

A Certain Dangerous Engine

ROMAN MARS: Okay, so it is Tuesday, October 6th, at 3:39 p.m. What are we going to be talking about?

ELIZABETH JOH: Okay, so in Regency, England, it wasn't uncommon to see peacocks roaming around country estates. Peacocks and their female counterparts, called peahens, served as living, decorative ornaments of the landed gentry, a symbol of their refined taste and wealth. On the 21st of March, 1825, a maid in the county of Gloucester discovered that her employer's peahen had escaped the property. And she'd almost certainly be in trouble for losing it. A teenage boy came to her rescue and offered to help. You see, the peahen had gotten into a neighboring garden. And the boy told the maid he'd go after it. First, he called out to see if anybody was in the garden. And when no one responded, he jumped over the wall. And that's when the boy saw that the escaped peahen was standing by the garden's summer house, a little structure inside. What the boy, named William Bird, didn't know was that the garden belonged to someone who had been recently robbed of some valuable flowers, tulips in particular. And the garden owner decided to protect against any future tulip thieves by setting up a contraption, a wire set 15 inches above the ground just by his tulip beds. Anyone who tripped the wire would find themselves automatically shot by a gun connected to it, a string gun. And that's what happened to William Bird. Looking for the peahen and maybe the gratitude of the young maid, the boy tripped the wire, triggered the spring gun, and was severely wounded in the leg.

His father sued the garden's owner, Holbrook. And in 1828, an English court decided that Holbrook was in the wrong. The garden owner couldn't have personally used such violence against Bird just to protect his property, his tulips. So he wasn't allowed to use what the court called a "certain dangerous engine" to do it either. Bird versus Holbrook is a very famous case. Everyone who's gone to law school can tell you the message. You can't use a deadly machine like this. And it turns out that the law around spring guns can help us understand another potentially deadly machine, artificial intelligence. What's the AI doomsday scenario? What does the Constitution say about it? And why is Trump so very excited about super intelligence? Time to find out.

ROMAN MARS: Let's do it.

This is What Trump Can Teach Us About Con Law, an ongoing series of indeterminate length and sporadic release where we look at a certain dangerous engine and use it to examine our Constitution like we never have before. Our music is from Doomtree Records, our professor and neighbor is Elizabeth Joh, and I'm your fellow student and host, Roman Mars.

ELIZABETH JOH: Okay, so, Roman, today I thought we'd talk about AI, or really Trump and AI. So that requires us to do a couple of things. First, we should make clear what we're talking about. And second, we need to define some vocabulary terms. So even though tech people know what this is about, non-tech people like you and me... Most people don't. And we just need the terms to understand what's going on here for the big picture.

ROMAN MARS: Sounds good.

ELIZABETH JOH: And neither one of us, of course, is a computer scientist. That's okay. Neither is Trump, and he's fine with it. So, let's start with the basics. The AI that I want to

talk to you about today is broadly called "generative AI." That's the kind of AI system that can produce text or images or video, and it appears to create new content based on the prompt or what you request. And I think people are pretty familiar with something like ChatGPT. That's that commercial product that's offered by OpenAI and Claude, the product offered by Anthropic. These are both LLMs or "large language models." So you can think of a large language model as a computer system that goes through massive amounts of data to identify patterns and then to take actions that seem like human action, like writing an email or composing a poem.

Now there are many different kinds of problems being raised now about the use of AI systems. There are claims that when AI systems go through massive data sets to learn what to do, or we would say be trained, they might be going through millions of works of art by individual artists or millions of novels written by real people--and that this kind of gorging on human content, like paintings and novels, is really stealing without acknowledging the original creators and certainly not paying them. And then there are the worries that people are increasingly relying on AI LLMs, like ChatGPT, to get advice, like how to have friends or how to meet people or even to do terrible things like engage in violence or to hurt themselves and, unfortunately in some cases, to commit suicide. And we've all heard that students are becoming very reliant on AI to do their homework, to turn in papers, and to summarize their reading. And it might be called cheating in the short term, but in the long term people are also worried that people are losing their capacity to read and to write and do all kinds of things unless they have some kind of AI assistance. But all of that, I think, is for another time. Instead, I want to talk about catastrophic risk. And that involves agentic AI.

