AI and You

Transcript

Special Episode: 2022 Retrospective and 2023 Prediction Panel

Episode 132

First Aired: Monday, December 26, 2022

Hello, and welcome to episode 132! Today, in what is fast becoming a tradition, we have a special end of year episode, recorded just before its air date of December 26, 2022. We are going to be talking about 2022 in retrospective and making some predictions for 2023, and I am glad to be joined by two people whom I am happy to have met through this work and who have been on the podcast before. Dan Turchin, who was on episodes 122 and 123, and who is CEO of PeopleReign and is host of the "AI and the Future of Work" podcast, on which this discussion will also be aired. And David Wood, who was on episodes 20 and 21 and who is the chair of the London Futurists and author of the recent book, "The Singularity Principles: Anticipating and managing cataclysmically disruptive technologies."

I've written about the folly of trying to predict the future when it's more important to be prepared for unpredictable disruption; so take what follows for entertainment value more than betting your life savings on, but here you have three people who are very well informed on trends in artificial intelligence, so as far as predictions go, these should be some of the best you'll find. Let's get right to it.

Peter: All right. So, I'm very excited to be joined here by David Wood and Dan Turchin, and welcome to the show, guys.

David: It's great to be here. Thanks for the invite.

Dan: Peter, I'm looking forward to this. Thanks for having me.

Peter: And so, I want to start with a retrospective of 2022 as it was, and I just want to focus this actually for a moment on the hottest topic at the moment of ChatAI, ChatGPT. And so, I thought of a question along these lines, because I really want to know the answer to this. So, I want you to imagine that I've got a time machine. And we are conducting this show one year ago, and it's 2021, and we are making predictions for 2022. But I am gifted with some foresight, and I describe to you precisely the capabilities of the foundation models that power image synthesis, like DALL-E-2, MidJourney, Stable Diffusion. And number two, I described to you the capability of ChatGPT and I communicate that those capabilities perfectly to you. And then I asked you, what do you think the chances are, that that technology will be invented and deployed to the internet in the following year? I want to know what your answers to that would be.

David: Well, I went on record in June 2021 in a chapter on a book that I wrote, in which I said here are fifteen areas in which it is possible that AI can make significant improvements soon. Now, I did not give particular timescales and I am cautious of any futurist who does offer specific timescales for these things. But I said there is so much commercial interest in

improvements to AI and there are so many ideas for how AI could indeed be extended, that it is quite likely that at least some of these fifteen areas will see advances and that include better analysis of text, I referred to GPT-3 and said we're likely to see some progress and it also referred to progress. I didn't talk about Stable Diffusion. But I did talk about generative adversarial networks and said I'm sure we haven't seen the end of that yet. But what would have surprised me and would have been difficult to forecast is exactly how much has been accomplished by it. Apparently some fairly modest changes in the algorithms that for just a bit of extra scale, and for some significant tweaks in the algorithm, some new ways of using transformers and so on, the effects have gotten so much better in so many ways, that would have taken me by surprise, and I think that reflects the state of knowledge of the industry generally. I have not been the only force observer of AI to have been surprised by what's happened. Many others have said some of the disappointments are disappointing, but some of the accomplishments are breathtaking.

Peter: Thank you. Dan, what does 2021 you have to say?

Dan: The 2021 me would say that generative AI as both a technology and investment theme were nascent, and David shares great perspectives and based on some of his work from a year back. I'm sitting in the cradle Silicon Valley, so, I'm going to think about this from the perspective of investors and there were some early leaders in the space that were getting a lot of venture capital funding; copycat AI comes to mind, using generative AI to write marketing copy, using generative AI to author summaries of large documents, PDFs, that sort of thing. Those were early ideas, DALL-E-2 had just been introduced, I think maybe came out of public beta around a year ago and it was a novelty and it was fun to be able to see a picture of a panda riding a rocket ship, in the style of Renoir and we all got giggles out of that, and it seemed to a lay person like magic and it was probably a bit of a conundrum to everybody who didn't have all the imagination of some technologists, certainly at least in Silicon Valley, to envision what would become of the panda rocket ship generator. So, I think we would have said, it's compelling, it's a novelty. It's a parlor trick and we were the 2021 version of me, didn't have the foresight to see either the hype that would surround generative AI technologies with the launch of ChatGPT, call it a year later, or certainly all of the innovation around generative AI that would be spawned from some of the early capabilities of simple text prompt to image generators. That's a kind of the most humble assessment of I think what the 2021 version of me would have said.

