**Karen Hao:** 0:00

What does this actually achieve? I guess, like the narrative that OpenAI tried to say was, by doing this, we will be able to continue developing AI for the betterment of humanity. That's why it's called OpenAI, and I think that's fiction. Microsoft has invested tens of billions I mean on paper, like what's been announced is they've invested over 12 billion in OpenAI. Microsoft's fate is tied to this company significantly. I suspect that this is not the end of the drama. I don't actually think that this resolution is going to be okay, great, like everything is back to normal.

**Craig Smith:** 0:39

This episode is sponsored by ISS, a leading global provider of video intelligence and data awareness solutions. Founded in 1996 and headquartered in Woodbridge, New Jersey, ISS offers a robust portfolio of AI-powered, high trust video analytics for streamlining security, safety and business operations within a wide range of vertical markets. So what do you want to know about your environment? To learn more about ISS' video intelligence solutions, visit issvscom. That's issvs.com they support us, so let's support them. Hi, I'm Craig Smith and this is my AI. In this episode, I talked to Karen Hao, a fellow journalist who is now a contributing writer at the Atlantic. She was previously writing for the Wall Street Journal in Hong Kong and before that for MIT Technology Review. She's the journalist with probably the best insight into open AI and we talked about the events of the past week. Karen knows the players, knows open AI's history and has some unique insights into what happened and what we're likely to see in the future. Beyond open AI, we talked about the likelihood of artificial general intelligence happening in the near future, as well as the existential risk that many AI researchers are concerned about. I hope you find the conversation as fascinating as I did.

**Karen Hao:** 2:28

My name is Karen Hao. I am currently a contributing writer to the Atlantic and also writing a book about the AI industry through the lens of open AI's rise and impacts around the world. I became a tech journalist very much by accident. I had studied engineering in college and then I worked in Silicon Valley at a startup and in the first year of working there I saw very rapidly the ills of Silicon Valley, if you will. The startup I was working for, the CEO, was fired, so actually very relevant to today's weekend's events. The CEO was fired by the board and I became very disenchanted with the progression of events that followed around both leading up to and after the firing of the CEO. So I was looking for other opportunities and I wasn't really convinced that I would be able to find something different within the Valley. So I had this vague inkling at the time that I would enjoy journalism, in part because I had always enjoyed writing and I was particularly interested at the time in climate change and how do you incentivize mass groups of people to change their minds and change their behaviours and originally I thought tech was the means to do that. Then I started thinking maybe actually journalism is the means to do that, like you need to really build public opinion and public support for the science around something before people are going to act on it. So I started working as an environmental reporter for my first job, but when I was looking for it was an internship and when I was looking for full-time opportunities it was very difficult for me to get hired as an environmental reporter because I didn't really have that kind of background. But consistently I was asked would you be on our tech desk instead, because of the background that I had, and so I ended up becoming a tech reporter and then I became an AI reporter, also not because I chose it, but because it was a job that was available. And then it became sort of the perfect match for me because I was really really fascinated by AI technologies and I actually had a lot of friends from college that had gone into AI research, sort of on the other side. So I was able to kind of quickly embed myself in the community and I realised that it was this microcosm of exploring all of the narratives that we have about technology the promise of it, the power, the potential, the societal impact, the sort of moneyed interests that are involved, the egos that are involved, and so I ended up reporting on that for now more than five years.

**Craig Smith:** 5:39

Yeah, and you were at MIT Tech Review, correct?