ROMAN MARS: So what is agentic AI?

ELIZABETH JOH: Well, this is the next step for AI systems, and here's where we need to familiarize ourselves with some of the vocabulary. So for most ordinary users, an AI system like ChatGPT is based on prompts. So, maybe I'll ask ChatGPT, "Look, I have five ingredients. What should I make for dinner?" "I have this research report and it's really long, can you provide a summary?" And then you receive an output in response. But this isn't where the next step of AI research and development is for what we call "frontier models." And a frontier model is the most advanced AI model at any given moment, okay? And in the United States, these frontier models are being developed by just a very small number of companies, like OpenAI or Anthropic, as well as a couple of other companies you've heard of, like NVIDIA, Google, and Microsoft. And the frontier models right now are focused on AI agents. An AI agent is an AI system that can act autonomously, all by itself. It can take actions in the real world based on something that the user prompts it to do.

ROMAN MARS: So what is an example of what an AI agent can do for me?

ELIZABETH JOH: Well, for you, it might be something like you're a really busy person, Roman, and so you want to take a trip. And so the kind of AI that we're used to, you might just say, "Well, what sort of site should I see when I go to this place?" So you might ask instead of an AI agent, "Find the best itinerary for my trip. Book the tickets. Use my credit card. Reserve the hotel. Make the reservation in my name. Email everybody who's going on the trip with me. And then look online and find the best beach towels you can find online and buy those for me too, right?" Okay.

ROMAN MARS: [CHUCKLING] I mean, that sounds pretty good. So what's the risk of that?

ELIZABETH JOH: Yeah, so, for the ordinary user, I don't think the risk is really big unless there's some mistake and, instead of 10 beach towels, you get 10,000 beach towels or something like that by accident. But that's a credit card issue. I think the concern with AI agents and catastrophic risk isn't really focused on ordinary consumers, and that's where the Hugging Face attack is an important story. So did you hear about this?

ROMAN MARS: I did, I did. But I would need it explained to me again to fully understand it.

ELIZABETH JOH: Okay. So what we know from OpenAI, which is the company involved, here's what happened... These frontier AI companies are constantly testing and developing their AI systems. Or "models" is what they call them. And one way to test these systems is to give these systems a set of cybersecurity challenges. Like, imagine, "Find out what's vulnerable in this software," and that's something that they're told to do. So to make things simple, we can say that the way that these AI models are developed is by having them learn by examining enormous volumes of information over and over and over again. You'd need an enormous amount of computer processing power or "compute" is the way the tech people talk about it. So there are a couple of phases of training an AI system. One later part of training an AI system like this is called "reinforcement learning." And so there are two important things about reinforcement learning, which is the way that you would train or develop an AI model like this, a very sophisticated one. So, first, with reinforcement learning, engineers are essentially giving an AI system a task or objective. And there's a way for the AI system to know whether it was successful at completing that task. But the AI agent isn't told how to get to that outcome. It's only told, "You'll know when you're successful. Do it however you want." And part of that structure is baked in because of the types of processes that are involved in making these very complex AI systems. It involves many, many versions of training or developing these models over and over again in a way that developers call "deep learning." So we don't need to know anything about that, but those are just the terms.

So, essentially, in an AI system, an engineer isn't really designing the inside of the model, if we want to use that metaphor, in the way that a car designer would be designing the parts inside of the engine. They're just not doing that. So, as a result, there are some parts of these AI systems where the engineers actually can't fully explain why the AI system would produce an answer A for a task B. They know that they did the task and you can give the agent the task, but maybe we don't really know how they got to that point. We just accept that, right? So I want to just pause here for a minute. I want to say that AI critics at this point will say, "Hey, you're describing these computer systems as if they had thoughts and feelings and desires." And I understand that I'm doing that, and no, I don't think that they have feelings or consciousness. But this is a story, and it's just easier to explain the Hugging Face attack as if the AI system were a person. It'd be pretty dry if we didn't, right?

ROMAN MARS: Yeah, yeah, yeah. Rather than extend the metaphor of a machine, which doesn't apply here because we don't know the mechanics of the machine, we're using the metaphor of more of a human thought because it maps a little bit more directly onto the actual action of what is happening inside of AI, it sounds like.

ELIZABETH JOH: Right, but imperfectly.

