PLEASE NOTE: This transcript has been auto-generated and has not been fully proofed by ISEOF. If you have any questions please reach out to us at info@thegreatsimplification.com.

[00:00:00] **Connor Leahy:** What I personally am most concerned about are extinction risks. If we have systems or things or species that are more intelligent than us and they are not wise, kind, often called aligned with humanity, well then I think humanity just doesn't have a long future. If there was a more intelligent species on the planet and that species wanted our resources, who would win?

[00:00:22] And I expect the more intelligent one wins.

[00:00:28] **Nate Hagens:** Today I'm pleased to be joined by Connor Lahey, an expert in AI development and a leading advocate for awareness and regulation of the existential risks that unfettered artificial intelligence presents for society and our future. Connor Lahey is the founder and CEO of conjecture, who is working on aligning artificial intelligence systems by building infrastructure that allows for the creation of scalable.

[00:00:53] Auditable and controllable ais. Previously, Connor co-founded a Luther ai, which was one of the earliest and most successful open source, large language model communities, as well as a home for early discussions on the risks of those same AI systems. If you've ever wondered about some of the basics behind how artificial intelligence software is created, or what exactly about that process makes AI innovation both rapid and risky, this conversation with Connor provides a fantastic primer on these topics, although it is not for the faint of heart.

[00:01:31] This episode also compliments our previous content on the discussions of ai, including the recent episode with Zach Stein on artificial intelligence and education. With that, please welcome Connor Leahy. Connor Leahy, great to meet

you. Thanks for driving me on the show. So, um, as longtime viewers of this show, no, I am no AI expert.

[00:01:55] But, as anyone awake and paying attention to the world, AI is part of our future, and our present, whether we like it or not. We've covered the topic of AI on the podcast. I will have, my team link those episodes in the show description, but I invited you on the show because you've become a vocal leader in raising awareness about the risks to humanity and the biosphere from the race to develop artificial general intelligence.

[00:02:23] So, so let's start there. Um, for those who might not be aware yet, or aren't convinced. Start with a definition, brief definition of artificial intelligence, artificial general intelligence, and artificial super intelligence. And from there, well use that as a springboard to explain what you think are some of the worst case scenarios for AI and humanity.

[00:02:45] **Connor Leahy:** Yeah. Start with easy questions. Just define intelligence. yeah. So it already starts with that these terms are contentious. There's no universally agreed bond definition, even for the word ai. Like people disagree quite a lot on what the word AI means, on what intelligent means, on what a GI means, what a I means.

[00:03:04] So instead of dig, digging into this whole, you know, valid controversy, I'm just going to ignore all of that and just make my own definitions. And if people don't like them, that's fine. These are just the words I am using for the sake of this podcast.

[00:03:15] **Nate Hagens:** Please. I do that all the time.

[00:03:18] **Connor Leahy:** Exactly. So for me, AI is generally software that can do things.

[00:03:26] We don't know how to write as code. Very simple definitions. It's like everything humans do or learn or whatever that we don't know how to write as like a formal algorithm. Um, you might be confused. How can there be software that we don't know the algorithm, but still does something? I'm sure we're gonna get it back to that in just a bit.

[00:03:45] A GI, artificial general intelligence is what I would define as something that can do anything a human can do. At least as good as a human can do it.

[00:03:53] **Nate Hagens:** But not play football or dig a trench. Oh yeah. Okay. Yeah. Sorry. I mean

[00:03:59] **Connor Leahy:** in on a computer. Okay. So any intellectual task that like a human could do on a computer, um, AI could do that.

[00:04:06] Obviously. I think such systems could also have, you know, robot bodies or whatever, but it's not necessary. Yeah. I don't think it's required. Some people think it's required. I don't think it's, um, we could talk about that as well. A cuter way of defining a GI is, it's the, it's a thing that has what humans have and chimps don't.

[00:04:25] Humans build nuclear bombs and go to the moon. Chimps don't, despite us sharing 99.9% of our genome having mostly the same brain structure, like down to the lobes and everything. So, but for some reason something happened with our ancestors where suddenly we can build nuclear weapons, chimps can't. So a lot of people expect that there is like something that differentiates from chimps that allows us to do that.

[00:04:49] And, but the truth is we don't actually know what that thing is. A lot of people have theories about what that thing is, whether it's language or the merge

operation or whatever. But the truth is we don't know. no one actually knows. 'cause we don't understand television that well.

[00:05:05] **Nate Hagens:** So the difference between chimps and humans, even though we share most of our DNA, you're making a comparison that between humans and ai, there will be a similar sort of leap

[00:05:15] **Connor Leahy:** I think between AI and a GI.