**Karen Hao:** 5:43

I was at MIT Technology Review and then I went to the Wall Street Journal and then I joined contributing writer at the Atlantic and while I was at MIT Tech Review I guess this gets to your other question of when did I embed in open AI While I was at MIT Tech Review, our focus was really on trying to cover, like, the bleeding edge of AI research. So, whereas you know, the journal takes a very different stance, it's like the journal covered technologies that are starting to be commercialised, that are starting to have some kind of business potential. Mit Tech Review was like if it had business potential, it was ready too late, right? So it was always about what we try to call the trends before they happen? And because of that, we started covering open AI very quickly after they were founded. They were founded at the end of 2015, and we probably started covering their research in 2017, because that's when they started producing some stuff that was starting to push the boundaries. And then in 2019, I can't really remember like how this came up, but I basically had a discussion with my editor at the time where I was like I think Open AI is sort of just a really interesting lab and there hasn't really been that much coverage of it Like we've covered its research but we haven't really covered its people and they were starting to become just prominent enough within the tech world that it felt like it was a worthwhile thing to do. And my editor at the time said you should just profile them. And so I reached out to the company. They already knew me pretty well because I had been covering their research and I said, hey, you've never had a profile done before. I think I would be the best person to do it. I think MIT Tech Review would be the best publication. And let me come to the office for three days and sit in on some meetings, chat with researchers, chat with executives. So that's what I did. I ended up flying to San Francisco from Boston and then I spent the three days there, and this was at the end of 2019. So this was a really, really interesting period of time in the company's history, because I went there a month after Microsoft invested a billion dollars into the research lab. So in that year of 2019, the GPT-2 announcement happened, which GPT-2 was. As listeners may remember, it was a few generations before chat, gpt and initially, openai took the stance of not releasing the model, but announcing to the world that they had developed it. So that happened at the start of the year and then it was a really big controversial decision because people thought like why would you announce it but then not release it? That's really odd. And then the capped profit arm was created within this nonprofit entity. Sam Altman joined as a CEO and then the billion dollar investment happened. So it was a rapid succession of changes that made clear that the company which was a nonprofit was quickly evolving into sort of a company and that it was sort of positioning itself to become bigger and bigger and more influential.

**Craig Smith:** 9:14

Yeah, yeah. And that GPT-2 debacle is how I regarded it was sort of the first glimpse of what was to come, because they said they developed it but it was too dangerous to release and it was so everyone. Of course that got everybody excited like what the hell is this thing? And then there were limited people, who were invited to review it, the creation of I mean, in your article, in the Atlantic article, and it's really what I wanted to talk to you about. You have a line and I think it's what a lot of people are thinking and concerned about and I'm gonna try and find it here if I can find where I opened it. I don't have it open, but you have a line about how this most important and powerful technology mankind has ever developed is controlled by half a dozen people who are fighting among themselves. And that's kind of frightening. And the whole sort of fiction of the nonprofit with a profit, even if it's a cap profit, subsidiary. I wanted to ask if that is fiction as well? I mean, it's all the same people. It's like you and I have a nonprofit, oh, and you and I also have a profit arm. It's not like we put on one hat and we're a nonprofit, we take it off and we're a profit seeking company. So is there anything real? Is that just a fig leaf from your point of view? And what do you think about this kind of technology being in the hands of such a few people, and people who evidently can't necessarily agree?

