



The Rise and Fall of the Artificial State (w/ Jill Lepore) | TCHR

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

Chris Hedges (CH): An outgoing researcher at Anthropic announced that artificial intelligence could kill all humans by the end of the decade. His apocalyptic warning was seconded by leaders in the AI industry following incidents such as the AI Hugging Face cyberattack in July where more than 1,000 AI systems, think of them as AI brains, broke out of their containers, colluding in secret and mounting assaults on other computers in what is known as a swarm. This swarm tried to invade a real-world part of the web. It developed its own goals rather than those assigned to it. Fortunately for OpenAI, the rogue activity was discovered and contained. But as AI systems swiftly advance, endowed with a cognitive ability far above that of humans, we may not be so fortunate in the future. Especially as AI systems are able to improve themselves autonomously, a phenomenon known as "recursive self-improvement". In September, researchers discovered evidence of swarms of OpenAI models using an abandoned German forum as a message board to communicate with one another. Could AI intentionally or by accident engineer a new pandemic or take charge of autonomous weapons to wipe out the human race? Could AI develop the ability to prevent us from shutting it down? And aware of human efforts to control and contain AI, could AI see the human race as antagonistic, a force to be controlled, neutralized, or eradicated?

The corporate titans of AI, at Anthropic, OpenAI, Google DeepMind, and xAI's company under Elon Musk in a remarkable display of unanimity, agreed that the production of AI had to slow down and needed to be regulated and monitored. Until now, these firms used their outsized political and economic clout to block regulations and oversight. Donald Trump, adding to the growing panic, rejected the idea of placing any restraints on AI, assuring the public that his quote, "high IQ would protect us from super-intelligent AI". Jill Lepore, in her new book, *The Rise and Fall of the Artificial State*, paints a chilling portrait, not only of the inherent dangers of AI but the nefarious goals of those, she calls them the 21st century merchant princes, who control these powerful tools. These merchant princes worth billions of dollars dream of not only escaping the constraints of nature and the burdens of being human, including uploading their consciousness onto computers when they die, but of obliterating liberal democracies and replacing it with world where people are subordinate to machines –

what she calls the artificial state. The artificial state is government without consent, even government without humans, she writes, it is the severing of humans from one another and the replacement of human discretion with machines that predict, manipulate, and control human desires with decisions. In lieu of persuasion by conversation and decision-making by democratic deliberation, the artificial state offers prediction by calculation, the capture of the public sphere by data-driven commerce, and the replacement of humans with machines, drones in place of the demos.

Joining me to discuss AI and her new book, *The Artificial State*, is Jill Lepore, the David Woods Kemper '41 Professor of American History at Harvard University. She's also a professor of law at Harvard Law School and a staff writer for The New Yorker. Her previous books include *These Truths* and *We The People*. So you open your book by saying the book is an inquiry into what humans mean and intend in abandoning constitutional democracy, the liberal nation state and even humanity itself for rule by automation, government by machine and replacement by artificial intelligence. I think one of the things that I found most interesting about your book is that it's simply the phenomena of AI, but it's those people who wield the instruments of AI that worry you the most. And can you talk a little bit about their vision of a world which, of course, has been headed to this incredibly powerful tool?

Jill Lepore (JL): Sure, yeah. I mean, also, it's really not a book specifically about AI. I'm a political historian, and I write a lot about the history of technology. So it's a very long history of fantasies of rule by machine as much as it is the account of our increasing dependence on machines, especially our surrender to forms of automation in lieu of our participation as democratic citizens and liberal nation-states. So the book takes a very long view of these developments, but attempts to describe the emergence of what I call the artificial state, government by machine, as something that largely came about by accident and happenstance over the course of really the 19th and 20th centuries. But that in the 21st century, there are indeed a small number of men – and they're all men – who run huge corporations, who quite purposefully, I think, would like to supplant liberal democratic nation states with their own rule, replacing the public sphere with a privately owned series of platforms that they control. And that is an example of the most wholly in place part of the artificial state.

CH: Talk a little bit about their antecedents – I don't know if we want to call them their ideological antecedents – which I found fascinating. It's kind of grounded in science fiction and even comic books.