[00:05:17] **Nate Hagens:** There'll be such a leap. Okay,

[00:05:18] **Connor Leahy:** so like for me, a GI is a thing. if AI is a chimp, then a GI is a human.

[00:05:23] this, kind of how I think about it. Okay, got

[00:05:25] **Nate Hagens:** it. Yeah.

[00:05:25] **Connor Leahy:** So an A GI system was something where you would believe there is nothing it could, that a human could do that it couldn't do, you know, in a reasonable similar timeframe.

[00:05:37] And then there's a SI or artificial super intelligence. So this I define as a system which is more intelligent, more competent at all relevant tasks than all of humanity put together. So this system, it would be more capable than the economy, it would be more capable than all states put together, not just more competent than individual person.

[00:06:00] This is kind of how I tend to use these words now. There's a lot of, you know, valid questions to be asked of like how a GI exist? Can a GI assist? Can a SI exist? I personally think that it's. Overwhelmingly likely that a GI is definitely I mean it's possible 'cause we have an example humans of a system that is

generally intelligent and we have, you know, very powerful systems such as the economy, like the economy can build semiconductors.

[00:06:30] I can't, you can't, no individual human can make, you know, complex semiconductors, but the economy can,

[00:06:37] **Nate Hagens:** does that make sense? It does make sense. But the economy can, as long as we have copper and silicone and international supply chains and peace in the war and all those things,

[00:06:50] **Connor Leahy:** of course, you know, this is a very simplified model.

[00:06:52] Don't think this is like literally, but obviously if you had a system, you know, where the system is made of silicon chips and robots or of humans that can extract or refine or develop new blueprints for semiconductors and whatever, pull up new factories, et cetera, then you can. Expect it'll be able to do these, you know, you there, you can have a system that can do this.

[00:07:14] And so the thing that I am most concerned about is that I think the step from a GI to is very soon and the step from a GI to A SI is very fast. So what I mean by this is that AI has been on an exponential improvement curve. This is unintuitive in many ways. We're quite used to things progressing linearly.

[00:07:37] You know, every year it gets two units better, but there are some exceptions such as Moore's Law, where, you know, every two years or whatever the number of, transistors on a chip doubles. And AI can ride the Moore's law wave, but it can also ride other ways in improvement in algorithms like we've seen just or less like five years, you know, things going from, you can't talk to your computer to, you can talk to your computer and it can solve PhD level math for you and also generate full to realistic videos for you and, you know, manage your inbox.

[00:08:10] That's really fast. It took humans, you know, 3 million years from our ancest, from the first non chimp to the first guy who could open Microsoft office. Right? So this is an extremely fast level development already happening and we have a lot of people including, you know, Nobel Prize winners, Jeffrey Hinton, the CEOs of the major AI companies who themselves say that they think we're going to get to a GI within the next couple of years.

[00:08:36] And this seems plausible and it's something that I think we should take very seriously. And once you get to a GI, I think you get to a SI quite quickly because, well, what if you had say the best program in the world, but they never need to eat. They never need to sleep. They never get tired. They never get bored.

[00:08:54] They can run 24 7 on a data center chip. You can have a hundred thousand of them in parallel every time one of them learns anything. They can immediately copy this knowledge to everyone else. How fast could it something like that? Do science? And my expectation is it could do it really fast. And then so we can get from, we can get, if you had a AI system just as smart as like the smartest scientist and you run it at, you know, let's say a thousand X speed, well that means every day that system could do two years of scientific research.

[00:09:27] That's a lot. So I think you will get from a GI to a SI very quickly. So very quickly we will have systems that are just smarter than us.

[00:09:36] **Nate Hagens:** I'm just gonna show my na naivete, openly to you on a lot of these questions. If it's doing science, two years of, science in a day, like scientists go out and they measure plankton populations in the ocean and they take, pH levels in the blood and all this.

[00:09:55] So it would still need the inputs from actual scientists to analyze Yes.

[00:10:00] **Connor Leahy:** For some things, yes. But there's a pretty clear exception to this rule, which is software. Software can run at the speed of software. Andis are themselves software. So doing AI development itself is a task that a GI could do and could do very quickly.

[00:10:15] It could do it much faster than humans. We are very bottlenecked by the fact that humans only work eight hours a day before they get grumpy, you know?

[00:10:23] **Nate Hagens:** So is that what people mean when they say recursive meaning it, it improves itself?

[00:10:28] **Connor Leahy:** That's correct. So there's some, again, contention about the words, but what I mean is for me, when I use the word recursive self-improvement or RSI is the moment when you have a system that is as good as the best human engineer at doing AI research.