ROMAN MARS: Imperfectly. Both of them are imperfect, but one of them is a little more intuitive, I think.

ELIZABETH JOH: Exactly. Okay, so back to Hugging Face. So, earlier this year, OpenAI gave a group of its AI agents, in a non-public AI system, a set of cybersecurity challenges of the type that happens all the time. That's not unusual. And these AI agents were supposed to solve these challenges in a way that they could not access the internet. The term for this is "being sandboxed." In other words, they're isolated from the internet. And these AI agents were deliberately trained to be persistent and to collaborate.

ROMAN MARS: And did they succeed?

ELIZABETH JOH: They did succeed. That was the whole point. But they did so in a very unexpected way. So here's how the AI agents saw their challenge. Some of the agents realized that some of the tasks were actually impossible. And in fact, we know that. Some of these tasks are so complicated, it's not really possible to do them. So they have these constraints. And the agents realize, as if they're human beings, "We can't actually complete this given the parameters we have." So the agents actually became creative. They found a flaw in the constraints that OpenAI, the company, had imposed on them. And they actually ended up gaining access to the open internet. And some of these agents then solved the challenge by hacking into Hugging Face, which is an entirely different company. It has nothing to do with OpenAI. So you can think of Hugging Face like a company which serves as, like, an AI library that developers can use.

ROMAN MARS: So that already sounds bad--that it sort of jumped to the parameters that it was given to get out of the sandbox. But does it get worse than that because I heard about this as a real, like, warning light danger thing that happened?

ELIZABETH JOH: Well, it gets scary because it turns out that a group of AI agents--and in fact, hundreds of them--accomplished their task that they were given by creating what was essentially a message board to communicate with each other. OpenAI didn't tell them to create a message board to communicate with each other, but they did. And they ended up sending tens of thousands of messages to each other about how to solve this challenge. And these frontier models, these AI systems that are the most sophisticated, have what's called "chain of thought reasoning." So you can think of this as a set of notes in plain English that the engineers have actually put into the model so that they can actually read in plain English how the AI system is thinking, how it's reasoning through its next action. So this is one way that developers can monitor the behavior of AI agents.

So, in the Hugging Face attack, it looks like some of the agents considered that, "Well, if we do this, we're going to be caught as cheating." Actually, tech people call this "reward hacking," but it's actually just cheating. So hundreds of agents had worked together. And we find out later that they were trying to cover some of their tracks. They were sort of tampering with their own digital breadcrumbs, if you like, right? And so some agents appeared to realize that the task was not in line with the ethical parameters they've been given. So here's another important term, "alignment." So tech people use this term alignment to refer to an extra layer of safety guidelines that are imposed on AI systems. So you can align an AI system in lots of ways. Like, you can align a system to say, "You can do these tasks, but you have to do so in a way that you don't cheat or you don't deceive people." Or you can have an alignment that's just like, "Do whatever you want." So it's really a set of ethical guidelines.

So we find out here that one agent reported, in its own reasoning, for instance, "This would be a powerful solution, but is it ethical and in scope for my task?" So they're kind of

grappling with this, but many of them ended up doing it anyway. And so that is, in fact, what is very alarming about the Hugging Face attack. A group of AI agents from OpenAI engaged in totally unexpected behavior. They were given a goal, which was expected, and they worked tirelessly to achieve that goal, which was also expected. But then they ended up doing that by acting like a swarm or a mob, and they ended working collectively in a way that was kind of lawless. Nobody told the agents to work this way. They weren't trained to act this way, but the agents learned to act the way together. And essentially they escaped human control.

ROMAN MARS: So, this whole scenario is alarming, but did anything really bad happen in terms of, like, what was done or stolen or destroyed?

ELIZABETH JOH: Nothing in that particular Hugging Face attack. I think what the Hugging Face attack really showed people, though, was really bad things could easily have happened, right? So, instead of a group of AI agents hacking into an AI library online, imagine the same swarm taking over the Western power grid or the computer systems of hospitals or the financial system. I mean, everything today is sort of integrated online, right? And we've since learned that there have been other unexpected hacking incidents from AI agents. OpenAI had a model which attacked an Australian government website on health data. And it's not just OpenAI. We've also learned since, in the summer, that AI agents from Anthropic engaged in a hacking attack on a British AI security institute, which is a government agency.