Peter: And I think for my part, the things that surprised me were how both of those technologies generated more than what their predecessors promise should have been capable of; that they're essentially putting the internet in a blender and giving you part of it. But then, these tools are as though an orange emerged from the blender, or whatever it is that you asked for, so that the generative image model generators for instance, you can ask it for a corgi sitting in a house made of sushi. Now, nowhere in the training data is a house made of sushi. It's got sushi, and it's got dog houses and it's got corgis. Fine, but the progress in image generation up

until that point, I always think of as the sort of thing that Janelle Shane was doing and, in her newsletters, and she was doing things with CLIP and CAN and you would get these Salvador Daliesque type images that were clearly the result of convolving some image that she was looking for with something else and it might look like something that HR Giger had created. But it wouldn't have exactly the perspective - and this was the thing that got me about that dog house made of sushi, that it was all arranged in the right scale, perspective, consistent lighting, that imputes some sort of understanding and this is where I get a problem describing a technology like ChatGPT without using words that make it sound like here is something that has suddenly involved general artificial intelligence, become conscious/alive/aware. That I feel there's a word between analysis and understanding that ought to be applied here, but I don't know if that word exists and I've seen ChatGPT do things that I would have said were impossible before, and again, don't want to the audience taking away the message that we're sounding alarm here, just that I don't find the explanations of how that technology works adequate to tell me how it's doing some of the things that I have seen it do. I needed an answer once, for instance, where I said, I wanted to know, how do I rewrite this code from PHP 7 to PHP 8. That's a very specialized sort of thing. But I was stuck; the general answers weren't answering the specialized case that I had of this, and I needed to do this. And just for a lark - because I thought, no way is it going to know this - I stuck that into there and not only did it give me a complete explanation that was far better than anything I'd found on Stack Overflow or Google. But it gave three examples explaining this using the same variables I had in the code that I had quoted. That was far outside the capability of any previous model for chat and so, I want to segue now to the question about this, since we're talking about this, which is what sort of disruption can you see this generating vertically or horizontally? Now, that everyone is talking about this - beyond the term papers are dead as a pedagogical instrument, which is the main reaction at the moment?

Dan: I had two recent guests on the podcast Peter; after your great episode, that both served to introduce, call them antipatterns, into the discussion about generative AI, it's easy to get caught up in just kind of unbridled enthusiasm because a lot of what you see in the popular press are examples of where it truly does seem like magic or maybe to use slightly more cynical terms, it's a reminder of how vast the internet is. When I asked some of these questions to a gentleman named Eric Olson, who started a company called Consensus, which is basically using a GPT-3-like technology to summarize questions to research related results by indexing the world's research journals, all scientifically published papers. And so it can go and mine through and actually the idea is consensus will give some kind of a score. So you ask it a question, is, "What's the right amount of vitamin C," or, "Should I eat a lot of calcium before working out?" Things where there's some amount of controversy and some amount of valid scientific research. And it'll go and actually comb through voluminous scientific research articles and come back with a summary and some pros and cons and obviously intended to be very thorough and a balanced approach. And even Eric Olson, who's the CEO of the company, said we're keeping it in beta, because there are a lot of examples of where it performs really well on