[00:10:43] Then you tell it to make a better ai. And then once it makes a better ai, well that better AI can make an even better ai. And now this even better, AI can make an even better AI until, you know it, it'll bottom out at some point. But I think you'll be very far from where we currently are.

[00:10:56] **Nate Hagens:** So what, given all that, given you think artificial general intelligence is quite possible within the next few years and that artificial super intelligence would come soon after that plausibly, what do you see as the worst case scenario for, AI and humanity and the biosphere?

[00:11:16] If you want to go that far?

[00:11:18] **Connor Leahy:** I'm not sure it's worth us talking about worst case scenarios because I think the true worst case are actually kind of unlikely. Um, and they're truly horrific. Um, how about

[00:11:26] **Nate Hagens:** middle of the distribution case?

[00:11:28] **Connor Leahy:** This, yeah, exactly. So this I think is much more useful. I mean, to speak quite frankly, let's be clear here.

[00:11:35] If you have something that is smarter than you and you don't control it. It doesn't end well for you

[00:11:41] **Nate Hagens:** unless it's wiser and kinder than me,

[00:11:44] **Connor Leahy:** unless it wants to help you. If it wants to be good to you, if it wants to help you. If it is kind and wise, potentially, that's totally possible. Do you know how to make computers wise?

[00:11:56] No. Yeah. The problem is no one does, and no one is even really trying. some people are, but not really.

[00:12:03] **Nate Hagens:** Even if the computer itself was wise, if it takes its orders from someone who's not wise, then there's a problem.

[00:12:10] **Connor Leahy:** That's a big problem. I think it's even worse than that. I think that what's happening right now is that it's not even that we're giving, I think there's many ways in ways the worlds can go wrong.

[00:12:18] there's kinda three large categories of you know, middle distribution scenarios. One is the kind of um, what a lot of people expect is some, well, some is called the dominant stock trend, which is this idea that, well, if you build a

super intelligence, well then you could tell it what to do, and then you rule the world.

[00:12:37] A lot of people currently believe this. Um, I don't think this is true, but so this already leads to all kinds of dystopias. Is there a single person on this planet who you would trust with full power? I don't like, there's no person or group or entity or even ethical principle that I think should have that much power.

[00:12:56] So this will already go wrong. And this I think is the optimistic scenario. What's happening right now is that we don't know how to control. So earlier I mentioned that with ais or systems that we don't really understand very often, so it's very important to understand that ais as they exist today, which is quite different from like how they were done in like the eighties or the seventies, is they're based on neur methods.

[00:13:22] **Nate Hagens:** Ais were done in the seventies and eighties.

[00:13:24] **Connor Leahy:** Yes. Actually, the term was coined I think in the late fifties.

[00:13:27] **Nate Hagens:** They just didn't have the compute or the complexity of today's models.

[00:13:31] **Connor Leahy:** No. And used to mean a very different thing. Okay. So there's a really funny historical artifact from what's called the Dartmouth Conference, which is where kind of the word AI first started.

[00:13:41] Where, um, so this is like the sixties, and they were like, well, we think that we can make significant progress on, you know, analyzing images and generating, you know, good speech over a summer with about 10 students. They were slightly off. Slightly off. So they lack compute. They lacked a lot of methods.

[00:14:02] Neural networks kind of existed, but not really. And so at this time there was a different paradigm sometimes called good old fashioned AI or gofi, which was more based on logic. So it was more like you would code logic, logical rules that your system would follow. Um, this had some useful properties such as it made them much easier to understand 'cause you could follow the logic trains that these systems would use.

[00:14:26] But they were very brittle, very brittle. They were very hard to make and they just didn't really work very well for the most part. You know, sorry to the gofi researchers out there, um, most of them just didn't work very well. There were some applications, of course, so the real revolution in what we call modern AI or deep learning kind of really started like the early nineties or the late eighties with a invention of a algorithm called back prop.

[00:14:49] This is an algorithm that you can use with neural networks to teach them things. The exact mathematical details don't really matter, but basically it's a machine learning algorithm. This is different from you sit down and you write down logic or you write down code. It's much more, you give them a bunch of examples.

[00:15:07] You give your computer a bunch of examples of what you want them to do, and then the computer figures something out how to do that. And this is how all modern neural networks work. All, modern AI systems work. What ha what this means is, that they're more like grown rather than written like.

[00:15:27] AI today are not lines of instructions that you can read. You know, like a human wrote those. It's more like a blob, of numbers that kind of saw millions or trillions of examples of the problem and absorbed them and learned how to deal with them. But from a human perspective, we have no idea what's going on in these things.

[00:15:48] We know that if we execute those numbers, if we all you know, let them run on our computer, they do a bunch of great things. You know, they talk to us about poetry, they, you know, make funny videos and so on. But we don't really know on the inside what's happening. I.

