gives us something in return.

**CRAIG:** The limitation of large language models as they exist is, their knowledge is contained in the language that they're trained on. And most human knowledge I think everyone agrees is non-linguistic.

**CRAIG:** I'm not sure Noam Chomsky agrees, but there's a problem in large language models as I understand it. Their objective is to satisfy the statistical consistency of the prompt. They don't have an underlying understanding of reality that language relates to. I asked ChatGPT about myself.

**CRAIG:** It recognized that I'm a journalist, that I've worked at these various newspapers, but it went on and on about awards that I've never won. And it all read beautifully, but none of it connected to the underlying reality. Is there something that is being done to address that in your research going forward?

**ILYA:** Yeah.

**ILYA:** So, before I comment on the immediate question that you ask, I want to comment about some of the earlier parts of the question.

**CRAIG:** Sure.

**ILYA:** I think that it is very hard to talk about the limits or limitations rather of even something like a language model. Because two years ago, people confidently spoke about their limitations, and they were entirely different, right? So, it's important to keep this context in mind. How confident are we that these limitations that we see today will still be with us two years from now? And I am not that confident. There is another comment I want to make about one part of the question, which is that these models just learn statistical regularities and therefore they don't really know what the nature of the world is.

**ILYA:** And I have a view that differs from this. In other words, I think that learning the statistical regularities is a far bigger deal than meets the eye. The reason we don't initially think so is because we haven't -- at least most people, those who haven't really spent a lot of time with neural networks, which are on some level statistical, like what is a statistical model?

**ILYA:** You just fit some parameters like what is really happening. But think there is a better interpretation to the earlier point of prediction as compression. Prediction is also a statistical phenomenon. Yet to predict you eventually need to understand the true underlying process that produced the data to predict the data well, to compress it well, you need to understand more and more about the world that produced the data.

**ILYA:** As our generative models become extraordinarily good, they will have, I claim, a shocking degree of understanding. A shocking degree of understanding of the world. And many of its subtleties, but it's not just the world. It is the world as seen through the lens of text. It tries to learn more and more about the world through a projection of the world on the space of text as expressed by human beings on the internet.

**ILYA:** But still, this text already expresses the world. And I'll give you an example, a recent example, which I think is really telling and fascinating. So, we've all heard of Sydney being its alter-ego. And I've seen this really interesting interaction with Sydney where Sydney became combative and aggressive when the user told it that it thinks that Google is a better search engine than Bing.

**ILYA:** Now, how can we, like what is a good way to think about this phenomenon? What's a good language? What's, what does it mean? You can say, wow, like it's just predicting what people would do and people would do this, which is true, but maybe we are now reaching a point where the language of psychology is starting to be appropriated to understand the behavior of these neural networks.

**ILYA:** Now let's talk about the limitations. It is indeed the case that these neural networks are, they do have a tendency to hallucinate, but that's because a language model is great for learning about the world, but it is a little bit less great for producing good outputs. And there are various technical reasons for that, which I could elaborate on if you think it's useful, but at this, right now, like at this second I will skip that. There are technical reasons why a language model is much better at learning about the world, learning incredible representations of ideas, of concepts, of people, of processes that exist, but its outputs aren't quite as good as one would hope or as or rather as good as they could be.

**ILYA:** Which is why, for example, for a system like ChatGPT, is a language model that has an additional reinforcement learning training process. We call it reinforcement learning from human feedback. But the thing to understand about that process is this.

**ILYA:** We can say that the pre-training process, when you just train a language model, you want to learn everything about the world. Then the reinforcement learning from human feedback. Now we care about the outputs. Now we say, anytime the output is inappropriate, don't do this again. Every time the output does not make sense, don't do this again.

**ILYA:** And it learns quickly to produce good outputs. But now it's the level of the outputs, which is not the case during pre-training, during the language model training process. Now on the point of hallucinations. And it has a propensity of making stuff up. Indeed, it is true.

**ILYA:** Right now, these neural networks, even ChatGPT, makes things up from time to time, and that's something that also greatly limits their usefulness. But I'm quite hopeful that by simply improving this subsequent reinforcement learning from human feedback step, we could just teach it to not hallucinate. Now you could say is it really going to learn?

**ILYA:** My answer is, let's find out.

**CRAIG:** And that feedback loop is coming from the public ChatGPT interface, that if it tells me that I won a Pulitzer, which unfortunately I didn't, I can tell it that it's wrong and, will that train it or, create some punishment or reward so that the next time I ask it'll, be more accurate.