certain topics. But when you ask it physics questions, for example, it reveals some holes and gaps in the literature that are easily exposed when you ask it you know, when a scientist asked questions and expects to get an unbiased summary. And then the second guest, is doing something fascinating with generative AI that, also revealed some holes in the robustness of the technologies gentlemen, Michael Osterrieder, who started a company called vAIsual, a YC backed company, that is using generative AI to essentially replace the Getty images and the vast repositories of these images online that you can license with generative AI that will synthetically generate very similar images using text prompts. And you can imagine how disruptive technology like that is. And in the episode, he goes into a lot of detail about the technical challenges of being able to, first off, modify sometimes copyrighted content, or in his case, I think he exercises AI very responsibly, they're careful not to ingest anything that's already copyrighted. But the technology is, it's very complex, but it's easy to understand when he goes through it. They're essentially ingesting publicly available content, images, etc., or in some cases, taking their own photos in a studio. But then using text prompts, they're able to annotate them automatic or semi automate the annotation of those photos with things like expressions or limbs or settings, light settings, etc. And what they're basically doing is introducing based on text prompts, introducing new versions of those photos, using generative AI in ways that are 100% synthetically generated, but all licensed content. And essentially eliminates the need beyond a photoshoot with a small number of photos let's say, you know, a Middle Eastern woman in a certain kind of head dress, in a market. Well, Michaels' team will bring a Middle Eastern woman into a photo studio and with a green screen, and maybe, take 50 photos of her, you know, with various expressions on her face and from that, with the benefit of the green screen and various other, can spawn millions of images, all related to, , the, the core concept of the Middle Eastern woman various poses with various expressions. So, two examples where I got the benefit of having a deep dive with these guests in the podcast, but where I think the technology is certainly viable. But I definitely encourage listeners to understand the sophistication and the rigor required to really get it right and do it in a way that's ethically responsible.

Peter: And that sounds like the either modification or extension of an existing job or creation of a new type of job. And I'm aware, by the way that I've skipped forward here, we will get back to some of the a, retrospectives of 2022 shortly. But as long as I'm talking about this; David, what is your take on the disruptive effect of these two new technologies in AI: image generation and the ChatGPT that haven't been fully appreciated yet?

David: We're still early days with this. Dan mentioned some of the pain points, some of the complexity, some of the need to understand the sophisticated level what's going on, if users of these systems aren't to fall foul. But that's the curve that new innovations go through, they start off with lots of ragged edges with difficulties in getting the best out of them and over time, they are increasingly packaged in more accessible, more robust ways, ways that are usable by ordinary folk, as opposed to experts. And I predict this is what's going to happen with many of the systems we're already hearing about, indeed, ChatGPT, by some measures, has more or less

the same underlying engine - some modifications - as GPT-3. The main innovation is that it's got a simpler interface that's got some memory, and you can chat to it. Whereas before, you needed to be quite clever to get that kind of chat experience from it. So, there's one modification that's already making it easier to use. But I think it's going to be the same with for example, software that looks up scientific facts, answers questions about science. Famously, or infamously, there was the Galactica software that was released by Meta, parent of Facebook briefly, which stunned people briefly before people realized it was hallucinating, fabricating lots of reports, apparently, in scientific journals, which never existed and so, understandably, Facebook hurriedly withdrew it. But you can imagine a combination, that one software is doing what Galactica is currently doing, coming up with plausible stories and another piece of software is fact checking it, seeing whether or not these articles exist, seeing whether the summary of them is fair and it's going to be the combinations of technologies that's going to lead to the biggest and fastest transformations, and it won't to just be the blue-collar work, it's very much going to be the designer's work. Instead of design currently being done by talented human individuals with help from AIS, with help from spell checkers, with help from layout tools is going to largely become the other way around: you'll just give an idea to the AI tool and then you maybe have a human at the end to just doing a sanity check on it. So this is going to be incredibly disruptive, more so as combinations of different AIs are grouped together in packages that deliver more than the sum of their parts.

Peter: Let's look at the past year, then, and, thank you David, for one of the questions you suggested, which was, "What do you think, biggest disappointments or setbacks in 2022 were around AI?" Maybe you want to go first with that?