[00:16:02] **Nate Hagens:** I have, I'm just gonna pop in with some questions as my, um, curiosity pops up.

[00:16:07] So, so I've got, I haven't used AI much. I use chat, GPT or Claude, um, to do a book summary on someone's research or things like that. But when I have a Google Chrome or whatever browser open and I'm accessing the internet, but then I have clawed or chat GPT loaded on one of the browsers, and I ask it a question and it has to go somewhere and use compute to give me the, all the iterations that it researched on my behalf.

[00:16:39] Is that an additional draw and energy somewhere in the world when I ask a, a chat GPT browser on my Google Chrome?

[00:16:47] **Connor Leahy:** So it works very similar to other forms of software. So if you access a website, what your computer does is it pings a data center. Somewhere in the world, the data center is running, you know, hundreds and thousands of computers that are running the software of the website you're visiting.

[00:17:01] And it returns to you what you're supposed to see. And then every time, you know, you click a button or you submit something or you want new photos or whatever, the server that you're in contact with will give you that additional information. You know, it will retrieve it from its database or it will calculate something and it will send it to you so you can, you know, your computer can then show you.

[00:17:20] Um, AI works in quite a similar way. Um, it is unusually compute heavy compared to other applications. Websites are quite light, you know, they usually don't take that much even so then, give for very large websites like Facebook or something, there's. Just these massive data centers that, 'cause you know, you have millions or billions of people accessing your website at the same time.

[00:17:41] So you need massive amounts of computers to serve everybody. AI is kind of similar. AI is a little bit different in that they use very special chips called GPUs or graphics processing units, which are a bit different from the CPUs that most of our computers, you know, most of our computer tasks are done with, um, 'cause they are kind of focused on doing this kind of like blob calculations for the neural network.

[00:18:03] Um, these are very energy hungry. They're extremely energy hungry compared to other forms of chips. They often take on the order of, you know, two to four times as much energy depending on the exact chip and so on. And often for the very, very big AI systems such as the systems, you know, like open like chat, EPT and so on, you often need Ung got the amounts of these chips to both make these ais. So there's kind of two steps. There's the making of the AI or the growing stage, and then there's the deploying or the using of the ai. This is often called training and inference. In the training phase, you is when you feed the ai, all the data, you, teach it, all this stuff.

[00:18:42] This is ridiculously compute heavy. You need massive supercomputers, you know, that can take, you know, megawatts, gigawatts of energy to run these things. You know, kind of similar to the supercomputers we use for like physics simulations and stuff like that. Like kind of like very similar thing. And we run those for, you know, months at a time to build one ai, like one big ai.

[00:19:04] And then once you have the big blob, then you have to, then you do inference. So you expose it to customers so customers can send a queries and get a response from it. Do

[00:19:15] **Nate Hagens:** you ever. Just get this deja vu, wide boundary, shuttering sort of sense that this all is like our modern equivalent of the stone heads on Easter Island.

[00:19:29] **Connor Leahy:** Um, well, at least our stones actually talk back to us.

[00:19:34] **Nate Hagens:** So, so back to my question then. I'll let you get back to your main point. Um, when I ask a Claude, a question, is that in the world, burning more energy than if I ask Google the same question?

[00:19:48] **Connor Leahy:** Probably, yeah. I don't know, obviously, but probably, yeah, it's definitely it's, worth keeping in mind that the energy consumption here per user is quite low.

[00:19:59] Um, you know, well, yeah, no, but if everyone all of a

[00:20:02] **Nate Hagens:** sudden uses chat GPT for everything in their lives, that there's gonna be an uptick in Oh, yeah,

[00:20:07] **Connor Leahy:** yeah, Absolutely. Absolutely. I mean, this already happened with websites and data centers and so on. Like data centers nowadays are a massive. So the user of large energy, a convenient thing about data centers is you can locate them in like right next to renewable energy, which is a thing that happens very often.

[00:20:24] Data centers are often located right next to hydroelectric dams or in the middle of deserts and stuff like this, because you can guess where you can get the cheapest energy. So, because it's okay if it's a little bit farther away, you

know, that's not that big of a deal. They don't have, these are very rarely, like in or near cities, they're usually like

[00:20:39] **Nate Hagens:** out in the sticks.

[00:20:40] Does the path between here and a SI require a lot more energy.

[00:20:45] **Connor Leahy:** I expect. So I think the path between where we are now and any future society requires a lot more energy. You know, I, am not an expert at all on energy infrastructure, but for what it's worth, I do think a GI can be built with what is currently on the grid.

