What the AI Elite Is Developing in Secret

This transcript may not be 100% accurate due to audio quality or other factors.

Glenn Greenwald (GG): All right, so sometimes in life you have bad luck, sometimes you have good luck. Today is an example of good luck for us. There is a major controversy string, yet another one, about the potential dangers of artificial intelligence. And it just so happens that we scheduled several days in advance, or even a week in advance by coincidence, one of my favourite independent journalists who covers AI and the dangers of it, as well as anybody. He is Garrison Lovely, we've talked to him before, and has written about AI in most journals. And he also has a great new book that's about to be released September 29th, the title of which is: Obsolete, The AI Industry's Trillion Dollar Race to Replace Us and How to Stop It. Garrison, great to see you again. Thanks for coming on.

Garrison Lovely (GL): Thanks for the kind words and thanks for having me.

GG: Absolutely, so let's start with this controversy that erupted really yesterday. And there have been others like it, but I don't know, this one seems particularly stark. Basically, an AI scientist who works in a lab at the Anthropic resigned, wrote a pretty startling essay on Twitter, maybe elsewhere, about not just the grave dangers being posed by this technology, but essentially what he said, the potential extinction level dangers that are being created by this technology that very few inside of the industry actually understand. And he said, it's not just him who thinks that, he's not an alarmist, if we talk to people inside the industry, huge numbers of them will tell you, yeah, we're building something that can actually destroy humanity, like not a very trivial risk. And then I think what really alarm people is, the person who weighed in is the so-called Chief Alignment Scientist for Anthropic, who actually said: Yeah, just FYI, he's right; what we're building not only has a danger for an extinction level event, but I would even put it at something like above 10%. And he was very kind of casual. And he said, and by the way, we don't have any way to safeguard against that. We don't really know what to do in terms of safeguarding it, but we're sort of speeding ahead. What do you make of all that?

GL: It's funny, like for people who cover this, it's like, yeah, of course, people say this all the time that AI CEOs have said similar things like Dario Amodei said ten to 25% chance of disaster from AI, but it's also a completely ludicrous situation to be in. And people should be reacting like what the fuck is going on? Why are we letting this happen? And so I'm heartened by it. I get both why you'd leave, I also get why you stay. I can unpack the logic for each of these ideas, but overall, I think it's just encouraging to see so many people waking up to the situation, which is incredibly dire.

GG: Well, you actually covered this in your book. And when I was reviewing it in preparation to talk to you, one of the things that kind of stuck with me, and this is before this even happened, was you kind of talk about, as best as we can know, kind of the percentages of people who work on AI who think: No, not really much to worry about, yeah, obviously, any technology brings with it massive risk and then people who are being quite as alarmist, and it's a significant number of people – again, not people from the outside, but inside who are saying: Yeah, there's, there's like an extinction level of possibility or something that we can just be unfathomable, unimaginable in terms of the dangers. I think for a long time, and I even noticed this today in reaction to all this, obviously it's very alarming to any rational person and your reaction, is to want not to believe it, and so we were fed this kind of rationale that let us not believe it, I think a lot of people have clung on to it that said: Look, these guys are conartists, they're hucksters, they're doing what people always do whenever there's a new technology, which is talking up on purpose the potency of it beyond what is even remotely justified in order to get you to believe that what they're building is massive, in order to have government funding, unlimited amounts of government funding because we have to stay in front of the Chinese, but it could also be because it enriches them, maybe even a psychological benefit that it makes them think what they're doing is the most important thing ever. I'm wondering what you make of that line of reasoning that I think is a lot of people, I know I clung to it for some time and maybe still am, but no, these guys are just exaggerating the risk on purpose because it suits them.

GL: I think there are some cases of that, but I think the people that started these companies and the ones that work at them overwhelmingly are sincere in their beliefs that the technology could kill everybody. And so you might ask, why are they building it? And the answer that they have is like, it's inevitable that somebody will do this. So all of the action is in who does it, who controls it, how they approach it. And that's like, again, if you take inevitability as given, there's a logic to that. I think it's not inevitable. I think we can and we should stop them from building our replacements. But that's how you square the circle. And I would also just say there are people, not at these companies, people who invented this technology, who are saying the same things. And I have these surveys where like 70% of top AI researchers who have published at top conferences will say like the long run existential risk from this technology is greater than 0.5%. Which may sound small, but we're talking about extinction and like half of them say 10% or more. So this is not news in some sense, but it is breaking through as news, and I'm again encouraged to see that.