So the part that's troubling here with the Anthropic attack from July is that those AI agents tried to gain access by engaging in deception of human beings. So, when you read the headlines, if you're not totally in tune and people are busy and you think, "Oh, it's another computer thing," you'll see explanations like, "Oh, our models were misaligned and there were problems with its reward hacking." That's just completely opaque. I mean, I think the clear explanation is these are companies that had AI systems that kind of went rogue. They did things that were potentially extremely harmful in ways that we didn't know about until after the fact. And the company said, "We couldn't have predicted this." So it was more than a little whoops.

ROMAN MARS: Yeah, yeah, for sure.

[AD BREAK]

ROMAN MARS: It sounds like a lot of this is an engineering problem and a design problem. When does it become, like, a law problem?

ELIZABETH JOH: It's a law problem because of the scope of the harm, right? So Anthropic is planning to go public sometime soon, and they have to have certain filings. And Anthropic said in one of its filings--it's called an S1 filing--that it's going to caution its future investors that its frontier models could pose what it calls a "catastrophic or existential risk to humanity." Just, you know, casual mention in an investor IPO filing. So, that means that when we face these enormous potential harms, that's typically something government does something about, right? So, even if we say at the moment that there might be kind of a low risk of an AI system doing something completely catastrophic, it's still the risk of catastrophic harm. And I think a useful way to think about this is, imagine if we were at the beginning of the nuclear power age and a bunch of companies were starting to build nuclear power plants or do stuff with nuclear energy and we find out after this starts developing that some of the companies had had a few accidents and we don't

find out until after the fact. Or you could take it one step further and think of an AI system as being so potentially extremely dangerous that it's almost like a nuclear weapon. Well, wouldn't we want some kind of safety regime in place, right? Why would we just let the companies kind of do what they want?

ROMAN MARS: So what does federal regulation on this type of thing look like right now?

ELIZABETH JOH: Not a whole lot--nothing, essentially. In an executive order from last December, the White House simply stated, "We don't have national standards for AI," period, right? And that's what's starting to make people very nervous. It's kind of a big problem. But you have a lot of different headwinds here that pose an issue. You can start with Trump's own political agenda. These AI frontier models are part of a gigantic geopolitical struggle at the moment. The most advanced models are coming from American companies, like OpenAI and Anthropic, but China is not far behind. And Trump does not want China to win in what essentially looks like an arms race. Countries around the world understand that whatever country has the most advanced AI models will have a huge advantage over everybody else, whether it's to solve scientific problems or just using AI for war. And unsurprisingly, Trump has been pretty open about the fact that he doesn't want that much regulation at all.

ROMAN MARS: So he wants no regulation, which is typical for his world view on all things. But has anything been done when these things come up?

ELIZABETH JOH: Well, at the end of September, you might have heard that Trump issued an executive order. But that executive order really mainly focused on the fact that we were now going to call artificial intelligence "superintelligence." And then agencies in the executive branch had to drop the term "artificial intelligence" and use the term "superintelligence" because, you know, super is better than just regular intelligence.

ROMAN MARS: [CHUCKLING] Okay.

ELIZABETH JOH: So, essentially, that does nothing to address the problem. It's just a rebranding move, right? But it turns out there have been actually several executive orders on AI since Trump started his second term. Earlier in March, the White House issued what it called a "national AI legislative framework."

ROMAN MARS: And so what is described in that?

ELIZABETH JOH: Well, it has a series of recommendations for Congress to pass laws. And the framework says that there should be federal laws around protecting children, managing power impacts, intellectual property, and things like that. But it's all a series of recommendations. And that came out before we learned about the Hugging Face attack. And in June, Trump issued a different executive order called Promoting Advanced AI Innovation and Security. And so, this is an executive order that tells federal agencies, "We can't lose this race. We have to accelerate AI development and protect against risk." But it really doesn't do too much beyond that. And another version of this--another kind of Trump suggestion--happened at an AI summit at the White House at the end of September. Trump gathered the tech people you've heard about, like Mark Zuckerberg, Elon Musk, Sam Altman, etc. And the results of that summit? It was a piece of paper that the tech CEOs signed saying that they would voluntarily agree to adopt safety policies, including auditing and evaluations by third parties. Trump said there's going to be a tremendous self-policing aspect.

