

THE EZRA KLEIN SHOW

The Government Knows A.G.I. is Coming

March 4, 2025, 5:04 a.m. ET



By Ezra Klein

This is an edited transcript of an episode of “The Ezra Klein Show.” You can listen to the conversation by following or subscribing to the show on the NYT Audio App, Apple, Spotify, Amazon Music, YouTube, iHeartRadio or wherever you get your podcasts.

For the last couple of months, I have had this strange experience: Person after person — from artificial intelligence labs, from government — has been coming to me saying: It’s really about to happen. We’re about to get to artificial general intelligence.

What they mean is that they have believed, for a long time, that we are on a path to creating transformational artificial intelligence capable of doing basically anything a human being could do behind a computer — but better. They thought it would take somewhere from five to 15 years to develop. But now they believe it’s coming in two to three years, during Donald Trump’s second term.

They believe it because of the products they’re releasing right now and what they’re seeing inside the places they work. And I think they’re right.

If you’ve been telling yourself this isn’t coming, I really think you need to question that. It’s not web3. It’s not vaporware. A lot of what we’re talking about is already here, right now.

I think we are on the cusp of an era in human history that is unlike any of the eras we have experienced before. And we’re not prepared in part because it’s not clear what it would mean to prepare. We don’t know what this will look like, what it will feel like. We don’t know how labor markets will respond. We don’t know which country is going to get there first. We don’t know what it will mean for war. We don’t know what it will mean for peace.

And while there is so much else going on in the world to cover, I do think there's a good chance that, when we look back on this era in human history, A.I. will have been the thing that matters.

One of the people who reached out to me was Ben Buchanan, the former special adviser for artificial intelligence in the Biden White House. I thought Buchanan would be interesting to bring onto the show for a couple of reasons.

One, this is not a guy working for an A.I. lab. So he's not being paid by the big A.I. labs to tell you this technology is coming.

Two, he has been at the nerve center of the policy we have been making in recent years — particularly to try to stay ahead of China.

Three, because there has been a profound changeover in administrations. The new administration — with Elon Musk and Marc Andreessen and David Sacks and JD Vance — has a lot of people with very strong views on A.I.

We're at this moment of a big transition in policymakers, and they are probably going to be in power when artificial general intelligence, or something like it, hits the world. So what are they going to do? What kinds of decisions are going to need to be made? And what kinds of thinking do we need to start doing now to be prepared for something that virtually everybody who works in this area is trying to tell us, as loudly as they possibly can, is coming?

Ezra Klein: Ben Buchanan, welcome to the show.

Ben Buchanan: Thanks for having me.

After the end of the Biden administration, I got calls from a lot of people who wanted to tell me about all the great work they did. But when you called me, you wanted to warn people about what you thought was coming.

What's coming?

I think we are going to see extraordinarily capable A.I. systems. I don't love the term artificial general intelligence, but I think that will fit in the next couple of years, quite likely during Donald Trump's presidency.

There's a view that A.G.I. has always been something of corporate hype or speculation. And one of the things I saw in the White House, when I was decidedly not in a corporate position, were trend lines that looked very clear. What we tried to do under President Biden's leadership was get the U.S. government and our society ready for these systems.

Before we get into what it would mean to get ready, what do you mean when you say “extraordinarily capable A.I. systems”?

The canonical definition of A.G.I., which, again, is a term I don't love, is a system —

It'd be good if every time you say A.G.I., you caveat that you dislike the term. [Laughs.]

It will sink in, right? [Laughs.]

Yes, people will really enjoy that.

I'm trying to get it in the training data, Ezra.

A canonical definition of A.G.I. is a system capable of doing almost any cognitive task a human can do. I don't know that we'll quite see that in the next four years or so, but I do think we'll see something like that, where the breadth of the system is remarkable but also its depth, its capacity to, in some cases, exceed human capabilities, regardless of the cognitive discipline —

Systems that can replace human beings in cognitively demanding jobs.

Yes, or key parts of cognitive jobs. Yes.

I will say I am also pretty convinced we're on the cusp of this. So I'm not coming at this as a skeptic. But I still find it hard to mentally live in the world of it.

So do I.

I recently used Deep Research, which is a new OpenAI product. It's on their pricier tier. Most people, I think, have not used it. But it can build out something that's more like a scientific analytical brief in a matter of minutes.

I work with producers on the show. I hire incredibly talented people to do very demanding research work. And I asked Deep Research to do this report on the tensions between the Madisonian Constitutional system and the highly polarized nationalized parties we now have. And what it produced in a matter of minutes was at least the median of what any of the teams I've worked with on this could produce within days.

I've talked to a number of people at firms that do high amounts of coding, and they tell me that by the end of this year or next year they expect most code will not be written by human beings.

I don't really see how this cannot have labor market impact.

I think that's right. I'm not a labor market economist, but I think that the systems are

extraordinarily capable. In some ways, I'm very fond of the quote: The future is already here — it's just unevenly distributed.

Unless you are engaging with this technology, you probably don't appreciate how good it is today. And it's important to recognize that today is the worst it's ever going to be. It's only going to get better.

And I think that is the dynamic that we were tracking in the White House — and that I think the next White House and our country as a whole is going to have to track and adapt to in really short order.

What's fascinating to me is that this is the first revolutionary technology that is not funded by the Department of Defense, basically. And if you go back historically, over the last hundred years or so, nukes, space, the early days of the internet, the early days of the microprocessor, the early days of large-scale aviation, radar, the global positioning system — the list is very, very long — all of that tech fundamentally comes from Department of Defense money.

It's the private sector inventing it, to be sure. But the central government role gave the Department of Defense and the U.S. government an understanding of the technology that, by default, it does not have in A.I. It also gave the U.S. government a capacity to shape where that technology goes that, by default, we don't have in A.I.

There are a lot of arguments about A.I. in America. But the one thing that seems almost universally agreed upon, and seems to be the dominant controlling priority in policy, is that we need to get to A.G.I. before China does. Why?

I do think there are profound economic, military and intelligence capabilities that would be downstream of getting to A.G.I. or transformative A.I. And I do think it is fundamental for U.S. national security that we continue to lead in A.I.

The quote that certainly I have thought about a fair amount was actually from Kennedy in his famous Rice University speech in 1962:

Archived clip of John F. Kennedy: We choose to go to the moon in this decade and do the other things not because they are easy but because they are hard.