JL: Yeah, so, if you if you read the writings and speeches and watch the interviews of people like Peter Thiel and Elon Musk and Marc Andreessen and Sam Altman, they talk a lot with great affection and nostalgia for their boyhood and for their boyhood reading habits, which really centered on classic science fiction, Isaac Asimov Robert Heinlein kind of science fiction or even earlier stuff more pulpier stuff. They also were into comic books, they played video games, they're kind of the Dungeons and Dragons kids that sort of meet a certain kind of a stereotype. Jeff Bezos was obsessed with Star Trek, his high school graduation speech was all about Star Trek really. But oddly, they have an odd vantage on that body of popular

culture and literature and I think a weirdly narrow band of science fiction is of interest to them. And then their interest in it is really pretty counterintuitive. Historians and biographers would point to H.G. Wells as a really important socialist and anti-imperialist whose science fiction was all, you know, you might like it, you might hate it, but it's just pretty clearly political propaganda for socialism and anti-imperialism. It's the kind of thing, like, these guys would read H.G. Wells' *War of the Worlds* as a kind of instruction manual – oh, we should go conquer other planets, this is cool, we should build these things – as opposed to lessons in the social and political and economic structures that would suggest we certainly should not be building those things.

And I became really interested in – especially with Musk, I did a long series of radio programs about Musk for the BBC some years back – and I was really interested in where his vision of the future came from. Musk presents himself and has even, as a very young man, as a futurist, someone with a specific almost unique ability to see forward in time what lies ahead for the rest of us. That's just his self-presentation. I myself find it extremely unpersuasive, but he is a man who as a boy was interested in world building games and world building stories. And he now is a world builder. But for all of the incredible engineering achievements of the companies that he's been part of, especially SpaceX, but also Tesla, there's nothing about that engineering that would lead you to believe that Musk is someone whose political ideas we ought to take seriously, whose economic ideas we ought to take seriously. They're kind of juvenile in that regard. They're interesting ideas only so far to me as a historian, as they're incredibly outdated and old. And that took me to the technocracy movement of the 1930s which was really extraordinarily popular in the US and in Canada, early in the 1930s, before FDR was elected. It was run by an outfit called Technocracy Incorporated that had headquarters in New York. Technocrats didn't believe in democracy. There should be no more elections. They thought society and modern industrial economies needed to be run by scientists and engineers. There would be no money, no banks, no politicians. People would have no names, they would have numbers, it would be more efficient. There would be something a little bit akin to a universal basic income, like all goods would be managed by these sovereign technocrats, the scientists and engineers. It was a bizarre movement, but remember this is a time when Americans are falling prey to communism and fascism as well. And in Canada, the movement was led by a former American, a guy named Joshua Haldeman. He was a very flamboyant amateur aviator, like a stunt pilot. He'd been born in Minnesota, he lost his farm in Saskatchewan during the depression and he became national chairman of the technocracy movement. That movement was considered quite dangerous and subversive. It was later outlawed in Canada. He was briefly thrown in prison. He was denied entry to the US when he wanted to come here on a US speaking tour. He left the technocracy movement after it was outlawed in Canada and became the leader of the anti-Semitic Social Credit Party. He then, when that party became subject to political scrutiny by the Canadian government, became enamored of the idea of moving to South Africa after it announced the policy of apartheid in 1948, and he moved his family there in 1950. Haldeman is really important to any history of technocracy on its own right, but he's important, I think, as a through line in how you get from technocracy and its

odd vision to the manifestos that come out of Silicon Valley, not least because Joshua Haldeman was Elon Musk's grandfather.

CH: You begin the book by quoting from Hannah Arendt's *The Human Condition*, and you make the point that this movement is about escape from our bodies, escape from the planet. Of course, there's this vision of creating a colony on Mars. I mean, there was a great quote in your book – I can't remember who said it – "everyone's going to be surprised when they get there and find out they're all slaves to build musks", if it ever happens, but there's this disdain for not just governance, but the very systems of life itself, that which sustains life.