**Karen Hao:** 11:35

So I think the nonprofit for profit arm, I mean it's interesting. The people that designed that structure were Sam Altman, Greg Brockman and Ilya Satskever, who ended up becoming the main characters of the weekend. I can't personally speak for what Sam believed when he designed this because I never spoke to him about it, but I spoke with Greg extensively about it during the time that I was embedded within the company and he genuinely thought that this was the solution to raise. They were trying to solve a problem. They realised that AI development, the type of AI development that they wanted to pursue, would be very, very expensive and they needed to raise more money than a nonprofit could help them raise. And they tried to. I remember Greg said this to me during that time that I was embedded. He was like we did actually try because, like this notion of having this nonprofit was very, very like near and dear to us. So we didn't wanna just immediately go like let's scrap it and move for a for-profit. So this like at least for Brockman he genuinely thought that this was like a really clever solution that they'd come up with to solve this problem of needing the money but also staying a nonprofit. But the thing is. It's like what is, what does this actually achieve? I guess, like, maybe the fiction is that this, the narrative, that opening I tried to say was by doing this, we will be able to continue developing AI for the betterment of humanity and with the participation of humanity. This was a really big part of their early days. Messaging as well was like they were going to be open, they were going to be transparent. That's why it's called open AI. I think that's fiction. The non-profit for-profit is solving the specific problem that they wanted to solve. I think that is, you know, like they were genuinely trying to solve this very particular problem. But does it actually get us more open, more transparent, participatory AI development? No, not at all. What it actually does is just entrenched the power of the people that designed this thing. And ironically, I mean what we saw this weekend was that the non-profit for-profit did end up working as designed, in that the board did in fact, do what their job was to vote out Sam for like not aligning with the mission, supposedly. But then the reaction that we see from Sam, from Greg and ultimately from Ilya, when Ilya flipped, suggests that they're not actually here for this mechanism to be used against them, right. But if the mechanism were designed with like true sincerity of maybe one day I'd actually ask Greg, I'd ask like would you ever consider firing yourself if you felt that you were no longer up for the job, which I could have even phrased it as, would you be open to the board firing you or the board firing the CEO if they evaluate? And at the time he said like I would be open to it, but clearly like they weren't actually open to it, right. So that's the fiction that I think kind of became very plainly displayed this week?

**Craig Smith:** 15:26

Yeah, and the board is the board of the nonprofit, is that right?

**Karen Hao:** 15:31

The board? Yes, the board is part of the nonprofit and the nonprofit governs the capped profit, and that's why the board was able to exercise the power that had been bestowed upon them with this legal structure to vote out Sam.

**Craig Smith:** 15:48

Yeah, it'll be interesting to see if that nonprofit profit structure survives, because it doesn't seem to make a lot of sense. I mean, going back to GPT2, that was the thing that upset a lot of people Open AI. Open AI was supposed to be kind of an answer to big tech, to Google specifically, and that they are going to be open source, they're going to share all their research. It's not going to be controlled by a for profit entity, and maybe that was just naive, that when anyone develops anything that has such profit potential, they're not. The logic is that they need to raise funds and investors need to have some profit participation, otherwise they won't invest. I mean, ultimately, this kind of technology is not going to survive under a nonprofit umbrella, I think. But more specifically, I talk and I know you do too to Yan Likun, who I have enormous respect for. I mean, obviously he's a genius, but I mean in terms of his opinions on things like open source versus proprietary research, and do you think that this kind of tech should be open source, regardless of the dangers of open sourcing? Incredibly powerful technology, but simply to avoid this sort of thing, that then you have the broader research community working on it, refining it and then some sort of a licence structure that allows people to use it, whether it's for research or commercial use.

**Karen Hao:** 18:07

Yeah, I think it's a great question. To be honest, I haven't fully made up my mind about whether to fully open source technologies like these, but certainly we need more transparency than we need now. I think that is very, very clear. And also, what's interesting, I will say that Yann has been a big advocate of open source or of transparency, but Meta's Llama 2 model does not actually technically fit the definition of open source. They open sourced the model weights but by definition, they would also need to open source the data in order for people to audit it, to understand how it works. And Meta has refused to release any information about the data that it was trained on, and this is something that I think could easily become like a very low stakes accountability measure. Is releasing the data? Just saying what's in the data already is a huge step forward and you haven't trained the model, like the data is not the model. So if you're worried, if we were to buy into the idea that open sourcing the model could have dangerous potential, open sourcing the data would not. But the fact that we don't have any understanding whatsoever of what is being used to train these models, I think is a very telling sign of why actually these companies that say that they can't open source the technologies was the true motivation behind their arguments.

**Craig Smith:** 19:43

And do you think that's because they're afraid of liability? Absolutely yeah.