David: I'm disappointed by the level of public discussion on the remarkable breakthroughs that are happening. So, I sat through a couple of days ago, three hours of international conference on the future of AI. Lots of people there with job titles such as head of AI, AI Advisor, and most of what they were seeing, said ten years ago, at least 50%, maybe 80% of what they said could have been said in 2018; there wasn't a sufficient grasp of exactly how revolutionary the changes that are coming down the line from transformers, from large language models. One point that I didn't hear at all mentioned, which I think is revolutionary, is the way in which these systems can be trained once, in a very lengthy training process, that's the foundational training; and then can quickly be shown one example. And then they can acquire a whole new skill. Whereas previously, you had to retrain each model for every different task. Now they can pick up new tasks much more quickly. So, I see this as a transformation of what should be the correct understanding of AI. For a long time, we've had the idea that today we've got narrow AI that can only do one or two tasks and, in the future, we sometime in the future, we're going to have general AI, which has the same kind of general reasoning as humans, and there's a big gulf in between. Well, what we're now seeing with these foundation models, with Stable Diffusion and other things, is the growth with a what I call Broad AI. It's systems that can do many tasks, not just the ones that were pretrained for, like DeepMind's, Gato, which could do 604 different types of tasks in many different domains, but other foundation models that can learn whole

new tasks, much like the way that a child who's got a brain that was created by hundreds of millions of years of evolution, but that brain is now able to pick up lots of new tasks amazingly quickly. Well, it's assembled these foundation models, and this is truly astonishing and many people who have got AI in their job title, frankly, haven't got it yet, and they are holding the whole public understanding back by still being stuck in what is now an outdated and dangerously outdated view. So that's what disappoints me, but that's what I think will change in 2023. More people are going to get it and more people are going to wake up. And suddenly we're going to realize that a lot of what we thought would require some future AGI, general AI is already here, in what I'm now calling Broad AI.

Peter: I agree I have the feeling like I and you we've been saying for some time, "Look, you should expect this sort of disruption. That's just the curve we're on. We can't tell you exactly how, what, or when, but it's coming," and now it's here in a way that has grabbed people's attention, as though "Where did that come from?" Dan, where did you see AI either underperforming, not meeting the hopes that you'd had for it? Or otherwise, some sort of disappointment falling short in the last year?

Dan: I'll answer that. But before I've got to take an opportunity to ask David a question because on my podcast, I asked a lot of guests about their perspectives on AGI; specifically what it is, and best guesses about when we'll achieve it, referring to whatever they think it is and there's, as you might imagine, not even close to an agreement around a when or a what I you have much more expertise than most of the guests that I asked that question to what's your answer?

David: Well, I exclude consciousness from this consideration. Because I don't understand consciousness enough. I define AGI; systems that have as much general intelligence as much common sense as much flexibility of thinking as humans. So, they are indeed able to follow a conversation across all domains and be the master of every such discussion. When will we have systems of such complexity? Well, there is a whole community – communities - that are devoted to breaking down the definition of AGI into small chunks and estimating when each of these individual components will be here. I refer to the Metaculous community, which was set up a number of years back to aggregate in an intelligent way forecasts. And I think their current community forecasts for the arrival of what they call Weak AGI which they break down in having four different aspects which I won't go through now. But it includes comprehensively solving the Turing test; it involves being able to answer maths puzzles at the level of very high maths students just by looking at the images of the photographs, and so on. They now forecast sometime in 2028, I think as the median likelihood for that to arrive, and I don't think that's far wrong. I might push it out a bit, because I think there are still some puzzles to be solved. But I see so many people working on this, I see huge investments by China, at the state level at the regional level. I see investments all across the world and so I wouldn't be surprised if you have AGI by 2028. Though, if you want to be more like 50% sure, I might say about 2035.

Dan: Were you surprised at all by the reaction to Lamda and the conversation that spawned around sentience?