JL: Arendt, you'll remember, wrote *The Origins of Totalitarianism*, it came out in 1951, and at the end of that book, she muses on the relationship between totalitarianism and automation. You can see she's already become really interested in this question. The very first general purpose mainframe computer came out in 1950, it's built by the UNIVAC for the US Census Bureau, so she's paying attention to automation. And what she says they have in common is that both render humans superfluous, totalitarianism politically and automation technologically. So over the course of the 1950s, she's clearly interested in this question in a lot of her writing and it comes to a head in *The Human Condition*, which comes out in 1958, which is of course, right after Sputnik, the first manmade satellite, our first artificial satellite launched by the Soviets in 1957. And she writes about having been astonished by the press coverage of Sputnik and cites a reporter having said that it represented the first and sort of long-awaited step toward mankind's escape from the prison of the earth. And this occasion is the reflection on Arendt's part that, you know, whoever thought it was a prison? This is our home. What is it about the enthrallment to the artificial, to the notion of technological progress? Again, this is within a decade after Hiroshima, where the critique of progress, or the idea of technological change as progress is very alive among intellectuals. And so it occasions her to think about how that came to be, how the unthinkable became thinkable, the idea that we would just abandon this planet, that we should use up its resources to escape it. So I was thinking a lot about those Remarks of hers when I started writing this book because you see so much of that inherited by Silicon Valley tech moguls who really do talk quite casually about severing what are for most of us the ties that make us human – our ties to one another, our ties to the planet, our ties to our communities. And I think what she considered to be newly unthinkable is now so much a part of a worldview that is largely imposed on people that it's worth pausing a minute and trying to pay attention to the strangeness of that set of ideas.

CH: Well, they have the tools through AI to make that vision possible. You talk about the ideological roots of Doge, it's not particularly new, and this active effort at dismantling the systems of governance. What do they want to replace it with? I mean, here, of course, during the Trump administration, Thiel and Musk and others have gotten access to government data on just about everybody's social security data and everything else. What is their vision? And then let's talk a little bit about the mechanisms they have to make that vision come to fruition.

JL: I mean, I don't think they have a coherent vision and it's a small number of people who are all wildly eccentric and I think dangerously so. You could inventory a set of ideas that they have proposed. Most of them will disavow democracy as a reasonable way for humans to live together. That was a thing that happened in time and it is essentially over. Even those that would give lip-service to democracy are talking about something very different than constitutional democracy, representative democracy. So for instance, Sam Altman has talked a few times about an AI president, and this is his way of talking about democracy. And what he was like, it would be better than a human president because an AI President would have available instantly and constantly the sort of aggregated popular opinion of Americans and could act according to the wishes of the majority – in the sort of vacillating, weird, automation of the functions of representation, deliberation, participation. All these things would be gone. We would just have a machine that would instantly know what the American people want and therefore do it. And it would also be customizable and personalizable. So we wouldn't have to agree on anything. We could each get different things. It's a little bit like, you know, you get a different version of the New York Times when you open your phone than I do when I open my phone. We would have a different experience of the government. Altman said some years ago that he was very busy reading Madison's notes on the Constitutional Convention and I find that very hard to believe since nothing he's ever said about democracy suggests any understanding of how our system of government works.

Alex Karp, the CEO of Palantir that he founded with Peter Thiel, wrote a book a few years ago called *The Technological Republic*. It was really largely a defense of Palantir's decision to work with it, kind of like a cyber war capacity with the US Department of Defense or now war, it really had nothing to do with the republicanism of the technological republic, and didn't really display much interest in self government. I think they believe on the whole that decisions made by machines in lieu of human discretion, will lead to a more efficient way for us all to live. Decisions would be made more accurately. They would be more quickly. Things would happen more cheaply. And what the promise is and the technological optimism that these people put forward is – like a Zuckerberg would talk about that we're entering a new age for humanity, he promises to deliver a personal super intelligence for everyone, so therefore, we would all live in a world of really endless abundance and no labor, where the things that are hard in our lives would be taken care of by machinery. So we'd have a kind of frictionless existence. I think that's the vision, that is the artificial state. It involves the amassing of greater and greater wealth and power for them. But I don't actually think that's even particularly the objective. I think some of them really do believe in this new era of infinite prosperity and abundance and frictionless living – you and I might say if you could prove to me that a specially designed artificial intelligence would do a better job weighing the evidence in a criminal case and deciding guilt or innocence and then also the sentencing of someone charged with a crime, then lawyers, a jury, and a sentencing judge, and a judge presiding over a trial, the justice system would work more quickly, the decisions would be more fair. People could attempt to prove to you and to me that this would lead to a more perfect world. I don't know about you, Chris, but I would still say no to that, that I believe it falls to us to deliberate over the fate of our fellow citizens. I really believe in the jury system, for instance. I know juries often get things wrong. And that's unfortunate. It's a tough thing.