Everyone remembers it because he's saying we're going to the moon. But actually I think he gives the better line when he talks about the importance of space.

Archived clip of John F. Kennedy: *For space science, like nuclear science and all technology, has no conscience of its own. Whether it will become a force for good or ill depends on man. And only if the United States occupies a position of pre-eminence can we help decide whether this new ocean will be a sea of peace or a new terrifying theater of war.*

And I think that is true in A.I. There's a lot of tremendous uncertainty about this technology.

I am not an A.I. evangelist. I think there are huge risks to this technology. But I do think there is a fundamental role for the United States in being able to shape where it goes — which is not to say we don't want to work internationally, which is not to say we don't want to work with the Chinese.

It's worth noting that in the president's executive order on A.I., there's a line in there saying we are willing to work even with our competitors on A.I. safety and the like. But it is worth saying that I think, pretty deeply, there is a fundamental role for America here that we cannot abdicate.

Paint the picture for me. You say there would be great economic, national security and military risks if China got there first. Help the audience imagine a world where China gets there first.

Let's look at a narrow case of A.I. for intelligence analysis and cyberoperations.

It's pretty out in the open that if you had a much more powerful A.I. capability, that would probably enable you to do better cyberoperations on offense and on defense.

What is a cyberoperation? Breaking into an adversary's network to collect information — which, if you're collecting in a large enough volume, A.I. systems can help you analyze.

We actually did a whole big thing through D.A.R.P.A., the Defense Advanced Research Projects Agency, called the A.I. Cyber Challenge to test out A.I.'s capabilities to do this. And I would not want to live in a world in which China has that capability on offense, defense and cyber and the United States does not. And I think that is true in a bunch of different domains that are core to national security competition.

My sense is that most people, most institutions, are pretty hackable to a capable state actor. And now both the state actors are going to get better at hacking, and they're going to have much more capacity to do it. You can have many more A.I. hackers than you can human hackers.

Are we about to enter a world where we are much more digitally vulnerable as normal

people? And I'm not just talking about people whom the states might want to spy on. But you will get versions of these systems that all kinds of bad actors will have.

Do you worry it's about to get truly dystopian?

What we mean canonically when we speak of hacking is finding vulnerability in software and exploiting that vulnerability to get illicit access. And I think it is right that more powerful A.I. systems will make it easier to find vulnerabilities and exploit them and gain access. And that will yield an advantage to the offensive side of the ball.

I think it is also the case that more powerful A.I. systems on the defensive side will make it easier to write more secure code in the first place, to reduce the number of vulnerabilities that can be found and to better detect the hackers that are coming in.

We tried as much as possible to shift the balance toward the defensive side of this. But I think it is right that, in the coming years, in this transition period we've been talking about, there will be a period in which older legacy systems that don't have the advantage of the newest A.I. defensive techniques or software development techniques will, on balance, be more vulnerable to a more capable offensive actor.

Which is what most people use.

I don't know if that's right, actually. You have an iPhone in your pocket. Or Google picks it up.

People are often not that quick about updating. The less tech literate you are, the more vulnerable you're going to be with this.

Sure, I'm thinking more about legacy power systems and server mainframes that could be two decades old and that haven't been turned off all the time. So that is where I feel the risk most acutely. For all of the risks that come with the monoculture of most people's personal tech platforms these days, one of the upsides is they do push security updates pretty regularly. They push them with new emojis that get people to download the updates.

And on balance, people are probably better at patching their personal software now than they were 15 years ago,

Yeah, it gets very annoying if you don't.

The flip of that is a lot of people also worry about the security of the A.I. labs themselves.

It is very valuable for another state to get the latest OpenAI system. The people at these companies — and I've talked to them about this — say: On the one hand, this is a

problem. And on the other hand, it's really annoying to work in a truly secure way.

I've worked in a SCIF — sensitive compartmented information facility — for the last four years, a secure room where you can't bring your phone and all that. That is annoying. There's no doubt about it.

How do you feel about the vulnerability right now of A.I. labs?

I worry about it. There's a hacking risk here. Also if you hang out in the right San Francisco house party, they're not sharing the model, but they are talking to some degree about the techniques they use, which have tremendous value.

To come back to the intellectual through line that A.I. is national security relevant, maybe even world-changing, technology that's not coming from the auspices of the government and doesn't have the government imprimatur of security requirements — that shows up in this way, as well.

In the national security memorandum, the president's side tried to signal this to the labs and tried to say to them: We, as the U.S. government, want to help you in this mission.

Sign up for the Opinion Today newsletter Get expert analysis of the news and a guide to the big ideas shaping the world every weekday morning. [Get it sent to your inbox.](#)

This was signed in October 2024, so there wasn't a ton of time for us to build on that. But I think it's a priority for the Trump administration. And I can't imagine anything that is more nonpartisan than protecting American companies that are inventing the future.

There's a dimension to this that people bring up to me a lot that is interesting about processing information.

Compared to spy games between the Soviet Union and the United States, we all have a lot more data now. We have all the satellite data. Obviously, we eavesdrop on each other and have all these kinds of things coming in.

I'm told by people who know this better than I do that there's a huge choke point of human beings and the fairly rudimentary programs analyzing that data. So to have these truly intelligent systems that are able to inhale that and do pattern recognition is a very significant change in the balance of power.

I think we were pretty public about this, and President Biden signed a national security

memorandum, which is basically the national security equivalent of an executive order, that says: This is a fundamental area of importance for the United States.

I don't even know the amount of satellite images that the United States collects every single day, but it's a huge amount, and we have been public about the fact that we simply do not have enough humans to go through all of this satellite imagery. It would be a terrible job even if we did. And there is a role for A.I. in going through these images of hot spots around the world, of shipping lines and all that, and analyzing them in an automated way and surfacing the most interesting and important ones for human review.

At one level you can look at this and say: Well, doesn't software do that?

That, at some level, of course, is true. But at another level, you could say the more capable the software, the more capable the automation of that analysis, the more intelligence advantage you extract from that data. And that ultimately leads to a better position for the United States.

The first- and second-order consequences of that are also striking. In a world where you have strong A.I., the incentive for spying goes up. Because if we are currently collecting more data than we can analyze, then each marginal piece of data we're collecting isn't that valuable.