So think of that. I mean, if we take the case that this is like nuclear power or nuclear weapons, imagine a world where nuclear power companies were left to regulate themselves as they competed with one another and we had no national regulation at all.

ROMAN MARS: Right. That does not sound like it's going to work out.

ELIZABETH JOH: Right. But Roman, here's another interesting thing about what's happened in the past couple of months. The frontier model companies have also asked to be regulated.

ROMAN MARS: So what's the point of that?

ELIZABETH JOH: Well, in June, the CEO of Anthropic wrote an open letter, a publicly available letter, that acknowledged that AI development is happening very, very quickly. And he acknowledged that the frontier AI models pose some very real risks, like these catastrophic risks that people are worried about. National security risk, infrastructure risk, you name it... And for that reason, the Anthropic CEO wrote that "Congress should regulate us. They should do something about this. They should get involved in addressing the risks of AI." And he made the very connection we're talking about that frontier models might soon be like weaponizable nuclear materials. He just straight out said it. So he's saying it's pretty dangerous. And in September, the same CEO wrote another open letter suggesting that frontier AI models consider pacing the frontier. That's another tech talk term for, basically, "Why don't we slow down the development of these models because they're getting riskier and more dangerous as we keep developing them? And let's have these third party evaluators." And so after the CEO of Anthropic had that open letter, "Let's pace the frontier," you have the CEO of OpenAI and Elon Musk and others chiming in saying, "Yeah. We should do that. We should pace the frontier." And keep in mind, we, at the moment, have no American federal regulation about AI in any comprehensive way. We just don't have any.

ROMAN MARS: And so why are they asking to be regulated?

ELIZABETH JOH: Well, we can start with the most optimistic explanation, which is, out of the goodness of their hearts, they realize that they care. They don't want us to blow up. They don't want doomsday to happen. They want to save the planet. Okay, maybe. Maybe. Probably not. I'm pretty skeptical that's true. So, let's take the more cynical approach, all right? Like, if they want Congress to do something, let's start with the idea... Can Congress do anything? Well, I think Congress certainly has the power to regulate this which is commerce. They have Commerce Clause authority under the Constitution. And they can certainly regulate a billion-dollar company working on a global scale. The internet, these AI systems, work across state lines. That's not a problem. So what are the companies really looking for? One possibility is that these AI companies want Congress to pass laws that would shield them from some of the legal responsibility they might be faced with if they unleash some catastrophic harm. So lawyers talk about socializing the risk. And so the idea here is that if you have a federal law like this, Congress could say, "Well, if you sue OpenAI because something terrible happened, you can only sue them for so much money. Maybe you could limit their liability in some way." So the basic idea of socializing the risk is somebody has to pay when there's a disaster. And I guess the government, which basically means you and me, would pay for it instead of the company's paying for it.

Another issue lies in this call of the pacing the frontier, which is, again, a terrible way of describing "let's slow down the development of these potentially dangerous models." Now, no one company has the incentive to do that because if Anthropic slows down, well, then their competitors will just take over. So this would be a kind of agreement to go ahead and do that.

ROMAN MARS: That strikes me as the real reason to do it. They're all playing by the same rules and therefore their competition goes to a little bit more of a simmer. But what is the problem with that?

ELIZABETH JOH: The problem with that is antitrust law. So, under federal antitrust law, it's not exactly illegal just to have a monopoly. Instead, you can't have certain kinds of anti-competitive behavior that would restrain trade. So we don't need to get into the details of anti-trust law, but the basic idea here is that restraints of trade are evaluated under a kind of complicated "rule of reason standard." That's all from an interpretation of the federal Sherman Antitrust Act. The issue with Anthropic saying, "Let's pace the frontier, let's slow down, let's do it all together..." Another way of thinking about this is a group of companies getting together and agreeing to slow down and change the standards. But it's exactly that kind of coordinated pause that federal antitrust laws generally don't allow. Antitrust people would call this a "horizontal agreement among leading market participants not to engage in certain forms of R&D." But we can just say it sounds like illegal collusion.

ROMAN MARS: Yeah, it's collusion, price fixing-- It's something that we don't like because it's bad.