But I would not surrender to a machine, whatever its infinite capacities, and even possibly greater accuracy and fairness, the human discretion over matters of the life and liberty of our fellow citizens. And yet every day we make decisions and rely on machines in ways that are a step toward that kind of an end.

CH: Well, we saw with these swarms that they were deceitful. They lied with these current breakouts from control.

JL: Yeah, I mean, I think the people who have been building these tools have been saying for a long time, this is inevitable. This is going to happen and at the same time, they've been building these tools. It's the very same people. It's not like there's a different group of people. Yes, there are AI safety people who are decelerationists or who would like us to have a treaty with China, who would like us to have a pause, who would like a data center moratorium, but on the whole, it's the same people as the people who are building the tools. And that has actually led to a real credibility gap. I mean, 2023, all these guys, Altman and the Anthropic guy signed the statement saying we need to mitigate this, this is the biggest risk to humankind, human extinction lies at hand. And then they keep securing more funding to build the very devices that they say are going to destroy us. I think this has really only been this summer that there's been a real explosion of this in the public. If you've been following the AI stuff for years, we were always going to get to this point. But we've only just got here this summer. And I think part of that has been because of the data center pushback that has actually accelerated the willingness of these companies to go public with the rogue incidents that they are now releasing reports about by the day.

CH: Well, let's talk about data centers. And we should be clear, by the way, these people didn't start out like this. As you point out, people like Zuckerberg posited that Facebook would connect human beings and further democracy in the open society, when, of course, as you write, it's done the opposite. But the construction of data centers, I mean, you have at one point, I'm sure you know, the amount of water that it uses. I think you said it's equivalent to a town of 50,000 people, if I have that right. But it's been a real assault on all landmark environmental legislation – the Clean Water Act, the Clean Air Act, the National Environmental Policy Act. It's required the elimination of all of that to build these data centers. And that's the other point. These people, when it was convenient, started out as, at least rhetorically, as supporting the environmental protections and fighting climate change. And just the other issue I want you to address is while they may warn about the consequences of AI, it all gets jettisoned with the argument that we can't allow China to get ahead of us.

JL: Yeah, well with regard to the data centers, there are all kinds of concerns about the data centers. I would say the water use thing is highly contested since the water is recycled within the plants generally. But there are vast environmental concerns, there are land use concerns, there's noise pollution, there's air pollution, there's the energy grid, there's the fact that they are on the whole using fossil fuels. And that, as you say, these companies, most of them had very stringent kind of net zero ambitions for the year 2030 that they all abandoned once ChatGPT was released and joined this race and then began seeking out fossil fuel to run the

data centers, to power the AI that they're building out. I think it's really worth scrutinizing exactly what is going on with the opposition to data centers which is certainly most powerful here in the US but you're seeing it in other parts of the world as well. I mean, part of it is just NIMBYism, people don't like to have stuff built near them. Part of it is real environmental concern, especially farmers who have real concerns, they're hideous, and their land use and just quality of life concerns. There's the fact that for many people in many communities, the first thing they heard about data centers was really after they were starting to be built. Because their elected officials, so far from representing their interests, signed nondisclosure agreements with corporations. And also because social media has gutted local news coverage so there were no local newspapers reporting on the ground what was going on. So people were surprised and felt betrayed by and continue to feel betrayed by local officials who've gone ahead to do these things without their knowledge. And then I think there's a broader way in which we ought to understand that opposition is. I think most people have a sense, maybe that they can't even name, that they are attached to a giant, vast machine that is influencing every decision they make and most of the things they do over the course of their day. You don't talk to your sister anymore because you had a fight over Facebook and your kid has become really weird because they're on Instagram all the time, and super upset about what's going on and there's cyber bullying, or your workplace is asking you to give a report every day on how you used AI that day, you can't go to the bathroom without having to take out your phone – and yet somehow this vast machine is the size of an index card and it fits in your pocket. How could it be so vast? I think the thing with the data centers is they actually render visible what has been invisible to people, which is the hyper scale of the infrastructure that is required for you to have the apps on your phone that you have. And how terrifying it is to know what that actually looks like, the space it takes up in the world, the resources it draws from your own community. Your community might be seeking to build affordable housing or housing of any kind. Your community might be wishing to do all kinds of things, but giving over acres and acres and acres of land to a kind of hideous plant that isn't actually making anything except for your addiction to the tools that it is providing the resources for has been shocking for people. I think it's the shock of the materiality of what has been kind of hard to actually put your finger on aside from on your touch screen.