I think that's basically true. I firmly believe that you need to have rights and protections that hopefully are pushing back and saying: No, there are key kinds of data here, including data on your own citizens and, in some cases, citizens of allied nations, that you should not collect, even if there's an incentive to collect it.

And for all of the flaws of the United States intelligence oversight process and all the debates we could have about this, we do have those kinds of structures.

And that is fundamentally more important, for the reason you suggest, in the era of tremendous A.I. systems.

How frightened are you by the national security implications of all this, including for the possibilities for surveillance states?

Samuel Hammond, who's an economist at the Foundation for American Innovation, had this piece months back called "Ninety-Five Theses on A.I." One point he makes that I think about a lot is: If we had the capacity for perfect enforcement, a lot of our current laws would be constricting.

Laws are written with the knowledge that human labor is scarce. And there's this question of what happens when the surveillance state gets really good. What happens

when A.I. makes the police state a very different kind of thing than it is? What happens when we have warfare of endless drones?

You hear about the company Anduril a lot now. They have a relationship with OpenAI. Palantir has a relationship with Anthropic. We're about to see a real change in a way that I think is frightening, from a national security perspective.

I understand why we don't want China way ahead of us. But how do you think about the capacities it gives our own government?

I would decompose this question about A.I. and autocracy or the surveillance state into two parts.

The first is the China piece of this. How does this play out in a state that is truly, in its bones, an autocracy and doesn't even make any pretense toward democracy?

I think we could agree pretty quickly here that this makes very tangible something that is probably core to the aspiration of their society — of a level of control that only an A.I. system could help bring about. I just find that terrifying.

As an aside, there's a saying in both Russian and Chinese: "Heaven is high, and the emperor is far away."

Historically, even in those autocracies, there was some kind of space where the state couldn't intrude because of the scale and the breadth of the nation. And in those autocracies, A.I. could make the force of government power worse.

Then there's the more interesting question in the United States: What is the relationship between A.I. and democracy?

I share some of the discomfort here. There have been thinkers, historically, who have said that part of the ways we revise our laws is when people break the laws. There's a space for that, and I think there is a humanness to our justice system that I wouldn't want to lose.

We tasked the Department of Justice to run a process and think about this and come up with principles for the use of A.I. in criminal justice. In some cases, there are advantages to it — like cases are treated alike with the machine.

But also there's tremendous risk of bias and discrimination and so forth because the systems are flawed and, in some cases, because the systems are ubiquitous. And I do think there is a risk of a fundamental encroachment on rights from the widespread unchecked use of A.I. in the law enforcement system that we should be very alert to and

that I, as a citizen, have grave concerns about.

I find that this all makes me incredibly uncomfortable. And one of the reasons is that there is a — what's the right way to put this? — it's like we are trying to build an alliance with another almost interplanetary ally, and we are in a competition with China to make that alliance. But we don't understand the ally, and we don't understand what it will mean to let that ally into all of our systems and all of our planning.

As best I understand it, every company and every government really working on this believes that in the not too distant future, you're going to have much better, faster and more dominant decision-making loops once A.I. is more autonomous. We are rushing toward A.G.I. without really understanding what that is or what that means.

It seems potentially like a historically dangerous thing that A.I. has reached maturation at the exact moment that the U.S. and China are in this Thucydides' trap-style race for superpower dominance. That's a pretty dangerous set of incentives in which to be creating the next turn in intelligence on this planet.

Yes, there's a lot to unpack here. So let's just go in order.

Basically, bottom line: I greatly share a lot of this discomfort. Part of the appeal of the export controls is that it identifies a choke point that can differentially slow the Chinese down, create space for the United States to have a lead and, ideally, in my view, spend that lead on safety and coordination and not rushing ahead — including, again, potential coordination with the Chinese while not exacerbating this arms-race dynamic.

I would not say that we tried to race ahead in applications to national security. Part of the national security memorandum is a pretty lengthy description of what we're not going to do with A.I. systems and a whole list of prohibited use cases and then high impact use cases. And there's a governance and risk management framework —

Yes, but you're not in power anymore.

Well, that's fair. The Trump administration has not repealed this. But I do think it's fair to say that, for the period while we had power, the foundation we were trying to build with A.I., we were very cognizant of the dynamic you were talking about, of a race to the bottom on safety. And we were trying to guard against it, even as we tried to ensure a position of U.S. pre-eminence.

Is there anything to the concern that, by treating China as such an antagonistic competitor on this — where we will do everything including export controls on advanced technologies to hold them back — that we have made them into a more intense

competitor?

I do not want to be naive about the Chinese system or the ideology of the Chinese Communist Party. They want strength and dominance and to see the next era be a Chinese era. So maybe there's nothing you can do about this, but it is pretty damn antagonistic to try to choke off the chips for the central technology of the next era to the other biggest country.

I don't know that it's pretty antagonistic to say we are not going to sell you the most advanced technology in the world. That's not a declaration of war. That is not, in itself, a declaration of a cold war. I think it is just saying: This technology is incredibly important.

Do you think that's how they understood it?

This is more academic than you want, but my academic research when I started as a professor was basically on the Thucydides' trap, or what in academia we call a security dilemma, of how nations misunderstand each other. So I'm sure the Chinese and the United States misunderstand each other at some level in this area.

But I think the plain reading of the facts is that I don't think not selling chips to them is a declaration of war —

But I don't think they do misunderstand us. Look, I'm aware of how politics in Washington works. I've seen the turn toward a much more confrontational posture with China. I know that Jake Sullivan and President Biden wanted to call this strategic competition and not a new cold war. I get all that, and I think it's true.

But we have just talked about — and you did not argue the point — that our dominant view is we need to get to this technology before they do. I don't think China looks at this like: Oh, nobody would ever sell us the top technology.

I think they understand what we're doing here.

To some degree — I don't want to sugarcoat this — I'm sure they do see it that way.

On the other hand, we set up an A.I. dialogue with China. I flew to Geneva and met them, and we tried to talk to them about A.I. safety and the like. So I do think in an area as complex as A.I., you can have multiple things be true at the same time.

I don't regret for a second the export controls. And I think, frankly, we are proud to have done them when we did them because it has helped ensure that a couple of years later, we retained the edge in A.I. — for as good or as talented as DeepSeek is.

What made DeepSeek such a shock, I think, to the American system was that it appeared

to be trained on, much less compute, for much less money and was competitive at a high level with our frontier systems.