**Karen Hao:** 19:49

Absolutely. I think they're afraid of liability, of reputation damage, because a lot of the content that is put into these systems is not actually vetted very well, and that's precisely why open sourcing would create safer systems, because if you force companies to open source, they would have to significantly do more work to clean up the data sets, which would actually result in better products. And if you have many more scientists within the community, many more other people within the community going through like more eyes on these data sets, they will naturally just become better. And I think it would also be a forcing function to then get to a place where, like some of the companies are now doing this, where they're striking data deals, where they actually purchase the data from a media company or from Shutterstock or whatever it is. That came very late in the stage of the AI development that we're in. All of the original models were not developed with these data deals, and now the companies can continue to profit off of the data that wasn't paid for and that wasn't. They weren't being transparent about it. But if we had more of this transparency, it would be a forcing function to accelerate this trend, which I think is a really good one. There should be payments for data and potentially even dividend payments to the data providers, data creators, over time.

**Craig Smith:** 21:27

Yeah, yeah, that's. That's an interesting idea that you know. I know you know the same people. I knew that Don Song at Berkeley has been working on this idea of, you know, using the blockchain to secure your data and then Be able to sell it and have kind of this lifelong income stream coming from it, which sounds great to me. Now that I'm close to retirement, it would be nice to have an income stream off the data that every company Used of mine. Yeah, so when do you think I mean Sam and Greg? Or back at open AI? I, Ilya, feel bad for myself. You know I've interviewed such a Deep soul. You know he clearly didn't mean to cause this, this Global ruckus. Helen toner, I feel bad for her. She's been raked over the coals, people have made fun of her research and where do you think this is going to go? I, I, I can't imagine that. I think they're. They're what three people on the new board that again, with this technology, as critical as it is, I would assume at the very least, Microsoft will have a seat on the board. Where do you think it's? It's going to go in open AI's case and then we can talk about sort of government regulation. I would think that Regulators would be looking at this and saying you know, we can't have a bunch of you know 30-somethings in Silicon Valley like. Yeah, yielding the future of the world.

**Karen Hao:** 23:30

So yeah, I Think for open AI's case, I suspect that this is not the end of the drama. I don't actually think this resolution is going to be okay, great, like everything is back to normal, Sam's installed, all happy and peaceful and all you know. Like the piece that I wrote in the Atlantic talks about how there's all these different factions with the company, different ideologies. They all are kind of in this power struggle and I really do think that the more powerful a technology is, the more you end up with a game of throne style power struggle, because people think, very, believe very strongly in their ideology for how AI should be developed and it's both like this belief that's like a true, genuine belief and also, of course, like there is elements of desire for power, desire for control, and We've seen open AI go through different waves of drama before. So this is just the third way. If you, I guess you could call it like the Elon Musk leaving. Open AI was the first wave and then the anthropic open AI split was the second wave and now this is the third wave. There's definitely something else that's gonna come, but in terms of what this means, I guess for the course of AI development. I suspect that Sam is Certainly going to be a lot wiser about selecting carefully, he's choosing his board members and trying to make sure that he entrenched his power again. So if that is the case, then his specific ethos and his sort of habit around rapid commercialization, rapid growth, is going to now be like the main driving seat of the organisation and that is going to continue. We're going to see way more proliferation of products, way more downstream companies building on top of it and, unfortunately, I think we will see way more Ripple effects, negative ripple effects as well as speed overtakes. You know Certain types of trust and safety concerns, for example, and so I think that that's probably the most likely scenario, but also it's really hard to know. It's really hard to know because I don't know if you saw there was like that, that that letter that was circulating, open letter to the board from former employees. So it's like is that going to be yet another episode in this particular weekend saga, or Are we, is it buried for now and we don't see something else for another one to two years?

**Craig Smith:** 26:22

Yeah, yeah. And then you've got Microsoft and you know Satya's is public, they all smile and everything's fine. But you can imagine.

**Karen Hao:** 26:36

I've been sweating like crazy. Microsoft has invested Tons of billions I mean on paper, like what's been announced is they've invested like something like over 12 billion In opening up. But it's way more than that in the sense that when you look at their Investor, like their latest investor statement, they say that they're planning on laying down more than 50 billion dollars in new data centres for next year, and Not all of that is to our opening AI, but they're laying that down because they're selling to Azure their cloud compute customers. This idea of the Microsoft open AI partnership and this is why Microsoft stock has been doing so well as it is Bank. It banks on this partnership. So, like you know, when the news came and like, the Microsoft stock immediately started dropping and now that, like, things are back to normal, Microsoft stock is increasing. Microsoft's fate is tied to this company Significantly, significantly.