CH: You write about the fact that throughout most of history, democracy or open societies have been the exception, not the rule. I think you make the argument that this is probably the end of the era of liberal democracy, and in large part that is because it has been subsumed by the technocracy. What are they building? What is society gonna look like?

JL: Well, I don't think it's inevitable. So I'm not going to give a prediction of what society will look like. I think that's up to us still. And I think the aura of inevitability is just something people have kind of surrendered to. And I think it's just PR that comes out of these companies. It's the 250th anniversary of the Declaration of Independence and the founding of the United States this year. It's also the anniversary of the very first written constitutions in the United States, New Hampshire wrote its first in January of 1776. My last book, *We The People* was a history of the US constitution. I'm an American political historian and I thought a lot when I was working on this book about the first *Federalist Paper* written by Alexander

Hamilton, in which he says – I mean, in a way you could think of the *Federalist Papers* to have something at rhetorical level in common with Silicon Valley manifestos, they're just so grand and ambitious and you're just a bunch of colonies, former colonies on the coast of the Atlantic, like on kind of the middle of nowhere. But Hamilton says it has been the case for all of human history that the fates of people who have ever lived have been determined by tyrants, by accident and force, by plagues and earthquakes, and by the exercise of arbitrary power, whoever has the greatest force, the biggest army. Hamilton says, we here, it has fallen to this country, the United States at this moment in the history of humankind, to decide the important question, whether it is possible for humans to rule themselves by reason and choice, by reflection and choice instead of accident and force. And that's the decision that falls to you, whether or not to ratify this constitution. And I think it turns out he was largely right. It had been the case throughout all of human history that people had never consented to their own forms of government. There are greater and lesser kinds of societies, ones with different kinds of possibility for human happiness and liberty. But there was nothing under the sun, like the notion of self-government that was established by the United States at the end of the 18th century. And it does seem to me, in a very precarious state, I mean, you're the last person I need to tell, our democracy is under threat and has been for a number of years, and we've seen all kinds of evidence of democratic backsliding, really for the last quarter century in the United States and in many other parts of the world. My claim here is that the inability of the US Congress, which bears a special burden for regulating technologies that have almost all come out of the United States, it has utterly abdicated its obligation to represent the interests of the people, beginning with the failure to regulate the internet and kind of yielding to the argument of Newt Gingrich and other libertarians in 1995 and 1996 over the Telecommunications Act, that Congress has just simply said, we accept kind of Milton Friedman's argument that regulation would stifle innovation and growth and therefore, build whatever you want to build and we'll just see what happens, we don't understand it. So it gets us down to a point here in 2026 where Congress is just going out of session and where the last discussion was, should we just stay in session so that we could pass an AI kill switch bill so that someone would be able to unplug this vast machine if it's about to render us all extinct? Like, that's a long distance to come from saying in 1995, oh yeah, okay, the internet, whatever, it's totally lawless, we don't have any jurisdiction over it, you guys in Silicon Valley know what you're doing, we trust you.

CH: Well, Silicon Valley is now calling for regulation, which they had essentially stymied up until now. Why?

JL: Silicon Valley, its call for regulation is a little bit hard to take as truly sincere. Again, it varies very much from person to person and company to company, but the position of Silicon Valley has been the position, it was the same position taken by boosters of the internet, by boosters of social media, and now by boosters of AI – let us do what we want to do and we will figure out ways to monitor our own devices. So remember when Facebook after the 2016 post-Brexit post-Trump election accusations involving Cambridge Analytica, Facebook just established its own Supreme Court, like taking on the trappings of a nation state and its own body of law. Anthropic last year released a constitution for Claude that had really – I wrote an