How did you understand what DeepSeek was and what assumptions it required us to rethink?

Let's take one step back and track the history of DeepSeek here.

We had been watching DeepSeek in the White House since November 2023 or thereabouts, when they put out their first coding system. There's no doubt that DeepSeek engineers are extremely talented, and they got better and better at their systems throughout 2024.

We were heartened when their chief executive said that the biggest impediment to what DeepSeek was doing was not their inability to get money or talent but their inability to get advanced chips. Though clearly they still did get some chips — some they bought legally, some they smuggled, so it seems.

In December 2024, they came out with a system called DeepSeek-V3, which actually is the one that should have gotten the attention. It didn't get a ton of attention, but it did show they were making strong algorithmic progress and basically making systems more efficient.

And then in January 2025, they came out with a system called DeepSeek-R1. R1 is actually not that unusual. No one would expect that to take a lot of computing power just as a reasoning system that extends the underlying V3 system.

That's a lot of nerdspeak. But the key thing here is when you look at what DeepSeek has done, I don't think the media hype around it was warranted, and I don't think it changes the fundamental analysis of what we are doing.

They still are constrained by computing power. We should tighten the screws and continue to constrain them. They're smart. Their algorithms are getting better. But so are the algorithms of U.S. companies.

And this should be a reminder that chip controls are important, China is a worthy competitor here, and we shouldn't take anything for granted. But I don't think this is the time to say the sky is falling or the fundamental scaling laws have broken.

Where do you think they got their performance increases from? We read their papers. They're smart people who are doing exactly the same kind of algorithmic efficiency work that companies like Google and Anthropic and OpenAI are doing.

One common argument I have heard on the left — Lina Khan made this point — was that DeepSeek proved our whole paradigm of A.I. development was wrong: We did not need all this compute, we did not need these giant megacompanies, that DeepSeek was showing a way toward a decentralized almost solarpunk version of A.I. development. And, in a sense, the American system and imagination had been captured by these three big companies.

But what we're seeing from China is: That wasn't necessarily needed. We could do this on less energy, fewer chips, less footprint.

Do you buy that?

I think two things are true here. The first is there will always be a frontier, at least for the foreseeable future. There will be a frontier that is computationally and energy intensive. And we want our companies to be at that frontier.

Those companies have very strong incentive to look for efficiencies, and they all do. They all want to get every single, last juice of insight from each squeeze of computation. But they will continue to need to push the frontier.

Then, in addition to that, there will be a slower diffusion that lags the frontier, where algorithms get more efficient, fewer computer chips are required, less energy is required. And we need to win both those competitions.

The A.I. firms want the export controls, but the semiconductor firms don't. DeepSeek rocked the U.S. stock market by making people question Nvidia's long-term worth. Nvidia very much doesn't want these export controls.

You were at the White House, at the center of a bunch of this lobbying back and forth. How do you think about this?

Every advanced A.I. chip that gets made will get sold. The market for these chips is extraordinary right now, and I think for the foreseeable future. So I think our view was we put the export controls on —

Wait, but Nvidia didn't think that. The stock market didn't think that.

We put on the first export controls in October 2022. Nvidia stock has 10x-ed since then.

I'm not saying we shouldn't do the export controls. But I want you to take the strong version of the argument, not the weak one. I don't think Nvidia's chief executive is wrong to say that if Nvidia cannot export its top chips to China, that reduces the market for Nvidia's chips in the long run.

Sure. I think the dynamic is right. If they had a bigger market, they could charge on the margins more. That's obviously the supply and demand.

I think our analysis was: Considering the importance of these chips and the A.I. systems they make to U.S. national security, this is a trade-off that's worth it.

Nvidia, again, has done very well since we put the export controls out. Nvidia is currently trading, even post-DeepSeek, something like 50 times earnings. So the market is continuing to expect they will grow. And I agree with that.

The Biden administration was also generally concerned with A.I. safety. I think it was influenced by people who care about A.I. safety. And that has created a backlash from the accelerationist side of this debate.

I want to play a clip for you from Marc Andreessen, a prominent venture capitalist and a top Trump adviser, describing the conversations he had with the Biden administration on A.I. and how they radicalized him in the other direction.

Archived clip of Marc Andreessen: Ben [Horowitz] and I went to Washington in May of '24. And we couldn't meet with Biden because, as it turns out, at the time, nobody could meet with Biden. But we were able to meet with senior staff. And so we met with very senior people in the White House, in the inner core. And we basically relayed our concerns about A.I. And their response to us was: Yes, the national agenda on A.I., as we will implement in the Biden administration in the second term is: We are going to make sure that A.I. is going to be only a function of two or three large companies. We will directly regulate and control those companies. There will be no start-ups. This whole thing where you guys think you can just start companies and write code and release code to the internet, those days are over. That's not happening.

Were you part of the conversation he was describing there?

I met with him once. I don't know exactly —

Would that characterize the conversation he had with you?

He talked about concerns related to start-ups and competitiveness. And I think my view on this is: You look at our record on competitiveness, and it's pretty clear that we wanted a dynamic ecosystem.

The A.I. executive order, which President Trump just repealed, had a pretty lengthy section on competitiveness. The Office of Management and Budget memo, which governs how the U.S. government buys A.I., had a whole callout in it, saying: We want to buy from a wide variety of vendors.

The CHIPS and Science Act also has a bunch of things in there about competition.

So I think our view on competition is pretty clear. Now I do think there are structural dynamics related to scaling laws and the like that will force things toward big companies that I think, in many respects, we were pushing against. But I think our track record on competition is pretty clear.

The view that I understand Andreessen arguing with — which is a view I have heard from people in the A.I. safety community but is not a view I had necessarily heard from the Biden administration — was that you will need to regulate the frontier models of the biggest labs when it gets sufficiently powerful.

And in order to do that, you will need controls on those models. You just can't have the model and everything floating around so everybody can run this on their home laptop.

I think that's the tension he's getting at. It gets at a bigger tension, which is how much to regulate this incredibly powerful and fast-changing technology such that, on the one hand, you're keeping it safe, but on the other hand, you're not overly slowing it down or making it impossible for smaller companies to comply with these new regulations as they're using more and more powerful systems.

In the president's executive order, we actually tried to wrestle with this question, and we didn't have an answer when that order was signed in October 2023.