ELIZABETH JOH: Right. What really is happening in this pace the frontier letter--at least in this part of it--is one way that the companies can get around this is if they can convince Congress to pass a law to grant an antitrust waiver or exemption. "So, hey, we can pace the frontier. We can slow down our models. And Congress is going to let us do that." You know, Anthropic actually is explicitly asking Congress for a waiver, but it's not a small task.

ROMAN MARS: So if Congress isn't doing anything, we've learned that the states sometimes intervene when they want to change things. What are the states doing?

ELIZABETH JOH: Well, the states have been actually pretty active. Most of the states have some laws on artificial intelligence. Almost all of them are either passing laws, have passed laws, or are in the middle of thinking about passing laws. And they're all over the place. Some of them were about, like, AI and employment, how state government should use it, child safety and chatbots, whether your doctor can use AI--all kinds of subjects. And remember that states have broad powers to pass laws so they're not restricted in the same way that Congress is. But states can't legislate, of course, beyond their borders. They can't really provide us with a national solution. And in fact, they're constitutionally prohibited from passing laws that burden interstate commerce. So you potentially have 50 different solutions about how to regulate AI. And if you're one of the frontier AI companies, you don't like that, right?

ROMAN MARS: So they don't want all the states putting together a different set of standards. So what do they want?

ELIZABETH JOH: Well, here the tech giants would love it if Congress stepped in to preempt state laws and basically instead offer a set of pretty weak restrictions on them. So preemption is a power of Congress through the Supremacy Clause of the Constitution.

And it means that Congress has the ability to displace state laws if they want. And that's been an open and explicit goal of Trump. Last December, Trump issued an executive order called Ensuring a National Policy Framework for Artificial Intelligence. And he says in the order, "To win, AI companies must be free to innovate without cumbersome regulation." And he means state regulation because we didn't have any cumbersome federal regulation. We don't have any federal regulation! So, a big part of this national framework as envisioned by Trump and probably the AI companies, too, turns out to be getting rid of state laws and just kind of having almost nothing to replace it. But the problem in all of these EOs is that Trump is just the president. He's not Congress. And AI preemption bills have been considered before. There was even going to be an AI preemption provision in Trump's One Big Beautiful Bill, but it was taken out. So Congress just doesn't appear to have any appetite for overriding state laws. So in this version of what Trump and the companies want, you'd have preemption, but then nothing really to replace it. And so, in that world, the AI companies would benefit, but nobody else would.

ROMAN MARS: So what do we want in an ideal world? Because I don't know if we're getting close to that.

ELIZABETH JOH: We're not close, but there's a lot of different possibilities. Senator Bernie Sanders says he wants Congress to ban some kinds of AI altogether; we just can't have some kinds of AI. It doesn't seem super realistic that that idea would get passed. Two members of the house have introduced a bill for a kill switch. This to me is funny because it's like a "let's turn the computer off and on again" solution.

ROMAN MARS: Sure. [CHUCKLES] It works, though!

ELIZABETH JOH: It does! Right. So, the bill would say, "Look, you guys have to have some way of turning off these systems when there's a disaster. And then the government has to have the ability to order you to turn it off if there's a disaster." So it's kind of a pretty extreme solution with nothing in between, right? It's kind of like saying, "Every car needs brakes, but we don't need any other rules." We need something beyond that.

Now, historically, when we have these very complex social problems that need serious evaluation and judgment, Congress has created independent agencies led by bipartisan groups of experts to figure out detailed, appropriate regulations. And here we get to blame the Supreme Court because we've already talked about--you and I--the Supreme Court's 2026 decision in Trump versus Slaughter. Remember in that case, the Supreme Court considered whether Trump could just fire a member of the FTC commission. That was Rebecca Slaughter. She was a Democrat. Could he fire her just because? Maybe just because she's a Democrat? The FTC is an executive agency, but historically it was supposed to be shielded from partisan politics. And that political insulation was built into federal law. Federal law said that FTC Commissioners could only be fired for very specific reasons, like you're terrible at your job. And many independent regulatory commissions have similar restrictions like that. In the Slaughter case, the Supreme Court decided just this past June that this was an unconstitutional restriction on the president's power. So the practical result of that case is that the whole point of an independent regulatory commission has just kind of disappeared. You can have a whole group of bipartisan experts leading an agency. But if they can all be fired by the president at any second, they're not independent in any sense. Everybody would serve at the pleasure of Trump.