**ILYA:** The way we do things today is that we hire people to teach our neural network to behave, to teach ChatGPT to behave. And right now, the manner, the precise manner in which they specify the desired behavior is a little bit different. But indeed, what you described is the way in which teaching is going to like basically to be, that's the correct way to teach.

**ILYA:** You just interact with it, and it sees from your reaction, it infers, oh, that's not what you wanted. You are not happy with its output. Therefore, the output was not good, and it should do something differently next time. So, in particular hallucinations come up as one of the bigger issues and we'll see.

**ILYA:** But I think there is a quite a high chance that this approach will be able to address them completely.

**CRAIG:** I wanted to talk to you about Yann LeCun's work on joint embedding predictive architectures. And his idea that what's missing from large language models is this underlying world model that is non-linguistic that the language model can refer to.

**CRAIG:** It's not something that's built, but I wanted to hear what you thought of that and whether you've explored that at all.

**ILYA:** So, I reviewed Yann LeCun's proposal and there are a number of ideas there, and they're expressed in different language and there are some maybe small differences from the current paradigm, but to my mind, they are not very significant.

**ILYA:** And I'd like to elaborate. The first claim is that it is desirable for a system to have multimodal understanding where it doesn't just know about the world from text. And my comment on that will be that indeed multimodal understanding is desirable because you learn more about the world, you learn more about people, you learn more about their condition, and so the system will be able to understand what the task that it's supposed to solve, and the people and what they want better.

**ILYA:** We have done quite a bit of work on that, most notably in the form of two major neural nets that we've done. One is called Clip and one is called Dall-E. And both of them move towards this multimodal direction. But I also want to say, that I don't see the situation as a binary either or that if you don't have vision, if you don't understand the world visually or from video, then things will not work.

**ILYA:** And I'd like to make the case for that. So, I think that some things are much easier to learn from images and diagrams and so on, but I claim that you can still learn them from text only, just more slowly. And I'll give you an example. Consider the notion of color.

**ILYA:** Surely one cannot learn the notion of color from text only and yet when you look at the embeddings -- I need to make a small detour to explain the concept of an embedding. Every neural network represents words, sentences, concepts through representations, embeddings, high-dimensional vectors.

**ILYA:** And one thing that we can do is that we can look at those high-dimensional vectors and we can look at what's similar to what, how does the network see this concept or that concept? And so, we can look at the embeddings of colors and embeddings of colors happen to be exactly right. You know, it like it knows that purple is more similar to blue than to red, and it knows that purple is less similar to red than orange is. It knows all those things just from text. How can that be? So, if you have vision, the distinctions between color just jump at you. You immediately perceive them. Whereas with text, it takes you longer, maybe you know how to talk, and you already understand syntax and words and grammars, and only much later you say, oh, these colors actually start to understand them.

**ILYA:** So, this will be my point about the necessity of multimodality, which I claim it is not necessary, but it is most definitely useful. I think it's a good direction to pursue. I just don't see it in such stark either-or claims.

**ILYA:** So, the proposal in the paper makes a claim that one of the big challenges is predicting high dimensional vectors which have uncertainty about them.

**ILYA:** So, for example, predicting an image like the paper makes a very strong claim there that it's a major challenge and we need to use a particular approach to address that. But one thing which I found surprising, or at least unacknowledged in the paper, is that the current autoregressive transformers already have the property.

**ILYA:** I'll give you two examples. One is, given one page in a book, predict the next page in a book. There could be so many possible pages that follow. It's a very complicated, high-dimensional space, and they deal with it just fine. The same applies to images. These autoregressive transformers work perfectly on images.

**ILYA:** For example, like with OpenAI, we've done work on the iGPT. We just took a transformer, and we applied it to pixels, and it worked super well, and it could generate images in very complicated and subtle ways. It had the very beautiful unsupervised representation learning. With Dall-E 1, same thing again.

**ILYA:** You just generate, think of it as large pixels. Like rather than generate a million pixels, we cluster the pixels into large pixels, and they generate a thousand large pixels. I believe Google's work on image generation from earlier this year called Party, I believe they also took a similar approach.

**ILYA:** So, the part where I thought that the paper made a strong comment around where current approaches can't deal with predicting high dimensional distributions. I think they definitely can. So maybe this is another point that I would make.

**CRAIG:** And then what you're talking about, converting pixels into vectors, it's essentially turning everything into language. A vector is like a string of text, right?

**ILYA:** Define language though. You turn it into a sequence. Yeah, a sequence of what, like you could argue that even a human life is a sequence of bits.

**ILYA:** Now there are other things that people use right now, like diffusion models where they produce those bits rather than one bit at a time, they produce them in parallel, but I would argue that, on some level, this distinction is immaterial.