What we did on the open-source question in particular — and I think we should just be precise here, at the risk of being academic again: But what we're talking about are open-weight systems.

Can you explain what weights and open weights are in this context?

In the training process for an A.I. system, you run an algorithm through this huge amount of computational power that processes the data. The output at the end of that training process, loosely speaking — and I stress this is the loosest possible analogy — are roughly akin to the strength of connections between the neurons in your brain. And in some sense, you could think of this as the raw A.I. system.

When you have these weights, one thing that some companies like Meta and DeepSeek choose to do is publish them out on the internet, which makes them what we call open-weight systems.

The crucial thing about an open-weight system, on the good side, is that it's much easier to innovate with that system to use as a basis for future systems because you've got access to the raw thing.

Maybe the riskier side is: Any safeguards that were built into that system — say, to refuse when a user asks you to help develop a biological weapon — are pretty easy to remove.

I'm a huge believer in the open-source ecosystem. Many of the companies that publish the weights for their system do not make them open-source. They don't publish the code and the like. So I don't think they should get the credit of being called open-source systems — at the risk of being pedantic.

But open-weight systems are something we thought a lot about in '23 and '24. And we sent out a pretty wide-ranging request for comment. We got a lot of comments back, and what we came to in the report that was published in July or so of '24 was basically: There is no evidence yet to constrain the open-weight ecosystem. The open-weight ecosystem does a lot for innovation and the like, which I think is manifestly true. But we should continue to monitor this as the technology gets better.

Basically exactly the way that you described.

When you were getting those comments, not just on the open-weight models but also when you were talking to the heads of these labs, what did they want? What would you say was the consensus from the A.I. world, to the extent there was one, of what they needed to get there quickly?

And also — because I know many people in these labs are worried about what it would mean if these systems were unsafe — what would you describe as their consensus on safety?

I mentioned before this core intellectual insight: This technology — for the first time, maybe in a long time — is a revolutionary one that is not funded by the government. In its early incubator days, that was the theme from the labs — a sort of: Don't you know we're inventing something very powerful? Ultimately, it's going to have implications on the work you do in national security and the way we organize our society.

More than any kind of individual policy request, they were basically saying: Get ready for this.

The closest thing we did to any kind of regulation was, after the labs made voluntary commitments to do safety testing, we said: You have to share those safety test results with us, and you have to help us understand where the technology is going.

That only applied to the top couple of labs. The labs never knew that was coming and weren't all thrilled about it when it came out.

So the notion that this was kind of a regulatory capture — that we were asked to do this — is simply not true.

But in my experience, I never got individual policy lobbying from the labs. I got much more: This is coming. It's coming much sooner than you think. Make sure you're ready.

To the degree that they were asking for something in particular, it was a corollary of that: We're going to need a lot of energy, and we want to do that here in the United States, and it's really hard to get the power here in the United States.

But that has become a pretty big question. If this is all as potent as we think it will be, and you end up having a bunch of the data centers containing all the model weights and everything else, in a bunch of, say, Middle Eastern petrostates — because they will give you huge amounts of energy access in return for having some purchase on this A.I. world — there is something to this question.

Yeah. This is actually an area of bipartisan agreement.

We really started to pay a lot of attention to this in the later part of '23 and most of '24, when it was clear this was going to be a bottleneck.

In his last week or so in office, President Biden signed an A.I. infrastructure executive order, which has not been repealed, which basically tries to accelerate the power development and the permitting of power and data centers here in the United States, basically for the reason that you mentioned.

As someone who truly believes in climate change and environmentalism and clean power, I thought there was a double benefit to this, which is that if we did it here in the United States, it could catalyze the clean energy transition. And these companies, for a variety of reasons, are willing to pay more for clean energy and on things like geothermal.

Our hope was we could catalyze that development and bend the cost curve and have these companies be the early adopters of that technology so we'd see a win on the climate side, as well.

There are warring cultures around how to prepare for A.I. — I have mentioned A.I. safety and A.I. accelerationism.

JD Vance just went to the big A.I. summit in Paris. I'll play a clip of what he said:

Archived clip of JD Vance: When conferences like this convene to discuss a cutting-edge technology, oftentimes I think our response is to be too self-conscious, too risk averse. But never have I encountered a breakthrough in tech that so clearly calls us to do precisely the opposite.

What do you make of that?

I think he is setting up a dichotomy there that I don't quite agree with. And the irony of that is, if you look at the rest of his speech, which I did watch, there's actually a lot that I do agree with.

For example, I think he's got four pillars in the speech. One is about centering the importance of workers. One is about American pre-eminence.

And those are entirely consistent with the actions that we took and the philosophy that I think the administration of which I was a part espoused and that I certainly believe in.

Insofar as he is saying that safety and opportunity are in fundamental tension, I disagree. If you look at the history of technology and technology adaptation, the evidence is pretty clear that the right amount of safety action unleashes opportunity and, in fact, unleashes speed.

One of the examples that we studied a lot and talked to the president about was the early days of railroads. There were tons of accidents and crashes and deaths. And people were not inclined to use railroads as a result.

What started happening was safety standards and safety technology: block signaling so that trains could know when they were in the same area, air brakes so that trains could brake more efficiently, the standardization of train track widths and gauges and the like.

This was not always popular at the time, but with the benefit of hindsight, it is very clear that kind of technology and, to some degree policy development of safety standards, made the American railroad system in the late 1800s. And this is the pattern that shows up a bunch throughout the history of technology.

To be very clear, it is not the case that every safety regulation on every technology is good. And there are certainly cases where you can overreach and you can slow things down and choke things off. But I don't think it's true that there's a fundamental tension between safety and opportunity.

That's interesting because I don't know how to get this point of regulation right.

I think the counterargument to Vice President Vance is nuclear. Nuclear power is a

technology that both held extraordinary promise, maybe still does. And also you can really imagine every country wanting to be in the lead on it.

But the series of accidents — most of which did not even have a particularly significant body count — were so frightening to people that the technology got regulated to the point that nuclear advocates believe it has been largely strangled in the crib from what it could be.

The question then is: When you look at the actions we have taken on A.I., are we strangling it in the crib, and have we taken actions that are akin to —

I'm not saying that we've already done it. Look, if these systems are going to get more powerful and they're going to be in charge of more things, things are both going to go wrong and they're going to go weird. It's not possible for it to be otherwise, to roll out something this new in a system as complex as human society.