David: I think it was understandable. Humans have got a characteristic to impute consciousness to inanimate objects. That's why we imputed the actions to deities or hobgoblins or fairies in the past, it sort of made sense; there was a useful bias, if we thought there might be a tiger in the woods rustling it was better to be afraid of it than to discount the conscious that possibility. So, it is understandable and we have to beware anthropomorphizing our software too much. But at some stage that will come software that we tell it "Hey, you're not conscious," and it will fight back and it will give a strong argument as to why it has a rich consciousness and I don't know when that's going to happen; or could it be that the AGI that we create will tell us, "here's what consciousness is and you can see, I'm not conscious," it will say, "because I don't have that kind of architecture, I have a different architecture" and I await that with great interest. That's one of the intellectual tasks, I hope to understand better at the end of 2023, than at the beginning of 2023.

Dan: Peter, I promised to answer your question. I will now; and thanks, David, that's very, very interesting and I'd read could a little bit about your perspective on that very interesting to hear. The best way for me to give a retrospective on 2022 in terms of AI is by looking at two domains within AI, as kind of representative case studies. One is industrial robotics, and I think that's a case where the progress in 2022 was tangible. It led to some amazing innovations, supply chain efficiencies, beyond Instagrammable videos of Spot from Boston Dynamics, doing various kinds of calisthenics, which you know, are interesting and play well in the popular press. Beyond that, the innovation that I've seen for people with prosthetic limbs or the ability exoskeletons to you know, help human laborers lift heavy boxes, or do various kinds of tasks that improve supply chain efficiency, mobility, human mobility, that sort of thing. I think that's been a big win. So, I want to lead with the positive. I think, in some ways, it's easy to overlook some of the AI driven innovation that we've seen in the past year. And then I think that the case study, that is the poster child for disappointments in AI in the past year is self-driving cars. I think we do a lot of navels gazing in Silicon Valley, guilty as charged. But we've perpetually been at least for the last decade, one year away from fully autonomous vehicles, driving up and down 101 and I'd say by any measure, whether it's safety or reliability or accuracy, for a variety of reasons, every company that has tried to achieve that has either died trying or is in the process of dying trying. Tesla is notorious for having delayed or ratcheted back expectations for what its fully self-driving autonomous driving mode, can achieve. We've seen reports of nearly 400 crashes in the last eleven months alone, of self-driving vehicles from a variety of companies, this is not to place blame on any individual company. I think the only thing that is to blame is the hype that the global AI investment community is guilty of kind of feeding notions of how quickly we can expect the technology. By the numbers, there's been over $100 billion invested in self-driving automobile technology, organizations that have gone public, mostly the SBAC, or via direct IPO, direct listing, have lost over $40 billion in market cap since going public. Now, there's a general malaise in the markets that can, you can attribute a lot of that too, but by any stretch, or by any metrics, whether it's financial or the maturity, that technology, or certainly the performance of the safety of the technology in the marketplace. It's pretty clear to me - I'd love to hear your

opinions if you have different data, but we're at least at least a decade away from where I'd feel safe putting my kids in a self-driving vehicle and having commute to school work commute on 101. Perhaps my grandparents, you know, driving a golf cart in a protected community, you know, with known paths, that may be different a self-driving, wheelchair or something where it's AI assisted, yeah, I think that's those are safer use cases where we may, you know, be a small number of single digit years away, but certainly the hype and expectations around self-driving vehicles replacing driven vehicles by humans. I think that's an example of where we continually miss expectations and certainly that's the case in 2022.

Peter: Yes, absolutely agree, I had thought it possible that Tesla alone could have precipitated a new AI winter. But the developments in other fields have forestalled that. But while I don't think that level five, or level four, autonomy for consumer vehicles is remotely close, and this is also based on personal experience with the full self-drive beta in my own Tesla; what I have been expecting and was disappointed in, is the lack of permits, experience, regulation, or general advancements in long haul trucking, because I think that it's perfectly capable now of driving safely along North American interstates. And I don't know why that hasn't gotten further, I would have predicted by now that it would have. So just looking at our timeline here, and propelling this conversation along just to pick up on the thread about consciousness, I once loosely labeled a development in AI in the future to be "Conscious AI." I didn't mind co-opting the term 'conscious,' because to me, I don't know what it is. I'm not sure anyone does, and it's entirely possible to me that it doesn't exist at all. But when asked to define that I said, when we get to the point of AI that you program by arguing with it more than coding, I think that might be the point at which that is a useful definition. And some of the interactions with ChatGPT that I have seen are definitely along those lines. I'm not saying we've got to conscious AI by my definition or any other yet, but it's really interesting seeing how that's happened. So, to make the best of our time here, let's move on now to what would you like to put a stake in the ground; predicting or aspiring for 2023 in development of AI. David?