[00:21:00] I think we will likely want more because it's faster and more convenient, makes you more money. But I expect just the current amount of compute that exists in the world is like enough, a hundred times

[00:21:10] **Nate Hagens:** over. That was very, helpful, by the way, because I didn't know those things.

[00:21:13] **Connor Leahy:** Yeah, no, please. I think these things are like not obvious and are obvious, often not like kind of assumed but not explained very well.

[00:21:19] Exactly. I think are very important to understand, you know, that AI is different from other software. It's not just another website. It like it is a different technology. It of course, you know, is on computers but has some unique properties that make it like quite different. And if we totally understood Ouris, you know, we knew every line of code, we knew exactly how they worked.

[00:21:39] I would feel a lot better about that. Right. 'cause then because we, you know, okay, to be clear, I don't think we do a great job on cybersecurity right now. It's not like our software is bug free, but at least hypothetically, you know, we can build pretty good software if we really try. With ai, it's much harder.

[00:21:54] So we talked about this like idea of the dominance doctrine, this idea that well if I have a super genius that does everything, I say, well then I can run the world. I can, you know, gain decisive strategic advantage in militarily, geopolitically. And I can use this to enforce my vision of utopia upon the world

[00:22:12] **Nate Hagens:** with just a super genius.

[00:22:14] Wouldn't that super genius also need to have access to power and might.

[00:22:19] **Connor Leahy:** I expect if you have something that is intelligent, persuasive, you know enough, it can develop technologies, it can out outperform the market. It can, you know, gain political power. It can tell North Korea, Hey, I'll give you a super weapon in return.

[00:22:34] I want you to do this for me. You know, whatever. Right? Yeah. Yeah. Okay. Like I think it will probably be much less this sounds kinda like a Hollywood movie. I think it's gonna be way more boring. I think the thing that's probably gonna happen and we're already seeing happen is that they're just gonna make smarter and smarter ais and the eyes are gonna get smarter and smarter, and then they're just gonna.

[00:22:51] Put the AI in charge of the corporations on purpose. they'll just be like, wow, this AI is a better CEO than our CEO. So we're gonna kick out our CEO and let the AI run it, and then we're gonna let it run all our hedge funds and we're gonna let it, you know, give advice to all our politicians. Like we already had like examples of people in the White House and so on using chat GPT for policy advice.

[00:23:10] Right?

[00:23:10] **Nate Hagens:** Isn't there such a thing as, hallucinations, at least in today's, um, level of, ais and isn't that kind of dangerous in itself?

[00:23:19] **Connor Leahy:** I agree. I think these things should not be in charge of any of these things, but those are getting. As it gets smart, like humans hallucinate too. have you asked your buddy like any random factual question, he'll probably make some shit up, you know?

[00:23:31] Right. Um, it happens to me too. Sometimes I misremember or or I thought I remembered something, but it was actually wrong.

[00:23:38] **Nate Hagens:** but a human does that out of self-deception and evolutionary peacock status. But an AI should be at least conceptually able to say, I actually don't have a good answer to that question.

[00:23:50] try again or help retrain me or something. Do they admit they're wrong? or that sort of thing.

[00:23:56] **Connor Leahy:** So this is a really fascinating one. So this brings us into how AI's different from other software. So for example, with gofi, with old ai, this is very much how it worked. It could, it would take its premises and its logic and it would logically deduce.

[00:24:10] And if it didn't know something, it would tell you. It's oh, I, my logic does not contain what you want. But this is not how our AI work. Ouris are much closer to copycats. Than they are to logical reasoners. So the way they're trained, like chat, GPT, Claude et the way they're trained is actually kind of weird.

[00:24:27] The way it works is you take huge piles of just text, books, stories, dialogue, whatever, and you show it the start of the text and then you say, what

do you guess is the next word in the sentence? And then it makes a guess. And then you show it that word and, okay, now guess what is the next word?

[00:24:49] And then you so on. So on. And you do this trillions of times, you are like, just guess what the next word in this sentence is going to be. And for some reason this gives you cha, GBT, clot, et cetera, if you train on the right data. But what this also means is that these things learn to emulate. many things that humans do.

[00:25:06] So they, if you give them, like for example, recently I've been getting a lot of email from crazy people telling me that AI told them that their theory of everything in consciousness and space travel is correct because the AI of course has seen this before and knows it's supposed to say yes, that's good.

[00:25:24] **Nate Hagens:** So they not only emulate human brilliance and science and inference, but also our delusion and self deception. Absolutely. Overconfidence, absolutely macho testosterone. I need to be right for status and blah, blah blah. Yeah,

[00:25:39] **Connor Leahy:** they often do. They often do. And the cra, it's actually even crazier than that because with humans, you know, humans kind of have a personality, right?