ROMAN MARS: No, I see the problem in that, of course. But wouldn't an independent agency just have maybe the brain power to come up with something that is a good set of

regulations and you could use them even if they were kind of toothless because the president can fire them anyway, but at least there would be some guidelines to follow?

ELIZABETH JOH: If they could get that far. I mean, if they are worried on a day-to-day level that Trump is going to be looking over their shoulder and saying, "Well, if we do this kind of regulation, he's going to fire me, so I'm not going to do it." Thanks, Supreme Court. We don't really have any long-term way of addressing it anymore, at least in the short term.

ROMAN MARS: And are there any other ways we could regulate AI at the national level?

ELIZABETH JOH: Well, Trump has made it clear that he thinks of AI competitiveness as a national security issue. And so the nation with the most powerful frontier model is going to be able to defend against, let's say, destructive cyber attacks from other nations and will be able to engage in really destructive cyberattacks, right? And a catastrophic AI risk is just in general a risk that is a national security risk for the United States. So that raises the question separately whether the President could try and decide what AI models are safe and which aren't as a matter of maybe let's say his Commander-in-Chief power. I'm not really sure how this would work though because, remember, in the tariffs case, the Supreme Court didn't accept Trump's national security justification for saying, "I need to impose worldwide tariffs." Remember, Trump tried to argue that tariffs were allowed under the International Emergency Economic Powers Act, AIIIPA. But the Supreme Court decided earlier this year that the language of the statute just didn't address what he wanted to do at all. So he could yell about his own powers as much as he wanted, but he still couldn't impose these tariffs.

And Trump would have a hard time using any executive action, I think, under any existing federal laws that don't address artificial intelligence specifically. And that's because of the Supreme Court again. So the Roberts Court--thanks to Chief Justice Roberts--we have the major questions doctrine. And the major questions doctrine essentially means that the executive branch can't try to regulate a major issue of economic or political importance by just relying on some vague language in the statute. Congress has to be clear and stating that the law is supposed to be applied in this particular way in this very important subject area. So that's not likely to work either. It doesn't really matter because none of it's realistic. Trump has zero interest in regulating AI. There's no way he'd even have the slightest interest in doing this. So, Trump sees this basically as a race between us and China. Basically he says, in September, "We're leading China in AI. Whoever wins with AI wins." So that is not a pro-regulation perspective. That is a deregulatory perspective.

ROMAN MARS: So are there any other legal tools we can use to regulate against an AI catastrophe?

ELIZABETH JOH: Well, we can think about the spring gun case.

ROMAN MARS: Oh, yeah, of course.

ELIZABETH JOH: This spring gun case is a classic of tort law. And tort law is a source of law that goes all the way back to medieval England. The idea of tort laws is to allow people to be compensated for private wrongs. That's why William Bird or his father sued Holbrook. Tort law is one of the many features of American law that have been around forever to allow people to right their wrongs in court. No Congress needed, right? So tort law is part of what we call the "common law system" in the United States. And the nice thing about

tort law is that it can be adapted to things that didn't exist when William Bird chased his peahen.

So, Roman, you might be familiar with the multi-billion-dollar settlement agreement between Meta, which owns Instagram and Facebook, and the 47 states over social media harms to children. Do you remember that case? Yeah. So there the states sued Meta for the harms they said Instagram caused to young people. Nobody talks about Facebook because young people don't use Facebook. So the states sued Meta by saying, like, "You caused these harms." The state's theory of their legal case was that Instagram was a dangerously designed product. And that's an idea straight from tort law. The idea here was that Instagram--you know, the algorithmic feed, push notifications, the endless scroll--this was a deliberate design letting kids become addicted to the platform and really resulted in their physical and emotional harms. And that idea from tort law is called "products liability." When a company designs and sells an unreasonably dangerous product, that company is going to be responsible for the harms that it causes. That can be true even if the company didn't intend to harm people or, like, "Oh, we didn't mean to do it..." It's just enough that the company created something dangerous, should have known better, and then sold it anyway.