David: Well, I think there will be a change in the public mood; that the public is going to pay more attention to this. Currently, it's a little bit of a curiosity, it's a bit of a thrill. I think we may see more and more of the public waking up to the fact that these technologies are having big impacts on real lives. And also waking up to the fact that the regulators are potentially going to make things worse, as well as potentially making things better. So I think there's going to be a big discussion in Europe, and probably elsewhere. I say Europe first, because the EU is introducing an AI Act, which is well motivated. It's motivated to prevent harm. And it's right, we should be aware of potential harm if people are harmed by medical devices, driven by faulty AI, if they are excluded from getting a job because of an algorithm that is unfairly prejudiced against people with their types of surnames, then that should be addressed. But the mechanisms that are being proposed in the bill are said by some critics to be too heavy handed to be practical. Talking about avoiding all bias, being entirely fair: well, mathematically, there's always some kind of bias and there's always some kind of unfairness if you have a complex system. So, I think the public's going to wake up and there's going to be a lot of quite heated

discussion, to what extent do these regulations hold back innovation, and risk not getting the benefits of AI. And to what extent are they legitimate because AI algorithms can do a heck of a lot of damage as people who have been driven to suicide by despair as to the messages they read on social media. And even worse; there's been genocides, arguably, caused by irresponsible lack of monitoring of what was spent around on visible social media, such as Facebook, but also in the dark social media of WhatsApp back channels. So, there *is* a requirement to address this and initial attempts to address it may need a heck of a lot more discussion. So that's what I want to see and I think will take place in 2023, independent of the technical advances, which I think there will be many. I think there's going to be a much more vigorous and hopefully much better-informed public discussion of which podcasts like ours have a key role to play.

Peter: Dan, your answer here.

Dan: Yeah, David, I'm going to build on what you said, I applaud the EU, the leadership and EU community for the regulatory frameworks that have been imposed and some of the actions that have been taken, I think they stand in stark contrast to where we are in the United States with things like the recently introduced blueprint for an AI Bill of Rights. I think it's what you can expect from Washington, DC these days: it sounds good, but it's absolutely toothless, and at the pace that Washington DC manages important topics like this, it'll be essentially giving license to rogue operators who are introducing AI into the marketplace for another ten years of potentially irresponsible behavior. If there are if there are two things that I would like to see happen in 2023, they'd be these. One: I think every developer who's working on any technology that involves the use of automated decision making, and I'm especially using that term, not AI, because sometimes that's a that's an ambiguous term, but automated decision making, specifically where it affects lives, it's going to make a decision about your credit worthiness or whether or not you can be let out of jail, or whether or not you can have custody of your kids. Whether or not you can get your insurer to take a certain medication. This is real, these are decisions that software is being used to make using algorithms that humans write. I'd like to see every developer developing software that makes automated decisions take at least one ethics course. As part of getting your CS degree, it should be required in this day and age. And then two: I'd like to see us rapidly and by rapidly I mean, probably not introduced by but by our US federal government, some other body and for those listeners, if better-informed you want to help me accelerate a path to something like this, I'm willing to be part of the solution and certainly not stand on the outside and throw stones from a glass house. But a framework whereby every vendor, that introducing any technology that involves automated decision making, that affects consumer lives, I'd like to see that get some kind of a hygiene score. I've talked about this on my podcast before, but I think, when you go into a restaurant, it's very routine, that they'll get scored, the hygiene of the kitchen will get scored and I don't knowabout you, but I don't particularly want to eat in a restaurant that has anything less than an A hygiene grade. Because somebody had to audit the hygiene of that restaurant. And I think if we do that for to regulate the safety of a meal, I absolutely think that before any piece of