[00:25:46] you know, you have a personality, I have a personality. Rightis are kind of even crazier than that. They kind of are like super schizophrenic, where if you give them a slightly different phrase thing, they can just flip their personality entirely. Like sometimes they can be super nice and friendly and like other times they can be super aggressive and crazy.

[00:26:03] And a lot of what these companies do is they put a lot of work in hiding this from the user. So there's a thing called a base model. This is what you

remember, the training process where you crunch all the data and it spits out what is called a base model. Base models are crazy. They don't, there's very few of them that people nowadays have access to.

[00:26:23] Unfortunately, it's really hard to get access to these things. They're kind of like these crazy schizophrenic aliens. They're super smart, but they just regurgitate kind of like everything they learned, all the emotions, all the different characters. it's like really weird. Um, they're very, smart, but they're not, they don't really talk.

[00:26:40] They pretend to be things more like these things then get further modified. Um. Which is sometimes called chat fine tuning where they train. They then further modify these systems to act more like a single entity that you can talk to. This process is not super reliable, so sometimes there's something what's called nowadays, usually called a jailbreak, where you can give certain instructions to a machine, to a AI chat model, which will make it suddenly go crazy or do something it's not supposed to do or totally change its personality or output data.

[00:27:15] It's not supposed to output and stuff like this. All this I know sounds crazy. And the truth is, we have no idea really why these things happen or like how to prevent them. It's all super ad hoc.

[00:27:26] **Nate Hagens:** So is there some, not a kill switch, but um, some secret code words that the developers of chat, GPT, OpenAI or whatever, can put in that, that do something.

[00:27:39] Um, does any human have oversight over this or once the model is out there, it's kind of the AI on its own.

[00:27:46] **Connor Leahy:** So in general, of course, when these things are deployed in a corporate setting and so on, they will be monitoring the logs with

their users as they will with any product, right? you know, if you use any product on the internet, the developers are watching very closely what you're doing and how they can improve their product further.

[00:28:02] So that's definitely a thing where this becomes a problem is with open source. So there are open source AI systems. You can download open source models and run them yourself. And with those, obviously there is zero possibility of overseeing anything going on there. Like it's just out in the world. Anyone can do anything they want, who knows.

[00:28:23] So there's there's some amount of oversight on kind of like the output or like the interface between customer and model. But within the actual AI itself, there's no real oversight. there's no way to bake in an extraction that it will 100% follow. You can try, you can prod them, you can train them in a lot of data that tells them if I say this, do that.

[00:28:48] And that works decently well,

[00:28:50] **Nate Hagens:** but

[00:28:50] **Connor Leahy:** it doesn't

[00:28:50] **Nate Hagens:** work a hundred percent of the

[00:28:51] **Connor Leahy:** time.

[00:28:51] **Nate Hagens:** So, um, get back to your middle of the distribution. what are the big risks here, Connor? I mean, I can think of a dozen, but you're the expert, so please help me understand.

[00:29:01] **Connor Leahy:** I mean, unfortunately there are definitely over a dozen.

[00:29:04] So for, I mean, at least a dozen here, right? So we already talked about the first one, which like, you know, power concentration, dystopia kind of stuff. I do think these risk are real, right? I do know this, miss these at all. I think these are like a very big problem. What I personally am most concerned about are extinction risks.

[00:29:20] If we have systems or things or species that are more intelligent than us and they are not wise, kind, often called aligned with humanity, well then I think humanity just doesn't have a long future. Then I think the future would belong to these things.

[00:29:37] **Nate Hagens:** Please connect the dots on that. I've heard that before.

[00:29:39] Can you just outline the sequence of how that UN potentially could unfold

[00:29:45] **Connor Leahy:** and obviously, I dunno how it exactly unfolds, right, would I. So I can only give a very rough, very rough idea of what happened. The way I think about this is mostly just like if there was a more intelligent species on the planet and that species wanted our resources, who would win?

[00:30:03] And I expect the more intelligent one wins. You know, maybe it'll take a while. Maybe there'll be nice at first, and then later they'll be, they'll betray us. Maybe they'll betray us right away. I don't know. I'm not a smart species, you know, I'm no chess grandma, so I don't know what they would do. I expect what's gonna happen is kind of what's been happening for the last couple years.

[00:30:22] We'll just keep accelerating. These things will get better and better at doing tasks. They'll automate more and more labor, more and more programming. They will get better, better at doing science. They will get better, better at doing diplomacy, at business sales, um, persuasion, um, politics, et

cetera. Over time, it will become a competitive advantage to have these ais run more and more of the economy because they're just more effective than humans.