So there's going to be this question of: What are the regimes that make people feel comfortable moving forward from those kinds of moments?

I think that's a profound question. What we tried to do in the Biden administration was set up institutions in the government to do that in as cleareyed, tech-savvy way as possible.

Again, with the one exception of the safety test results sharing, which some of the chief executives estimate cost them one day of employee work, we did not put anything close to regulation in place.

We created something called the U.S. Artificial Intelligence Safety Institute, which is purely national-security focused — cyber-risks, biorisks, A.I. accident risks — and purely voluntary and that has relationships, a memorandum of understanding, with Anthropic, with OpenAI, even with xAI, Elon's company.

And basically I think we saw that as an opportunity to bring A.I. expertise into the government. To build relationships between the public and private sector in a voluntary way. And as the technology develops, it will now be up to the Trump administration to decide what they want to do with it.

But I think you are quite diplomatically understating what is a genuine disagreement here.

What I would say Vance's speech was signaling was the arrival of a different culture in the government around A.I. There has been an A.I. safety culture where we have all these conferences about what could go wrong.

And Vance is saying: Stop it. Yes, maybe things could go wrong. But instead we should be focused on what could go right.

And frankly, I would say this is the generalized view of the Trump-Musk administration — which I think is, in some ways, the right way to think about the administration. That if something goes wrong, we'll deal with the thing that went wrong afterward. But what you don't want to do is move too slowly because you're worried about things going wrong. That it's better to break things and fix them than to have moved too slowly in order not to break them.

I think it's fair to say that there is a cultural difference with the Trump administration and us on some of these things.

But we held conferences on what you could do with A.I. and the benefits of A.I. We talked all the time about the need to mitigate these risks, but you're doing so so you can capture the benefits.

And I'm someone who reads an essay like "Machines of Loving Grace," by Dario Amodei, the chief executive of Anthropic, that is basically about the upside of A.I., and says: There's a lot in here we can agree with.

The president's executive order said we should be using A.I. more in the executive branch. So I hear you on the cultural difference. I get that. But I think when the rubber meets the road we were comfortable with the notion that you could both realize the opportunity of A.I. while doing it safely.

And now that they are in power, they will have to decide how to translate Vice President Vance's rhetoric into a governing policy. My understanding of their executive order is they've given themselves six months to figure out what they're going to do. And I think we should judge them on what they do.

Let me ask about the other side of this. What I liked about Vance's speech is: I think he's right that we don't talk enough about opportunities. But more than that, we are not preparing for opportunities.

If you imagine that A.I. will have the effects and possibilities that its backers and advocates hope, one thing that implies is: We're going to start having a much faster pace of the discovery or proposal of novel drug molecules. A very high promise.

The idea here, from the people I've spoken to, is that A.I. should be able to ingest an amount of information and build the modeling of diseases in the human body that could get us a much better drug discovery pipeline.

If that were true, you can ask: What's the choke point going to be?

Our drug-testing pipeline is incredibly cumbersome. It's very hard to get the animals you need for trials, very hard to get the human beings you need for trials. You could do a lot to make that pipeline faster.

And this is true in a lot of different domains — education, etc. It's pretty clear that the choke points will become the difficulty of doing things in the real world. And I don't see society preparing for that. Maybe we're not doing that much on the safety side because we don't know what we should do.

But on the opportunity side, this question of how to actually make it possible to translate the benefits of A.I. quickly seems like a much richer conversation than I've seen anybody seriously having.

I basically agree with all of that. The conversation when we were in the government, especially in '23 and '24, was starting to happen. We looked at the clinical trials thing.

You've written about health care for however long, and I don't claim expertise on health care. But it does seem to me that we want to get to a world where we can take the breakthroughs, including breakthroughs from A.I. systems, and translate them to market much faster.

This is not a hypothetical thing. It's worth noting that, quite recently, Google came out with — I think they called it a co-scientist. Nvidia and the Arc Institute, which does great work, had the most impressive biodesign model ever that has a much more detailed understanding of biological molecules. A group called FutureHouse has done similarly great work in science.

So I don't think this is hypothetical. This is happening right now. And I agree with you that there's a lot that can be done institutionally and organizationally to get the federal government ready for this.

I've been wandering around Washington, D.C. this week and talking to a lot of people involved in the Trump administration — people from different factions of what I consider the modern right.

I've been surprised how many people understand either what Trump and Musk and DOGE are doing, or at least what it will end up allowing, as related to A.I. — including people who are not a part of the tech right, and from whom I would not really expect to hear that.

What they basically say is: The federal government is too cumbersome to take advantage

of A.I. as a technology. So if the whole point of A.I. is that it accelerates cognitive work, the government needs to be stripped down and rebuilt to take advantage of A.I. — which, like it or hate it, is what Musk and DOGE are doing. And the dismantling of the government allows for a creative destruction that paves the way for the government to better use A.I.

Do you buy that?

It feels orthogonal to what I've observed from DOGE. I think Musk is someone who does understand what A.I. can do, but I don't know how starting with the United States Agency for International Development, for example, prepares the U.S. government to make better A.I. policy.

So I guess I don't buy that as the motivation for DOGE.

Is there something to the broader argument, though? I will say, I don't buy the argument about DOGE — I would make the same point you just made.

But what I do buy is that I know how the federal government works pretty well, and it is too slow to modernize technology. It is too slow to work across agencies. And it is too slow to radically change the way things are done and take advantage of things that could be productivity enhancing.

I couldn't agree more. The existence of my job as the White House special adviser for A.I. — which David Sacks is now — and I had this job in 2023 — existed because President Biden said very clearly, both publicly and privately: We cannot move at the typical government pace. We have to move faster here.

I think we probably need to be careful, and I'm not here for stripping it all down. But I agree with you. We have to move much faster.

Another major part of Vice President Vance's speech was signaling to the Europeans that we are not going to sign on to complex multilateral negotiations and regulations that could slow us down — and that if they passed such regulations in a way that is penalizing our A.I. companies, we would retaliate.

How do you think about the differing position the new administration is moving into vis-à-vis Europe and its broad approach to tech regulation?

I think the honest answer here is we had conversations with Europe as they were drafting the European Union A.I. Act.

At the time that I was in the E.U., the A.I. Act was still nascent. The act had passed, but a

lot of the actual details of it had been kicked to a process that is still unfolding. So —

Speaking of slow-moving bureaucracies.