software, is going to tell me whether or not I can I get a loan or whether or not my kids are eligible to take a certain course in school, I think we should apply the same principles and certainly the same rigor and introduce some kind of a framework for scoring the hygiene of any product that's making automated decisions. So, Peter, that'd be my vision, I think those are two very pragmatic thoughts about where we might be able to go; perhaps ambitious, but like I said, I think when smart people have good ideas and a passion around changing the world in a way that could influence billions of lives, I think we can certainly accelerate how quickly we can make some of these solutions available.

Peter: So here's my take. Now, I'm going to go in some different directions and I'm not going to touch self-driving vehicles this time because I've been burned too many times before. I'm going to try and go out on a limb in a few places. I'm going to assert first of all that there *should* be - and hopefully *will* be - a revolution then microbiology, because of what AlphaFold did and it took the number of known protein structures from 190,000 to 200 million - in other words, *all* of them - and I just don't think it's possible to increase the knowledge space of some science by a factor of 1,000 without creating new industries, and I don't know why it hasn't happened yet, but I am not part of the microbiology community, so, maybe it's going on quietly. Microbiologists perhaps not known for being rowdy. nd I would expect that there will be whole new breakthroughs and industries created around that knowledge. I would, I can't wait to see what GPT4 is having experienced GPT 3.5 in the form of ChatGPT and I would say if the Loebner Prize was still running - and it appears to have stopped without any fanfare in 2019 - that the text version of it would be won easily next year. But I think there will be some form of Turing Test competition created, and that something based on ChatGPT or similar, will pass it and that there will be the sort of hype that draws a lot of attention and then the questions about sentience, again, raised a lot of heat, not so much light, a lot of smoke, not too much fire, but it will draw attention. And finally I think that right now there must be a cacophony of knocks on the door of OpenAI from large businesses with lots of data, anywhere from Procter and Gamble, Walmart, Goldman Sachs, you name it; saying, "We want to get a version of your AI that is trained on our intranet, data lakes and knowledge bases, and see what it can do as an internal resource." Because it could be clearly game-changing in the way that it could act as an adjunct, like an extra brain to their employees. I think anyone who's not thinking that, and has that sort of data available, needs to wake up and smell the coffee. That's clearly something that could be done, and if OpenAI is interested in making appliances for organizations like that, they could sell them as fast as they could make them. For just about any price too. So, that's my prediction. I think we're just about out of time here. So, a few words of conclusion, from both of you? David?

David: Well, I'd like to build on your thoughts about breakthroughs in science. That could be also breakthroughs with nuclear fusion. We've seen in recent very impressive reports that nuclear fusion, in some cases, is now generating more energy than was put into it. A few months earlier, a team from Deep Mind had reported that they could use AI to improve the design and operation of another fusion plant. So if AI is part of the solution to sustainable,

clean, abundant nuclear fusion, that's going to be an enormous breakthrough. I also think it's maybe about 10% likely that at some stage in 2023, there's going to be another scientific breakthrough, and I don't know what, but it might be that it's going to spit out the solution to superstrings, or the unification of quantum and relativity, it will have absorbed all the concepts and mix it up in its own vast mind and come up with some crazy proposal that two people are going to say, gosh you know what, this is a true super intelligence because it has solved this outstanding physics problem that has been in place - I don't know - since the time of Einstein, the unification of quantum mechanics and general relativity, which so many generations of physicists have failed to crack, and perhaps when AI does that, we're going to stop hearing complaints that it's not really intelligent. It's just statistics.

Peter: Dan, few final words.