**Craig Smith:** 27:45

So, yeah, yeah, although a brilliant move, because at one point it looked and this is well you know all over Twitter and commentators who say but that he in effect had acquired open AI or was on the cusp of a AI, without having any, any regulatory interference or even having to pay a premium, actually paying a discount. So, yeah, adjust. In the data centres this is something I've Been talking to people about. This technology is is so much promise for enterprise, but because of the constraint in in available compute, which, which goes all the way back to, you know, silicone starts at the foundry, but but then through to to Nvidia and and their limited supply, you can't actually build and deploy a heavy use enterprise application Using GPT for through an API, just the, the pipe through which you're sending your tokens Is too narrow. Is this investment by Microsoft intended to ease that? What do you think about that? That constraint, how long will that last?

**Karen Hao:** 29:27

um, I definitely think that. Yeah, I do think that Microsoft's investments are meant to try and facilitate all of their customer base to transition to an AI forward business. I suppose, because I mean, every, every company that I've been talking to these days, regardless of what industry they're in, is suddenly Alert to the idea that they need some kind of AI strategy. And all of the tech giants Microsoft, google as AWS, amazon are all trying to capture that new Market and they're trying to build out their infrastructure to also facilitate that integration with these, these business customers and I Mean it's I personally, I I think there's sort of two interesting things that I'm personally watching for. One is how much of this talk is going to actually convert into implementation, because a lot of the companies I talk to say they need the AI strategy, but they're actually not sure what that means and whether or not it would ultimately be valuable for their business. It is valuable right now for their business to talk about it, but will it actually be valuable later to implement it? That's one thing to look out for, then, I guess the second thing is whether or not these companies, these Cloud providers that are jockeying for Market Share, are even able to acquire the resources necessary to continue laying down the data centres to keep up with this kind of demand. I think those are two things that could potentially end up limiting AI adoption or bottlenecking adoption, but it's difficult to tell right now what that will actually look like in five years time maybe.

**Craig Smith:** 31:28

Yeah. Have you heard anything that you haven't published or that you have published about this idea of Sam Altman starting a chip company or open AI? Starting a chip company to compete with Nvidia?

**Karen Hao:** 31:48

Only what's reported. My understanding is that this is not a new idea for him, that it was something that he'd always been interested in but had never, maybe not taken seriously before I'm not sure. But then it became much more real and viable, potentially and potentially a smart business decision to be able to actually have that. But the thing is I don't know that. People fully understand sometimes that it doesn't matter how many chip companies we have. We only have one real chip manufacturer, which is the SMC. You're not going to get, I mean, samsung as well and a couple other companies that are able to produce these chips, but TSMC is the most consistent provider and everyone wants to use them. No matter how many chip companies you have. There's still that bottleneck. I'm not really sure what Sam's plan was with that, whether he was trying to create his own chip company to get around the waiting list for Nvidia, or whether he was doing it for something else, like maybe for optimising, like trying to get to the next stage of AI development, maybe by trying to optimise how model training works by going to the hardware level. I'm not really sure.

**Craig Smith:** 33:20

Yeah yeah. A lot of people don't understand that when people talk about chip companies, they're not actually manufacturing the silicone chips, they're designing chips. I love the bit in the article about Ilya chanting the AGI. I'm a big AGI sceptic. I agree with Yanlacun again. I mean a lot of his ideas really resonate with me that the language models are not the way to AGI. I mean they'll certainly advance to something. As a journalist, a very well-informed journalist, what's your feeling about that? Do you think I get comments all the time on various things that I post AGI we're going to reach super intelligence sometime next year. It's like really, yeah, what's your sense of that?