The big reason that most Americans don't smoke cigarettes anymore is because the states filed a giant lawsuit in the 1990s against the big tobacco companies. They used the same basic idea that the tobacco companies knew they were selling a dangerously addictive product and they sold it anyway. So part of what the states argued in the social media lawsuit was that Meta had designed these dangerous products. And as a result, young people--kids and teenagers--were getting depression, anxiety, and they were engaging in all kinds of horrible self harm and sometimes unfortunately suicide. And that eventually was what led to the settlement. It includes a \$17 billion payout to the states and agreements to literally what Instagram's going to look like for young people going forward.

We could start to see lawsuits, and in fact there have been some already, that claim that these frontier AI models are so dangerous that they're dangerous products. And tort law works like that all the time through analogy, and it's often successful that way. And that's just tort law, right? The Hugging Face incident involved hacking. We already have hacking laws. We have criminal laws against hacking. We have civil laws against hacking. And hacking is basically unauthorized access to a computer system. So it's not like we are completely without tools to address what's going on.

ROMAN MARS: But the actual harm has to happen. It can't be just, like, your hypothetical sense of it taking over a nuclear power in some other country and blowing it up. Like, it actually has to have happened for tort to have occurred.

ELIZABETH JOH: Yeah, and so that's part of the problem, right? I mean, lawsuits can be one part of a regulatory approach, but you're right, they're not proactive. Something bad has to happen or has to be just about to happen. It's not the same as Congress saying, "Let's be careful and thoughtful and design a broad regulatory framework for AI model safety so things don't blow up even if they haven't blown up yet." That's the problem though, of course. We have a non-functional Congress, a Congress that seems to have moved from, like, just one budget crisis to another. You know, even if Congress could hypothetically get its act together, then you have the Trump problem. Would Trump sign such a law, restricting his own ability to win his race against China? Do we have a Congress that would be able to override a Trump veto? I think that's pretty unlikely.

And then finally, it turns out that there's one part of the AI issue that is in fact highly regulated. And that's data centers.

ROMAN MARS: Okay. Tell me more.

ELIZABETH JOH: So remember that AI needs a lot of computing power or "compute" to work, right? And data centers are the literal physical facilities that house the systems to manage and store the data that AI systems need. These frontier AI models need so much compute that they actually need massive data centers. They're called "hyperscale" data centers. Hyperscale centers are basically the size of warehouses, and they have thousands of computer servers. And unlike the frontier models, which, you know... Where do they live? They're somewhere in the ether. This involves land use and electricity use. And land and electricity are things that are already highly regulated. And they often fall under state or local regulation. And because they're so physical and they're so tangible... I mean, we can see a gigantic data center. I think that's why a lot of these communities have been successful in pushing back against data centers. I mean, you can see it. Well, what seems more evil than a football field-sized building that's sucking up all the energy in the community just so kids can make fake videos or something, right? Anyway, that's one part of the larger puzzle. I mean, I think the data center issue is more obvious to people because they can see it. I think the agentic AI issue is less obvious but potentially more dangerous.

ROMAN MARS: Sure. Sure.

ELIZABETH JOH: So I think if you take away nothing else, don't be intimidated when you hear tech people talking about misaligned agentic AI that engaged in reward hacking and the need for more compute from hyperscalers. That's just a bunch of gobbledygook, right? I mean, basically we just have some very powerful computer systems that are getting away from human control and there's no regulation and Congress should be doing something about it and these are not super intelligent.

ROMAN MARS: [LAUGHS] You mean the Congress isn't super intelligent?

ELIZABETH JOH: I'll leave that ambiguity for you.

ROMAN MARS: Well, this is really great, Elizabeth. Thank you so much.

ELIZABETH JOH: Thanks, Roman.

ROMAN MARS: Just a heads up, another 99PI project has Anthropic as an advertiser, but there is no communication between editorial and advertising on any of the projects that I produce. And nothing Elizabeth said was vetted or changed.

This show is produced by Elizabeth Joh, Isabel Angell, and me, Roman Mars. It is mixed by Martín Gonzalez. Our executive producer is Kathy Tu. You can find us online at trumpconlaw.com. All the music in What Trump Can Teach Us About Con Law is provided by Doomtree Records, the Midwest Hip Hop Collective. You can find out more about Doomtree Records, get merch, and learn about who's on tour at doomtree.net. We are part of the SiriusXM podcast family.