[00:30:48] They're cheaper, faster, smarter, more sociopathic, um, so they will make more money. So if you replace more and more of your employees with AI systems, you'll make more money. So more people do it more and more politicians will start taking advice from these systems because they're smarter than their advisors.

[00:31:03] So they'll use the AI's advice for how to run their political campaigns. And then as this continues, the information ecosystem will continue to grade. it's already almost impossible to really know what's going on in the world, right? you look at social media, like what's true, what's false? It's like already very hard to determine.

[00:31:21] **Nate Hagens:** What do you think the risks are for job losses and how will AI replace many of today's common desk jobs and what sort of, um, jobs are at most at risk for this? And then I have some follow ups to that.

[00:31:36] **Connor Leahy:** Yeah. So my general opinion on this is kind of informed by my belief that I think AI and a GI is coming very soon and that a SI comes quickly after because what this means is that we can get to a SI long before people have bothered to automate all the jobs, even if it was possible.

[00:31:54] People just maybe haven't gotten around to it yet. You know, firing people takes time. Doing new processes takes time. While going from a Asia to a SI could take a year or less. It could take months. It could maybe only take a week. Who knows, right? who knows how fast those things can think. I don't know.

[00:32:10] So therefore, I think the, answer to your question is all jobs by definition, a GI can do anything a human can do, including, say, develop cheap robotic bodies that can be mass produced.

[00:32:23] **Nate Hagens:** So one of. I think one of the reasons that there are a lot of people that follow this podcast is they know I am flawed.

[00:32:31] I don't have all the answers. I'm hitchhiking a ride generally in the s sapient wise direction, as a human alive in these times. But I'm real and I make mistakes and I tell people my feelings and my thoughts and my opinions. It. How soon could AI replace even me? like in, in a, podcast? Or is that, that not likely?

[00:32:58] **Connor Leahy:** So I think there's two aspects to it. One is like, when could it do the literal craft that you do? come up with questions, ask them in a, you know, charismatic way and have a good conversation. Um. I think we're not, I mean we're, I think we're kind of there already, right? Yeah. I think the thing that makes you good, why people like you is because, you know, you're, you know, we like being with other humans.

[00:33:21] We like hearing other humans we like, et cetera. But in terms of the craft, you know, AI are already like really good at asking questions, organizing interviews, you know, stuff like this.

[00:33:32] **Nate Hagens:** But what I kind of meant there is a deep fake of me where people wouldn't be able to tell it wasn't actually Oh, definitely

[00:33:38] **Connor Leahy:** very close to that.

[00:33:39] I mean, you can already, I say dupe like 50% of the population for sure. and I think that number will keep rising.

[00:33:45] **Nate Hagens:** So here's, um, actually last night I recorded what I call a, frankly, which is my Friday, thing. And I did, um, out of the blue, before even talking to you, I did the eight, archetypes, of human and ai.

[00:33:59] And I, opined that AI is gonna act as a filter, not. For jobs, which was my initial or ongoing concern. But for our identities and some people who live in uncontacted, no internet areas of the world are gonna have their natural human cognitive structure remain intact. But it's gonna be like, I think you say algorithmic cancer is gonna spread around the world.

[00:34:27] And some people, because of their temperament and their insecurities or whatever else are gonna become. Totally, completely addicted to AI and their self, who they are as a human is gonna dissolve away. So I increasingly think that is a big risk. Like we lose what it means to be human because this is gonna be so powerful and so ubiquitous.

[00:34:51] What are your thoughts on that?

[00:34:52] **Connor Leahy:** This is absolutely true, and it's so much worse than you can imagine. like it's already happened. so much of this has already happened. I know how much, for example, you spend time in the third world, like North Africa or stuff like this. Not recently, but I have in my life.

[00:35:07] So, you know, I have friends and so on who have family in these areas of the world and you know, addiction to TikTok and so on, it's like already a problem in the West. But it's so much worse in like many of these areas where there's just like no antibodies to these kind of things. Oh, no. And now with ai, it's been gone into hyperdrive because you know, now you can have characters that you can talk to.

[00:35:31] You know, you can have individualized, personalized friends. Recently Mark Zuckerberg's had an interview. Is that the average? I think he said something like, the median American has three friends but has demands for 15, and this demand can be filled with ai.

[00:35:46] **Nate Hagens:** So not only would you have your favorite friend or girlfriend or boyfriend as an ai personality, but you could have 10 different personalities and call them Fred and Louisa and whoever and ha and interact with them.

[00:36:01] **Connor Leahy:** Absolutely. And there are companies who do that right now, and they make a shit ton of money.

[00:36:05] **Nate Hagens:** Why is it that I invite people on this podcast and in my first conversation, I just get the sense that they are gonna continually inform me that it's worse than that I'm surrounded by. It's worse than that.