Exactly. So maybe this is a failing on my part, but I did not have particularly detailed conversations with the Europeans beyond a general articulation of our views. They were respectful. We were respectful. But I think it's fair to say we were taking a different approach than they were taking. And insofar as safety and opportunity are a dichotomy — which I don't think are a pure dichotomy — we were ready to move very fast in the development of A.I.

One of the other things that Vance talked about, and that you said you agreed with, is making A.I. pro-worker. What does that mean?

It's a vital question. We instantiate that in a couple of different principles. The first is that A.I. in the workplace needs to be implemented in a way that is respectful of workers and the like.

And one of the things I know the president thought a lot about was how it is possible for A.I. to make workplaces worse in a way that is dehumanizing and degrading and ultimately destructive for workers. So that is the first distinct piece of it that I don't want to neglect.

The second is: I think we want to have A.I. deployed across our economy in a way that increases workers, agencies and capabilities. And I think we should be honest that there's going to be a lot of transition in the economy as a result of A.I.

I don't know what that will look like. You can find Nobel prizewinning economists who will say it won't be much. You can find other folks who will say it will be a ton. I tend to lead toward the side that says it's going to be a lot. But I'm not a labor economist.

The line that Vice President Vance used is the exact same phrase that President Biden used, which is: Give workers a seat at the table in that transition.

And I think that is a fundamental part of what we're trying to do here and, I presume, what they're trying to do here.

I've heard you beg off on this question by saying you're not a labor economist.

I am not a labor economist, Ezra. [Laughs.]

I will promise you the labor economists do not know what to do about A.I. You were the top adviser for A.I. You were at the nerve center of the government's information about what is coming. If this is half as big as you seem to think it is, it's going to be the single

most disruptive thing to hit labor markets ever, given how compressed the time period is in which it will arrive.

It took a long time to lay down electricity. It took a long time to build railroads. A.I. is going to come really quickly.

I think that is basically true, but I want to push back a little bit. I do think we are going to see a dynamic in which it will hit parts of the economy first. It will hit certain firms first. But it will be an uneven distribution across society.

Well, I think it will be uneven, and that's what will be destabilizing about it in part.

Let me give you an example of the kind of thing I'm worried about. There are a lot of 19-year-olds in college right now studying marketing. And there are a lot of marketing jobs that, frankly, A.I. can do perfectly well right now.

As we get better at knowing how to direct A.I., I mean, one of the things that's going to slow this down is simply firm adaptation. But the thing that will happen very quickly is you'll have firms that are built around AI.

It's going to be harder for the big firms to integrate it, but what you're going to have is new entrants who are built from the ground up, where their organization is built around one person overseeing these, like, seven systems. So you might just begin to see triple the unemployment among marketing graduates.

I'm not convinced you'll see that in software engineers. I think A.I. is going to both take a lot of those jobs but also create a lot of those jobs because there's going to be so much more demand for software. But you could see it happening somewhere there.

There are just a lot of jobs that are doing work behind a computer. And as companies absorb machines that can do work behind a computer for you, that will change their hiring.

You must have heard somebody think about this. You guys must have talked about this.

We did talk to economists to try to texture this debate in '23 and '24.

The trend line is even clearer now than it was then. We knew this was not going to be a '23 and '24 question. Frankly, to do anything robust about this was going to require Congress, and that was just not in the cards at all.

So it was more of an intellectual exercise than it was a policy —

But policies begin as intellectual exercises.

Sure, I think that's fair. The advantage to A.I. that is, in some ways, a countervailing force here — though I hear you and mostly agree with your side's argument — is that it will increase the amount of agency for individual people.

So I do think we will be in a world in which the 19-year-old or the 25-year-old will be able to use a system to do things they were not able to do before. And insofar as the thesis we're batting around here is that intelligence will become a little bit more commoditized, what will stand out more in that world is agency and the capacity to do things. And I think that could, in the aggregate, lead to a pretty dynamic economy.

And the economy you're talking about of small firms and a dynamic ecosystem and robust competition, on balance, at an economy scale, is not in itself a bad thing. Where I imagine you and I agree — and maybe Vice President Vance, as well, agrees — is we need to make sure that individual workers and classes of workers are protected in that transition.

I think we should be honest: That's going to be very hard. We have never done that well.

I couldn't agree with you more. In a big way, Donald Trump is president today because we did a shitty job on this with China.

The reason I'm pushing on this is that we have seen this coming for a while. But as I look around, I do not see a lot of useful thinking here. And I grant that we don't know the shape of it. But, at the very least, I would like to see some ideas on the shelf for what we should do if the disruptions are severe.

We are so addicted in this country to an economically useful tale — that our success is in our own hands — that it makes it very hard for us to react with either compassion or realism when workers are displaced for reasons that are not in their own hands — because of global recessions or depressions because of globalization.

There are always some people with the agency and the creativity who become hyperproductive. There's often this sentiment of: Look at them. Why aren't you them?

I'm definitely not saying that.

I know you're not saying that. But that's such an ingrained American way of looking at the economy: You should do some retraining.

Are all these people going to become nurses? There are things that A.I. can't do. How many plumbers do we need? More than we have, actually. But does everybody move into the trades?

What were the intellectual thought exercises that all these smart people at the White House who believe this was coming — what were you saying?

Yes, we were thinking about this question. We knew it was not going to be a question we were going to confront in the president's term. We knew it was a question that you would need Congress to do anything about.

Insofar as what you're expressing here seems to be a deep dissatisfaction with the available answers, I share that. I think a lot of us shared that.

You can get the usual stock answers of a lot of retraining. But I share your sort of doubts of that as the answer. You could probably talk to some Silicon Valley libertarians or tech folks, and they'd say: universal basic income.

I believe, and I think President Biden believes, there's a kind of dignity that work brings. And it doesn't have to be paid work, but there needs to be something that people do each day that gives them meaning.

Insofar as you have a discomfort with where this is going on the labor side, I share that. Even though I don't know the shape of it.

More than that, I guess I have a discomfort with the quality of thinking right now. Sort of across the board, but particularly on the Democratic side — because I have you here as a representative of the past administration.

I have a lot of disagreements with the Trump administration, to say the least. But I do understand the people who say: Look, Elon Musk, David Sacks, Marc Andreessen, JD Vance — at the very highest levels of the Trump administration are people who have spent a lot of time thinking about A.I. and have considered very unusual thoughts about it.