**ILYA:** I claim that on some level it doesn't really matter. It matters as in like you can get a 10x efficiency gain, which is huge in practice. But conceptually, I claim it doesn't matter

**CRAIG:** On this idea of having an army of human trainers that are working with ChatGPT or a large language model to guide it in effect with reinforcement learning.

**CRAIG:** Just intuitively, that doesn't sound like an efficient way of teaching a model about the underlying reality of its language.

**ILYA:** Yeah.

**CRAIG:** Isn't there a way of automating that and, to Yann's credit, I think that's what he's talking about is, coming up with an algorithmic means of teaching a model the underlying reality without a human having to intervene.

**ILYA:** Yeah. So, I have two comments on that. I think. So, the first place, so I have a different view on the question, so I wouldn't agree with the phrasing of the question. I claim that our pre-trained models already know everything they need to know about the underlying reality. They already have this knowledge of language and also a great deal of knowledge about the processes that exist in the world that produce this language.

**ILYA:** And maybe I should reiterate this point, it's a small tangent, but I think it's so important, the thing that large generative models learn about their data -- and in this case, large language models -- about text data are some compressed representations of the real-world processes that produced this data, which means not only people and something about their thoughts, something about their feelings, but also something about the condition that people are in and the interactions that exist between them. The different situations a person can be in. All of these are part of that compressed process that is represented by the neural net to produce the text. The better the language model, the better the generative model, the higher the fidelity, the more, the better this, the better it captures this process.

**ILYA:** So that's the first comment that we make. And so, in particular, I will say the models already have the knowledge. Now, the army of teachers, as you phrase it, indeed, you know when you want to build the system that performs as well as possible, you just say, okay, like if this thing works, do more of that. But of course, those teachers are also using AI assistance.

**ILYA:** Those teachers aren't on their own. They're working with our tools together. They're very, efficient. It's like the tools are doing the majority of the work, but you do need to have, you need to have oversight; you need to have people reviewing the behavior because you want to have to eventually to achieve a very high level of reliability.

**ILYA:** But overall, I'll say that we are at the same time, this second step after we take the finished pre-trained model and then we apply the reinforcement learning on it, there is indeed a lot of motivation to make it as efficient and as precise as possible so that the resulting language model will be as well behaved as possible.

**ILYA:** So yeah, there are these human teachers who are teaching the model desired behavior. They're also using AI assistance. And the manner in which they use AI systems is constantly increasing, so their own efficiency keeps increasing. So maybe this will be one way to answer this question.

**CRAIG:** Yeah. And, so what you're saying is through this process, eventually the model will become more and more discerning, more and more accurate in its outputs.

**ILYA:** Yes. And it's -- that's right. There is an analogy here, which is it already knows all kinds of things, and now we just want to really say, no, this is not what we want.

**ILYA:** Don't do this here. You made a mistake here in the output. And of course, it's exactly as you say, with as much AI in the loop as possible so that the teachers who are providing the final correction to the system, their work is amplified. They're working as efficiently as possible.

**ILYA:** So, it's not unlike an education process, how to act well in the world.

**ILYA:** We need to do additional training, just to make sure that. The model knows that hallucination is not okay ever. And then once it knows that, now you are in business. And it's that reinforcement learning human teacher loop that will teach it.

**ILYA:** Human teacher loop or some other variant. But there is definitely an argument to be made that something here should work. And we will find out pretty soon either way.

**CRAIG:** That's one of the questions, where is this going? What, research are you focused on right now?

**ILYA:** I can't talk in detail about the specific research that I'm working on, but I can mention a little bit. I can mention some of the research in broad strokes and it would be something like -- I'm very interested in making those models more reliable, more controllable, make them learn faster from lesson data, less instructions. Make them so that indeed they don't hallucinate. And I think that all this cluster of questions, which I mentioned, they're all connected. And there's also a question of how far in the future are we talking about in this question?

**ILYA:** And what I commented here on is the perhaps nearer future.

**CRAIG:** You talk about the similarities between the brain and neural nets. There's a very interesting observation that Geoff Hinton made to me. I'm, sure it's not new to other people, but that large models, or large language models in particular hold a tremendous amount of data with a modest number of parameters compared to the human brain, which has trillions and trillions of parameters, but a relatively small amount of data.

**CRAIG:** Have you thought of it in those terms? And can you talk about what's missing in large models to have more parameters to handle the data? Is that a hardware problem or a training problem?

**ILYA:** This comment which you made is related to one of the problems that I mentioned in the earlier questions of learning from less data.