Dan: Yeah, well said David. So, in my podcast we talk a lot about AI and the intersection of AI and the future of work. And so two perspectives on how AI is influencing or defining the future of work One, gentleman and Gary Bolles, who is the chair of the future of work at Singularity University. David, great person for you to meet if you haven't met Gary yet, but he talks about the evolution of a work net, from what we now call a workforce, where careers are essentially stitched together by a combination of tasks, or capabilities, or passion projects, as opposed to entering an office, kind of a grind of the nine to five where you work for one employer your whole life. There are things like skills matching technologies and various innovations that are making it easier for employees to be kind of free agents and stitch together careers based on the things that they're most passionate about. And from the perspective of how we use technology to make humans better, I really believe that we're just on the cusp of this new era, when we spend 100% of our productive working time when we're earning a wage, doing things that we love. And that's vastly different than today where most employees report that they're in some way disengaged or demotivated or feel like their values aren't aligned with their employers; I think that's going to go away. The second thing I firmly believe is that AI and AI-related automation technologies are going to give every employee back about an hour a week. And I think it's incumbent for us as leaders, think about how that changes whole civilizations, whole societies; the future of what it means to be human. This kind of magical fusion of the best of what humans can do with the best of what machines can do, is truly going to change the nature of how we think about why we're on this planet. Even an as little as an hour a week back to be better spouses, friends, caregivers, pursue a hobby, get to the kid's soccer game - football game for those international listeners. That in and of itself frees us up to feel like we're our best selves at work. When we also feel like there's time available to be at the best versions of humans or at the best versions of ourselves that we can be after work. So, when we talk about some of the potential risks, potential safety implications of automated decisions, those are real, and I think we need to take very seriously the ethical implications of how these AI technologies are used. I don't want to miss an opportunity to say how enthusiastic I am about the role of technology in improving life and I truly believe that the next billion employees, the whole next generation, whole next civilization around the world, will be made better when we get these

technologies right and I'm going to get on every soapbox every time I have an opportunity and try to remind as many people as possible that that's the future of humanity, that that I think we're on the cusp of experiencing.

Peter: Really well said. Well, there we've got fascinating, useful, monetizable, insights from experts in the field and two people that I'm proud and grateful to have had the privilege of interacting with and getting to know more over the last year: David Wood and Dan Turchin, thank you for coming on AI and you.

David: Thanks for making this happen. Peter. It's been a great idea.

Dan: Pleasure speaking to both of you, thanks.

That's the end of the discussion, which I could have happily gone at for hours longer. I felt so many ideas come on during that brief period.

Of course we had a lot about ChatGPT there, since not much more than a week before the recording date it took the Internet by storm and has done some things even I find frankly amazing and outside my ability to explain to my satisfaction. So ChatGPT is going to be this week's news ripped from the headlines about AI, and boy, is it making headlines. It's an understatement to say that we could do a whole show about it; we could make an entire podcast series out of it. I've had many interactions with it that have broken my previous expectations for what a large language model could do and it would take hours to deconstruct them all. For instance, I asked it to, "Explain Kant's Critique of Pure Reason," and it gave me a nice academic screed of the sort that has got people declaring the death of the term paper as an instrument of teaching. Then I said, "Explain Kant's Critique of Pure Reason to a five-year-old," and it gave this response:

> Imagine that you know a lot of things because you have learned them from other people or from books. But sometimes, you also just know things without being told, like how two plus two equals four. Kant's Critique of Pure Reason is a book that talks about how we know things and why we can't always know everything. Kant says that our minds are like little boxes that help us understand the world, but sometimes our boxes can't hold all the things we want to know. But even though we can't know everything, our minds are still really good at figuring out a lot of stuff!

That wasn't even the most dramatic demonstration of how ChatGPT is breaking expectations, but it's a fairly relatable one. Obviously there's a lot more to come in this area.

Next week, my guest will be futurist Tony Czarnecki, a returning guest who has a new article called "How Might Transhumans Control Superintelligence?" which really calls for discussion on what he means by transhumans, superintelligence, and control, and that's just what we'll do, next week on *AI and You*.

Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

http://aiandyou.net