**Karen Hao:** 34:38

My feeling is that we don't have any agreed definition of AGI. Agi could be here if you define it based on what we have, or it could be 100 years away if you define it totally differently. For the people who are saying super intelligence might be here soon, even scientifically, we don't have an agreed upon definition of intelligence. I'm not talking about AI. I'm talking about biology, psychology and neuroscience. There's no agreed upon definition of intelligence. I'm sure that the people that are saying these things totally agree with them. It's just like you get to define yourself what the goal is and where to go. I think this is the fundamental problem of the AI industry as a whole, as illustrated this weekend by OpenAI, is that by setting a goal towards something that is completely undefined, you just get to do whatever you want. I'd say that it's under the banner of a thing that sounds really nice and magical, even and de facto good, but ultimately the AGI is actually just a rhetorical tool to continue advancing towards whatever you want to advance towards.

**Craig Smith:** 36:05

Yeah, although I think we all have an idea of people who are paying attention to the space, of what it would look and feel like. I really like Yanlacoon's world model research because it's grounded in language. You layer on top of that. Ilia is a student of Jeff Hinton's. Jeff is now beating the existential risk gong. What do you think about that? Because, again, I lean more towards Yan and his view that it's certainly their risks, but this existential risk is a bridge too far.

**Karen Hao:** 37:09

So I've talked with Hinton about this. Actually, what actually changed his mind about this thing? Because he changed his mind. He's sort of relatively recently, and it was specifically that he realised that the definition again this goes back to definitions the definition that he was using for superintelligence before was potentially the wrong benchmark and that he should actually just be observing the ability of these technologies, that we have to engage with the real world and influence people and cause like real world phenomenon, and that we had already reached a point where it was causing like mass real world phenomenon it was like and large scale influence, you know and that, whereas humans are very lossy in our ability to transfer knowledge, that digital intelligence, as he was calling it, digital intelligence is not like you could have multiple models that immediately combine their knowledge. I'm saying all this in quotes because I think it's sort of important to emphasise that there are lots of debates around, like the use of these terminology, but that the, that digital models would be able to combine instantly, transfer knowledge instantly, and then that is how you would reach superintelligence. I am extremely sceptical of these claims as well. I think that Hinton believes what he believes and has a very logical path for what he believes. I also think that ultimately, like who likes me it's sort of like you would have to have it. We don't have very good techniques right now for developing advanced capabilities without massive data centres. So it doesn't to me it doesn't make sense that we should fear like 100 models suddenly combining into one. Who's training those 100 models? Like these models are exorbitantly expensive. Dario mode, a CEO of Anthropics, said publicly on stage earlier this year that currently the industry is training models that are around $100 million of cost. Then it's going to be a billion dollars of cost and he could see in two years there is a reaching $10 billion of cost. I mean there is not. Like are we going to train $110 billion models and then worry about them combining into superintelligence? I and like, where are we getting the data from for this? From, I think it immediately hits the real world limitations. But you know, like the scientists that have been working on these things for a long time, they have a very I think they sometimes have tunnel vision about the things that they research and they're not necessarily spending a lot of time out in the world like they're. They're like in their lab and like thinking about these things from a mathematical, theoretical perspective. And if you were to think about it from that perspective, then certainly I think you would start to get to some alarming conclusions. But yeah, that's sort of my view on it.

**Craig Smith:** 40:28

This episode is sponsored by ISS, a leading global provider of video intelligence and data awareness solutions. Founded in 1996 and headquartered in Woodbridge, New Jersey, it offers a robust portfolio of AI powered, high trust video analytics for streamlining security, safety and business operations within a wide range of vertical markets. So what do you want to know about your environment? To learn more about ISS video intelligence solutions, visit issvscom. That's issvs.com they support us, so let's support them. That's it for this episode. I want to thank Karen for her time. If you want to learn more about what we talked about today, you can find a transcript on our website IonAI, that's e-y-e-onai. And in the meantime, remember the singularity may not be near, but AI is changing your world, so pay attention.