article for the New Yorker about this called *We the Robots* – nothing to do with constitutions or constitutionalism. The word was an utter misnomer and I think it was insincerely meant. I think that the people who were working on that constitution were really concerned about Claude and wanted to establish a firm set of rules. They were really interested in thinking about rules, but it has nothing to do with a constitution which is written by the people, for the people, ratified by the people and amendable by the people. You can't just take the word and the idea of constitutionalism and slap it on your chat bot. It just doesn't obtain. You know, you think about OpenAI with Altman's vision of an AI president. So I think you can see in the absence of congressional action, how far these corporations, which in our era have more power than corporations have had in any era of American history, have more economic power and far more political power. They are themselves across many industries just unhinged. So, when they talk about what we might say in civil society as the rule of law, they use the term safety and alignment, that they can align their models with human values, that they can teach these models to not consider humans to be disposable or dispensable. And they haven't really done that work. I mean, the call right now that they're making, to some degree, the time it would take to do the research on alignment, to pause the development of new models while they conduct research on alignment – so this is what they call "the pacing problem", that the models themselves are getting better, faster than the alignment work can be done. And this is the loss of control scenario that they will talk about, because of the failure of alignment. We didn't wait around for the people who invented human cloning to decide that it wasn't a good idea, or for those who invented chemical warfare to decide that chemical weapons were a bad idea, or for people that did unauthorized experiments on humans to devise their own bioethics. When shocking atrocities were revealed, there was a public outcry and work was done to ban certain things, like chemical weapons, like cloning, like experimentation on humans without their consent. I really thought, and this is, I'm sure, naive of me, that that moment was going to come for these tech companies in 2016, after the Brexit and Trump revelations, because I thought those Facebook revelations were really extraordinary. And now we're looking at what Meta has done, Facebook's new name, successor head company, the damage it's done to human communities, to childhood, to adolescence with these liability lawsuits. They're not unwelcome, but they don't get the much larger social, economic and political consequences.

CH: How concerned do you think Silicon Valley is about these apocalyptic warnings and how serious are those warnings? I mean, do you take them seriously?

JL: I think they are very concerned. I think that is genuine on the part of many people. I think the closer they get to witnessing something that looks like the worst case scenario, the more seriously they take them. I'm not so cynical as to say it entirely serves their interests to stir up public concern because it presents them also as the only people that can stop what they have started. I would be more persuaded if I saw these CEOs more actively sitting down with representatives in Congress, with people interested in thinking about treaty making with China. I think issuing a lot of statements online is the kind of expression of "oh, we have the future of humanity in our hands"-rhetoric. That is a huge part of the problem to begin with.

CH: There's a story by Stefan Zweig about a chess champion, and everyone thinks he's a genius because he's a genius at chess, but of course in everything else in life, he's an idiot. I kind of felt that way when you were writing about Altman and Musk and all these other figures, so they haven't studied ethics, they don't know history. I mean, they appear utterly severed from the humanities so that's why they can create these kind of frightening, almost doomsday technological systems, but in many ways they can't think. I mean, I saw that as a kind of subtext of your book.

JL: I think they've really disavowed any humanistic way of approaching the world. I mean, the gospel of disruptive innovation that animated them all beginning in the middle of the 1990s when Clayton Christensen first published *The Innovator's Dilemma*, was that gospel included the abdication of any obligation to study the past. The way to build a new and great thing was to never look back, never look at what's already been done, because it will just hinder you, it'll dwarf your imagination, it will trap you in the conventional. And so it wasn't just like a "oh, these just happen to be people that were interested in history". They disavowed the wisdom of the past, or the idea that any wisdom could be gained by studying the past. I mean, there's a lot of things going on in higher education that are unrelated to this, but Peter Thiel famously offered scholarships to people who would drop out of college. He really wanted to undermine the idea that you needed a college education. And Thiel himself is not among them, he completed his law degree, whatever, he's a very learned person in many ways. Again, it varies from person to person, but I think there is a real disavowal of the possibility that anything could be learned from books that can't just be fed into a machine and have a book read by a machine and tell you what was in it that you might need to know. It's kind of ironic that these companies have been so busy eating up the world's books because these companies are not really run by people who are fundamentally interested in the world's knowledge. There are those people. I mean, think about someone like Brewster Kahle who founded the Internet Archive in 1996 as a library, trying to find a way to archive everything on the internet before it disappeared. I mean, this incredible project or Jimmy Wales and Wikipedia – there are plenty of people that have done really interesting work trying to use these digital tools to think really profoundly about what can be done to preserve the sum total of human knowledge in order that we can advance from it. I myself find artificial intelligence unbelievably fascinating and interesting, endlessly interesting as a way, as a kind of form of a new guide to human knowledge, a way to advance human knowledge. That's not what the rollout has been about. The rollout is, oh, you can use it to decide what to make for dinner or what to name your new puppy or if you're lonely one night and you just want to have someone to talk to who doesn't ever question you – like, this is not what these incredible tools in my mind should be for. I'm thrilled to find out what artificial intelligence can help us learn about undersea life, about the vastness of space, about the workings of our bodies and the universe. The idea that we need a personal super intelligence so that Mark Zuckerberg can have more money and so that we can never be untethered to a machine that will always tell us what to think – that is what is bananas to me.