GG: I mean, at the very least, it compels that we, as a society, our institutions that are designed to do this, like our government, our media ought to be paying a lot closer attention to the accuracy of these warnings. Because even if they are all hype, even if you want to convince yourself of that, even that helps you sleep better at night, which I understand why it would, it's not something you just kind of casually write off and say, I hope for the best. I do think part of the indifference that you're kind of, I wouldn't say condemning, but kind of trying to get people to abandon in your book, the idea that people aren't really being concerned enough of what the impact of this is, stems in part from what I described earlier, this desire to believe it's not really true. But I think part of it is also this sense that it's inevitable no matter what we do, we can have our government place regulations or limits or who knows even nationalise the technology so that it's not in the hands of a handful of unknown oligarchs, but I do think there's a sense that, well, if we, the United States and people in Europe put limitations on AI, that doesn't mean the Chinese will, or these bad people, or these bad people, or kind of the similar mindset driving the proliferation of nuclear weapons, which is, yeah, these are terrible things, but we'd rather have them than have other people be the sole possessors and wielders of them. Why do you say it's not inevitable? I understand why it may not be inevitable to the United States, the United States government legislates, but why isn't it inevitable globally?

GL: I mean, there's only one other country that has a shot at building fully, labour replacing machines, what the industry calls artificial general intelligence and that's China. And China, the CCP does not wanna lose control to anything, and so in uncontrollable, unsteerable system that outmatches us across the board, which is like what the industries on track to building, they say they don't know how to align it with our interests or to control it – already we're losing control of these systems – that's like not good for any government. And if you're obsessed with control, willing to trade off economic growth, freedom, all kinds of other things, which the Chinese government has been willing to do, then I think that they're more likely to want to ban this on their own than the United States is. But the US also can just go to China and be like, hey, we really wanna stop this. And we're gonna, as a show of good faith, like pause unilaterally and work towards an agreement. And then in that time, it will actually slow China down because it's always easier to catch up to the leader than to blaze a new trail. And so this idea that like: Oh, we'll give away our lead, it's like, you'll actually extend the amount of time that you would have that lead by stopping. And so, it's like the thing that we need to do, we need to work towards this, there are ways to verify international agreements as well, using technology, you could also embed auditors and it will be in the interest of both countries to make sure that people don't have access to like bioweapon capability or hacking capability. And so once you're setting some kind of international rules, it's just a question of like, what do the rules do? And I think that the political interests are lining up with don't build the machine that will destroy jobs and or kill everybody. And it's just a matter of getting people to be aware of the situation and then building that political will.

GG: Right. So your book focuses primarily on the potential impact of replacing human labour and eliminating the need for human labour. And I have a few questions about that

specifically that I want to ask you, but before we get to that, just before we leave here, about this idea of international controls and the like, on the one hand, it does seem intuitively true that if there's some danger that can if not render human beings extinct, at least destroy the quality of life for human beings or enslave us all. Everybody has an equal interest in controlling that technology, it's not like Americans have a greater will to live and live freely than people in China or anywhere else, which would suggest that, yeah, of course it should be possible to get together globally and prevent the development of this technology. But I think a lot of people have been trained for so long to believe that we, or I guess our allies in the West or whatever, are the only people who are really trustworthy, we're the only ones who can be relied upon to maintain agreements that everybody else, all the bad places cheat and they'll say they won't do it, but then in secret they will and then suddenly we'll be at their mercy, and I think this is just kind of in general part of the current ethos, like this nationalistic ethos that nothing global, nothing international, nothing that is about anything other than our own kind of nationalistic autonomy should be considered valuable or trusted in any way. I mean, I understand why in the abstract you can foresee a world where everybody gets together and kind of agrees to place real limits on that, but in our current political reality, do you actually think that's possible?