**ILYA:** Indeed, the current structure of the technology does like a lot of data, especially early in training. Now later in training, it becomes a bit less data hungry, which is why in the end it can learn very, not as fast as people yet, but it can learn quite quickly.

**ILYA:** So already that means that in some sense, do we even care that we need all this data to get to this point. But indeed, more generally I think it'll be possible to learn more from less data. I think it's just; I think it requires some creative ideas, but I think it is possible, and I think learning more from less data will unlock a lot of different possibilities.

**ILYA:** It'll allow us to teach our AI the skills that is missing and to convey to it our desires and preferences exactly how we want it to behave more easily. So, I would say that faster learning is indeed very nice. And although already after language models are trained, they can learn quite quickly.

**ILYA:** I think there are opportunities to do more there.

**CRAIG:** I heard you make a comment that, that we need faster processors to be able to scale further. And it appears that the scaling of models, that there's no end in sight, but the power required to train these models, we're reaching the limit, at least the socially accepted limit.

**ILYA:** So, I just want to make one comment, which is I don't remember the exact comment that I made that you're referring to, but you always want faster processors, of course you always want more of them. Of course, power keeps going up. Generally speaking, the cost is going up.

**ILYA:** And the question that I would ask is not whether the cost is large, but whether the thing that we get out of paying this cost outweighs the cost. Maybe you pay all this cost, and you get nothing, then yeah, that's not worth it. But if you get something very useful, something very valuable, something you can solve a lot of problems that we have, which we really want solved, then the cost can be justified.

**ILYA:** But in terms of the processors, faster processors. Yeah. Any day.

**CRAIG:** Are you involved at all in the hardware question? Do you work with Cerebras, for example, the wafer scale chips.

**ILYA:** No, all our hardware comes from Azure and GPUs that they provide us.

**CRAIG:** Sure. Yeah. Yeah. You did talk at one point I saw about democracy and about the impact that that AI can have on, democracy. If you had enough data and a large enough model, you could train the model on the data and it could come up with an optimal solution that would satisfy everybody.

**CRAIG:** And do you have any aspiration or, do you think about where this might lead in terms of helping humans manage society?

**ILYA:** Yeah, let's see. It's such a big question because it's a much more future looking question. I think that there are still many ways in which our models will become far more capable than they are right now.

**ILYA:** There's no question. In particular, the way we train them and use them and so on, there's going to be a few changes here and there. They might not be immediately obvious today, but I think in hindsight it will be extremely obvious that will indeed allow it to have the ability to come up with solutions to problems of this kind. It's unpredictable exactly how governments will use this technology as a source of getting advice of various kinds. I think that to the question of democracy, one thing which I think could happen in the future is that because you have these neural nets and they're going to be so pervasive and they're going to be so impactful in society, we will find that it is desirable to have some kind of a democratic process where, let's say the citizens of a country provide some information to the neural net about how they'd like things to be, how they'd like it to behave, or something along these lines.

**ILYA:** I could imagine that happening. That can be a very like, a high bandwidth form of democracy perhaps, where you get a lot more information out of each citizen and you aggregate it, specify how exactly we want such systems to act. Now it opens a whole lot of questions, but that's one thing that could happen in the future.

**CRAIG:** Yeah.

**CRAIG:** And I can see in the democracy example, you give that, that individuals would have the opportunity to input data. But, and this sort of goes to the world model question, do you think AI systems will eventually be large enough that they can understand a situation and analyze all of the variables. But you would need a model that does more than absorb language, I would think.

**ILYA:** What does it mean to analyze all the variables? Eventually there will be a choice you need to make where you say, these variables seem really important. I want to go deep because a person can read the book.

**ILYA:** I can read a hundred books, or I can read a book very slowly and carefully and get more out of it. So, there will be some element of that. Also, I think it's probably fundamentally impossible to understand everything in some sense. Let's, take some easier examples.

**ILYA:** Anytime there is any kind of complicated situation in society, even in a company, even in a mid-size company, it's already beyond the comprehension of any single individual. And I think that if we build our AI systems the right way, I think AI could be incredibly helpful in pretty much any situation.

**CRAIG:** That’s it for this episode. I want to thank Ilya for his time. I also want to thank Elie Georges for helping arrange the call. If you want to read a transcript of this episode, you can find one on our website, eye-on.ai. We love hearing from listeners so feel free to email me at craig@craigsmith.ai. I get a lot of emails, so put Listener in the subject line so I don’t miss it. Believe it or not, we have listeners in 170 countries and territories.

**CRAIG:** Remember, the Singularity may not be near, but AI is changing your world, so, pay attention.