And I think sometimes Democrats are a little bit too institutionally constrained for thinking unusually. So I take your point on the export controls. I take your point on the executive orders and the A.I. Safety Institute.

But to the extent that Democrats imagine themselves to be the party of the working class, and to the extent we've been talking for years about the possibility of A.I.-driven displacements, it's true that when things happen, you need Congress. But you also need thinking that becomes policies that Congress implements.

So I guess I'm trying to push. Was this not being talked about? There were no meetings? You guys didn't have Claude, Anthropic's A.I. assistant, write up a brief of options?

Well, we definitely didn't have Claude write up a brief because we had to get over government use of A.I.

See, but that is itself slightly damning.

Ezra, I agree that the government has to be more forward-leaning on basically all of these dimensions. It was my job to push the government to do that. And I think on things like government use of A.I., we made some progress.

So I don't think anyone from the Biden administration, least of all me, is coming out and saying: We solved it.

What we're saying is: We were building a foundation for something that was coming that was not going to arrive during our time in office and that the next team would have to, as a matter of American national security — and, in this case, American economic strength and prosperity — address.

This gets to something I find frustrating in the policy conversation about A.I.

You start the conversation about how the most transformative technology — perhaps in human history — is landing in a two- to three-year time frame. And you say: Wow, that seems like a really big deal. What should we do?

That's when things get a little hazy. Maybe we just don't know. But what I've heard you kind of say a bunch of times is: Look, we have done very little to hold this technology back. Everything is voluntary. The only thing we asked was a sharing of safety data.

Now in come the accelerationists. Marc Andreessen has criticized you guys extremely straightforwardly.

Is this policy debate about anything? Is it just the sentiment of the rhetoric? If it's so [expletive] big, but nobody can quite explain what it is we need to do or talk about — except for maybe export chip controls — are we just not thinking creatively enough? Is it just not time? Match the calm, measured tone of this conversation with our starting point.

I think there should be an intellectual humility here. Before you take a policy action, you have to have some understanding of what it is you're doing and why.

So it is entirely intellectually consistent to look at a transformative technology, draw the lines on the graph and say that this is coming pretty soon, without having the 14-point plan of what we need to do in 2027 or 2028.

Chip controls are unique in that this is a robustly good thing that we could do early to buy

the space I talked about before. But I also think that we tried to build institutions, like the A.I. Safety Institute, that would set the new team up, whether it was us or someone else, for success in managing the technology.

Now that it's them, they will have to decide as the technology comes on board how we want to calibrate this under regulation.

What kinds of decisions do you think they will have to make in the next two years?

You mentioned the open-source one. I have a guess where they're going to land on that. But there's an intellectual debate there that is rich. We resolved it in one way by not doing anything. They'll have to decide if they want to keep doing that.

Ultimately, they'll have to answer the question of: What is the relationship between the public sector and the private sector? Is it the case, for example, that the kind of things that are voluntary now with the A.I. Safety Institute will someday become mandatory?

Another key decision is: We tried to get the ball rolling on the use of A.I. for national defense in a way that is consistent with American values. They will have to decide what that continues to look like and if they want to take away some of the safeguards that we put in place to go faster.

So I think there really are a bunch of decisions that they are teed up to make over the next couple of years that we can appreciate are coming down the horizon without me sitting here and saying: I know with certainty what the answer is going to be in 2027.

And then always our final question: What are three books you'd recommend to the audience?

One of the books is "The Structure of Scientific Revolutions" by Thomas Kuhn. This is the book that coined the term "paradigm shift," which is basically what we've been talking about throughout this whole conversation — a shift in technology and scientific understanding and its implications for society. I like how Kuhn, in this book, which was written in the 1960s, gives a series of historical examples and theoretical frameworks for thinking about a paradigm shift.

Another book that has been very valuable for me is "Rise of the Machines" by Thomas Rid, which tells the story of how machines that were once the playthings of dorks like me became in the '70s and the '80s things of national security importance. We talked about some of the revolutionary technologies here — the internet, microprocessors and the like — that emerged out of this intersection between national security and tech development. And I think that history should inform the work we do today.

And then the last book is definitely an unusual one, but I think it's vital: "A Swim in the Pond in the Rain" by George Saunders. Saunders is this great essayist and short story writer and novel writer. He teaches Russian literature and, in this book, takes seven Russian literature short stories and gives a literary interpretation of them.

What strikes me about this book is that it's fundamentally the most human endeavor I can think of: He is taking great human short stories, and he's giving a modern interpretation of what those stories mean. And when we talk about the kinds of cognitive tasks that are a long way off for machines, I hope, at some level, this is one of them — that there's something fundamentally human that we alone can do. I'm not sure if that's true, but I hope it's true.

I'll say I had Saunders on the show for that book. It's one of my favorite-ever episodes. People should check it out.

Ben Buchanan, Thank you very much.

Thanks for having me.

You can listen to this conversation by following "The Ezra Klein Show" on NYT Audio App, Apple, Spotify, Amazon Music, YouTube, iHeartRadio or wherever you get your podcasts. View a list of book recommendations from our guests here.

This episode of "The Ezra Klein Show" was produced by Rollin Hu. Fact-checking by Michelle Harris, with Kate Sinclair and Mary Marge Locker. Mixing by Isaac Jones, with Efim Shapiro and Aman Sahota. Our supervising editor is Claire Gordon. The show's production team also includes Elias Isquith, Kristin Lin and Jack McCordick. Original music by Pat McCusker. Audience strategy by Kristina Samulewski and Shannon Busta. The executive producer of New York Times Opinion Audio is Annie-Rose Strasser. Special thanks to Switch and Board Podcast Studio.

The Times is committed to publishing a diversity of letters to the editor. We'd like to hear what you think about this or any of our articles. Here are some tips. And here's our email: letters@nytimes.com.

Follow the New York Times Opinion section on Facebook, Instagram, TikTok, Bluesky, WhatsApp and Threads.

Ezra Klein joined Opinion in 2021. Previously, he was the founder, editor in chief and then editor at large of Vox; the host of the podcast "The Ezra Klein Show"; and the author of "Why We're Polarized." Before that, he was a columnist and editor at The Washington Post, where he founded and led the Wonkblog vertical. He is on Threads.