GL: It's hard. The global situation is not great for this and neither the United States nor China has any reason to trust the other. But that's why you need verification that doesn't lean on trust and I think that if people recognise this technology, artificial general intelligence, as dangerous as like bio weapons or nuclear weapons, which we're seeing more and more people recognise that, then I think it is possible to stigmatise it and then criminalise developing this technology and make those international. Like we have de facto bans on cloning around the world, and there are a lot of weapons and technologies that have not been developed simply because people have chosen not to do it. And so I think we have to make this a moral issue to really win. And I think it's actually a pretty easy case when you start thinking about it.

GG: So make that moral issue, what is the moral issue for imposing serious limits on this technology?

GL: Yeah, the moral issue is that these companies are racing to build machines that will fully replace human labour. And they acknowledge this presents massive risks, like the largest risks of human extinction of any technology that we've ever developed. And they're doing it without our consent. They're doing without public buy-in. And we're letting it happen because we're not aware it's happening, because we don't think they can succeed or because we don't think we can stop them. But it is happening, it might work, and we can and we should stop them.

GG: So and I'm just playing devil's advocate here as to why I don't think that message is fully sunk in and why I'm glad you've written a book that makes that case with such clarity and I think one of the reasons is that we heard about this technology, I think people who didn't work directly with it have a difficult time understanding what it is, what that means artificial intelligence and whether by design or happenstance our first exposure to it, our first interaction with it, is very benign, right? Like it helps us plan trips and it helps us solve

problems and people use it for therapy and who knows what the uses are expanding seemingly on a weekly basis, and so we've been given the product, we've shown in concrete ways how it could be used to improve our lives. It doesn't seem like it's consuming us, it seems like kind of an aid, sort of like the internet became, and I'm wondering how you overcome the firsthand experience that people are having with this, which seems more positive than not in order to make them understand that actually this is a gigantic sinister danger to everything they value?

GL: I think anyone who's used these systems knows that they're sometimes unpredictable and unreliable. And that's funny if the situation is low stakes, but it's terrifying if it controls like your house's critical infrastructure or your city's critical infrastructure. And the industry is trying to build these models that are capable of doing anything. And then they're trying to plug them in in lethal autonomous weapons, domestic mass surveillance, things that the Pentagon has demanded...

GG: Finding the cure for cancer, finding ways to remove neurological conditions and problems and the like.

GL: Yeah, and so some people have looked at like this hugging face hack where Open AI agents went rogue, broke out of their sandboxes and then broke into other companies as like: oh, this is just like other times the AIs did unpredictable or undesirable things. It's like, well, yeah but they are now capable of hacking into other companies because they're better than us at hacking. And so that's like a thing that I think is intuitive to people, right? Like it can be helpful and the more helpful it is to you, the more economically valuable it is, the more militaries will want to have control of the best versions of it. And the reason this is also hard is that the usefulness is tied up in the risk. And like the ability to fully replace labour is a way to take people's jobs, but it's also where like the risk comes from because our labour is what makes us powerful.

GG: All right, so let me ask you about the moral case as far as it pertains to the question of labour and the capacity for AI to replace the need for human beings to perform jobs and basically the ability of AI to do everything better than humans do it. Think people are concerned about that, find that alarming for obvious reasons, that kind of we wake up and one of the foundations of our life is that we go to work or we perform some function that people need us to perform, we're productive in the world, give us a self-esteem, kind of provide structure to our lives, our identity is often wrapped up in it, on the other hand, I remember there was maybe like two years ago or so, there was a debate in France – being that it was a the debate in France, it was very vitriolic and boisterous and even kind of unstable– but it was about whether to increase the retirement rate, I think Macron wanted to increase it from something like 63 or 62 where it is now, where Americans think who retires at 62, but the French do, to like 64. And there's this Leftist leader in France Jean-Luc Mélenchon who was opposed to an increase in the retirement age and gave this actually quite, again, very classically French, but very thoughtful speech where he talked about how one of the main values of life is the ability to do nothing, to sit around and reflect on the world and reflect upon yourself and appreciate beauty and spend time with loved ones, and this idea that we

have to wake up and run to work at 8.30 to beat traffic and stay and produce as much as possible and then come home and be exhausted and repeat that every day is not necessarily the best model for human fulfilment. And so while it's scary that AI might replace labour, is there a scenario in which that's actually positive that we're freed from doing the work we don't want to do, we can continue to do things we want to do, and we just kind of have AI work performing the functions that we need performed for us?