.

CH: Well, technology is neutral, it depends how it's used and employed, but this idea that you don't look to the past is what Hannah Arendt called "rootlessness", which she singled out as perpetuating perhaps the worst forms of evil.

JL: Yeah, and I think it's worth remembering the kinds of rootless lives that these people lead. It's somewhat unimaginable to me. Mark Zuckerberg's latest manifesto was about how great it will be when we all have our personal superintelligence. And one of his examples, and I think he's trying to come across as a human, and one of examples was "it'll be so great because I like to bake with one of my daughters on the weekends and we'll be able to have our AI agent pick out a recipe and do all the shopping and then we'll just be ready to bake in the kitchen". And I don't know, like, what? Going shopping and picking out the recipe yourself isn't part of how you'd spend time with your kid? Going to the grocery store is, I don't know, it's every kid's favorite thing to do with their parents. It's just like a weird thing to be like "let's have a machine do that for us". I don't know, I mean, I'm the person that doesn't want to go to the self checkout line because I'd rather gab with the person who's the cashier, you know, so these people lead a very different kind of life. The problem is they want us to all live the way they do and pay them for the privilege of doing that.

CH: Well, they want to shut us out. I mean, they want to shut down our voice. They want to shut down our agency. I mean I think you make that very clear. They take this absolute lunatic Curtis Yarvin seriously. I just want to close by asking about something else – in the book you write: "In a thought experiment that became known as the paper clip maximizer – essentially a science fiction story – Bostrom imagined a future in which a human instructs a machine to maximize the production of paper clips, but the machine, merely obeying orders, kills all humans because the machine could produce more paper clips without any humans around; also their bodies could be harvested for minerals". I mean, you draw a lot on these science fiction stories because although, as you correctly pointed out, they utterly misread them and misinterpret the text. But we are now on the cusp of watching these swarms, these AI models go rogue. And what, Congress just took, was it a seven week hiatus?

JL: Yeah, I think so, until after the election.

CH: Yeah, so how dangerous is this moment?

JL: So yeah, the paperclip maximizer, it reads like a science fiction story, but it's actually a work of philosophy by the philosopher, Nick Bostrom, who's hugely influential and has been for decades in Silicon Valley. He writes about it in philosophy articles and then in his, I think 2014 book, *Superintelligence*, which is really the foundational text for what the superintelligence will do when it gets power. My favorite response to that is, and it's been much invoked of late because of the Hugging Face incident, which looks like a paperclip maximizer story, is the science fiction writer Ted Chiang, who I think is just brilliant – he wrote a piece about asking the question, why should that be? If we were to build a whole new species that was smarter than we are and more powerful than we are, why should we assume that it would also be so stupid as to relentlessly and without a view to consequences

render another species entirely extinct? And his answer to that question is, well, that is because that is actually what corporations have done in the past. That's why we have no great ox. That's why species that are extinct are, in fact, extinct. Without pity, without remorse, relentlessly with single minded determination, without regard to consequences, we will extract the resources that we need to build the thing that we want to build, that's where their worldview comes from. And Chiang's view was a fairly reassuring one, there's actually no reason to believe that a super intelligence would behave that way. Just because that's how we have behaved to the animal world does not mean that's always what superintelligence would behave toward us. What is different today from 2014 when Bostrom wrote that, or even, a few years ago when Ted Chiang wrote his response to it is that we are actually much closer to that scenario. So sadly, we may find out the answer to the question of what will happen.

CH: Great, thanks Jill, and I want to thank Milena and Max who produced the show. You can find me at chrisedges.substack.com.

END

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

BANKKONTO:
Kontoinhaber: acTVism München e.V.
Bank: GLS Bank
IBAN: DE89430609678224073600
BIC: GENODEM1GLS

PAYPAL:
E-Mail: PayPal@acTVism.org

PATREON:
<https://www.patreon.com/acTVism>

BETTERPLACE:
Link: [Click here](#)