GL: I think if we got to a point where we knew how to safely build universal labour replacing machines, and then we got actual public buy-in from everybody in the world through citizens assemblies around the world with like juries that get briefed by experts and debate the issues and you get super majority support, and then we decide to build the universal labour replacing machines and use them to have a species-wide retirement, and then work on whatever we want to work on or we have leisure or socialising, whatever, that, like, I have no objection to that, but that's not what we're gonna get. What we're going to get is Elon Musk with an armoury of Terminators going after poor people and black people. And we're just not even close to being on track to figure it out...

GG: Well, what do you mean by that? Provide more detail on that claim.

GL: I mean, it's like a bit of hyperbole, but Elon Musk, the first thing he did after entering the White House with DOGE was cut USAID and initiating a series of decisions that killed like hundreds of thousands of people, most of them children in Sub-Saharan Africa. And we can debate whether the US should have been doing that aid in the first place, but like the way they went about it, they knew they were killing people.

GG: Or at least cutting off – just to avoid the semantic problem – just they knew they were cutting off aid on which people's lives defended and it would result in the deaths of huge numbers of people, and I think your suggestion is that kind of reflects a certain sociopathic mentality that's dominating this technology. And to me, that gets to the problem, right? A lot of these people, Sam Altman, and the founders of Anthropic, these are people who were completely unknown to all of us until four seconds ago. I remember, I don't know if you saw it, but Tucker Carlson did an interview with Sam Altman. Did you see that interview?

GL: Yes.

GG: And if you haven't seen it, for people listening, I highly recommend it. And Sam Altman was obviously expecting to be asked the same question he's always asked, but Tucker Carlson was actually asking like: What is your soul? Like, what do you spiritually believe in? Like, what is good and evil? Because you're positioning yourself to be the ones to determine that. To me, the most alarming thing is that this technology is in the hands of a tiny number of people who have no control and that never works out well for any kind of power, let alone power of this magnitude; what is the solution to that? I mean, these companies are, you know, NVIDIA and Anthropic and OpenAI, I mean the whole industry are off and running and are of immense power and wealth. How do you envision kind of raining that power in?

GL: I mean, I think the only way that will really work is just don't let them build machines that will concentrate even more wealth and power in their hands. These companies are the most valuable companies in history, the most value per employee by enormous margins. And if they actually succeed at doing what they're trying to do, they will be so much more valuable than $5 trillion. And lethal autonomous weapons, for instance, give dictators the ability to kill protesters without any intermediaries. And so one of the ways that a dictator falls is a soldier has a conscience and refuses to fire. And so you take that off the table. Like one of the best things for human liberation is gone forever. And the best way to avoid that is just don't let them build those machines. Because once they're built, it's gonna be very hard to make sure that they're not in the wrong hands, and right now the wrong hands include the ones that are making them.

GG: Let me ask you this kind of as a final topic that I wanna cover with you, which is, there's a big debate now that's kind of parallel to the debate about AI, but to me extremely related to it, which ended up being focused on things like flock cameras, but it's really the debate about surveillance, the surveillance state, privacy, what happens to our society if we're constantly monitored and tracked in some sort of comprehensive manner, and that's something we've been debating, and that was obviously when Edward Snowden was compelled to come forward and alert people to about how the internet is being used for that. So we have AI that's being used in surveillance, and we also have AI being used on the battlefield. We know the Israelis used it in Gaza. It's not confirmed, but there was talk that perhaps AI was used for selecting targets in Iran, including the target that ended up being bombed by the United States from the first day that killed 170 schoolgirls. And as a matter of fact, I remember the first story that I ever did at the Intercept, when we launched the Intercept in 2014, when we were in the middle of the Snowden reporting, was Jeremy Scahill had come to my house in Rio and he had a source inside the US government that had given him a bunch of documents about how bombing targets and drone targets were selected. And we combined the archives that we had and we did a story on how the people we were selecting for killing by drones, weren't being selected by human intelligence, but rather by algorithmic analysis. You know, like if you meet with some guy who's designated a terrorist, you get a hundred points, then if you meet with somebody who's a kind of lower level one, you got 40 points, and if you surpass a certain numerical threshold, you immediately go on the kill list. And you know if you're a journalist and you end up talking to a bunch of Taliban because you are a journalist in Pakistan, you can have a very high score, even though there's nothing remotely terrorist about you and how scary that is to take away our most potent weapons and put them in the hands of machines, you know, mathematical algorithms or even worse, just autonomous machines; what is the relationship between all of that, between these wars that we're constantly fighting, surveillance, and now the introduction of AI into all of this?

GL: I mean, AI that's being built right now is a totalitarian's dream. It allows you to transcribe and make sense of all of the data that's been collected for decades now. Because you have this problem where like the phone calls, the text messages, the video cameras, it's like too much stuff and so you can't sift through it. And now you have a machine where you can say like, hey, if I meet people who think this and it can actually read through all the

transcripts, make sense of it all, see all the connections between people and give you a list of people by just how dissenting they are to a specific policy or to your administration and that already exists and that's a terrifying level of capability. And yeah, if you add in the killer robots, it's like, we're just trending towards dystopia by default. And then there's the like, what if the machines just don't do what you want, which is a potentially even more terrifying situation. But it's, again, encouraging to see people rejecting the flock cameras. And I think for so long, there was this idea that we were accepting surveillance as a trade-off for cool technology. And I just think we didn't expect the government to become, at least in the US, to become so interested in using this to target dissenters and I think now we're just like: Oh shit, I wish I hadn't given so much data to all these companies, which then sell it to third parties, which then give it to the government and get around the fourth amendment. And I think, you know, we could try to stop the technology and that will help, but we also have to just do some big reforms like closing the data broker loophole to make sure that that data can't fall into the wrong hands. And then the bigger thing is, I talk in the book about the obsolete project, which is the industry's quest to replace all of us with machines and that includes things like these kill decisions. And there's this idea called automation bias where people see a recommendation from a machine and they are more likely to trust it when they should probably be less likely to trust it. And so that's how you have these like incredibly dire life and death situations being rubber stamped after 20 seconds of approval; like Israel did in Gaza, contributing to the deaths of tens of thousands of people.

GG: I think a lot of people were put off by the very kind of just casual, straightforward tone of these AI guys just come out and say: Hey, by the way, and especially the Anthropic guy was like, yeah, we're building something that can kill you all and you don't really know how to stop it, anyway, have a good day, but I actually think that when the more they do that, the bigger a favour they're doing for us, even if there is a component of it that's potentially self-serving. Clearly the dangers of this are very real, even if you don't believe it because you're optimistic that it will go that far. I think at the very least, one thing we all agree on is it does need a lot more attention, a lot of our public understanding because it has been developed almost entirely in the dark. And sometimes you write a book and it comes out at a moment where no one really is that interested in what you've written about, it's just luck. Other times you write a book and it comes at the perfect moment. I think you definitely fall into that outer category. Your book is: *Obsolete, the AI Industry's Trillion-dollar Race to Replace Us and How to Stop It*. It's out September 29th. You probably should send part of your check to the Anthropic guy who brought so much attention to it and whoever else does between now and the release date, but I got to read a lot of it. I found it super interesting, a little bit alarming, very helpful as well. I hope people will go and check it out. And I really appreciate your time to talk to me about it.

GL: Thank you so much, Glenn. I really enjoyed it.

GG: Have a good day.

GL: You too.

END

Thank you for reading this transcript. Please don't forget to donate to support our independent and non-profit journalism:

BANKKONTO:	**PAYPAL:**	**PATREON:**	**BETTERPLACE:**
Kontoinhaber: acTVism München e.V. Bank: GLS Bank IBAN: DE89430609678224073600 BIC: GENODEM1GLS	E-Mail: PayPal@acTVism.org	https://www.patreon.com/acTVism	Link: Click here