[00:36:18] humans, who are all, um, swinging for the fences to make the future better than the default. So thank you for all your work on this. Um. So what else are you, I mean, okay, so getting back to my point, I mean, is there. We're gonna talk about governance and your, work on regulation and, the software hierarchy of these things.

[00:36:43] But is there any resistance, that humans can have in their own life? One of the eight categories, I labeled simply AI Luddites, which is those that understand the things that you are saying. And I had Zach Stein on a couple weeks ago talking about the impact of, AI and education, that we understand that.

[00:37:03] So we use AI when we need to, but we cordone it off with a Chinese wall, that we don't do it beyond X, Y, Z, and then we go out and go out in nature and meditate and cook our own meal and are not using it. Are there protocols?

Are we gonna be able to, um, distance ourselves using discipline or is this like social media and Facebook on crack cocaine and meth at the same time?

[00:37:31] **Connor Leahy:** I think it's like social media and crack cocaine and meth on the same time. That doesn't mean we can't do something about it, but it is like really worth understanding. Just like the fact that social media is unregulated is insane, like it is as addictive as gambling or like hard drugs and yet it is completely legal to, you know, pipe it directly into a two year old's brain 24 7, advertise it, you know, no problem.

[00:37:58] Right? Like we have a mechanism in. Western democracies for handling dangerous pollutants and dangerous addictive substances. And it's called regulation. We regulate gambling well, I mean, has been unregulated recently and has been a huge disaster, especially in, the US but also here in the uk.

[00:38:17] **Nate Hagens:** Wait, there, just real quickly on that, what do you mean?

[00:38:20] Recently it's been a huge disaster. I mean, I'm unaware of the history of gambling and sports books and all that, but What do you mean recently?

[00:38:27] **Connor Leahy:** Oh, recently they've just legalized sports betting. Okay. Which has just been. And, again, another rabbit hole. So there's a whole rabbit hole here where recently in the US many states have legalized sports betting.

[00:38:39] Yeah. And now it's being super optimized by you know, social media apps to like ping you on your phone, get a free, you know, bet right now. Yeah. and not only that,

[00:38:50] **Nate Hagens:** but what I noticed last football season is they want you to bet on the next play, what is the next play gonna be? Exactly. And we've lost the action.

[00:38:58] It's actually a microcosm for this whole conversation 'cause we've lost the inherent, cooperative sport team dynamic of watching the Packers beat the Bears. And we're super focused on our fantasy lineup or the next play, which is a little bit of a microcosm of what AI is splintering our human experience.

[00:39:17] **Connor Leahy:** Absolutely. fundamentally, humans have some just flaws in hard brains are designed. There are just a couple flaws. And the big, one of the big ones is addiction. It's like. Everyone knows addiction is a thing. It can happen to anyone. It can happen for many different reasons. It's very bad for us. Yet once you're in it, it's very hard to get out.

[00:39:38] And it's even if you know it. Even if you know, right? Yeah, exactly. and it can happen for all kinds of things. Right. And having addicted customers is so profitable. It is just so profitable. You know? Have you ever heard the word like wa term whale before in the context of like this? In

[00:39:55] **Nate Hagens:** Bitcoin? I've heard the term whale.

[00:39:56] **Connor Leahy:** Yeah. So in gaming and betting and so on, you know, whale is the word they use for the big spenders. You know, people who, right, right. Who spend a large money and everyone is targeting whales. And like whales usually make up light, you know, 5% of the user base, but make up 99% profit or something like this.

[00:40:14] And so these are people who are, so all these apps are not optimized for, Hey, let's have make everyone have a fun time. You know, doing a little bit of betting for fun. you know, gambling can be fun, right? You know, have poker with

the boys, whatever. Right. You know, it's good fun. You are not the target audience.

[00:40:29] You know, if you have a little bit of fun, you know, putting a couple dollars on a match here and there, you're not the target audience. The target audience is people who get their paycheck and immediately put everything into the next play within four hours. Those are the target audience. And these people are optimizing with like actual clinical psychologists to target, manipulate and groom these people.

[00:40:50] And create more of them. Like it's so cynical.

[00:40:53] **Nate Hagens:** And is that, I mean, Connor is this, so what's the other? Addiction was one of the flaws of the human brain. What was the other?

[00:41:01] **Connor Leahy:** there's a couple other ones. tribalism is a big one. Yeah. You know, in group Outgroup, ingroup. Outgroup, that's a big one. Yeah. You know, there's various like biases in that direction.

[00:41:10] **Nate Hagens:** I mean, is this whole thing that as humans have, ridden up the carbon pulse, which is drawing down all these, fossil and material inputs millions of times faster than Mother Earth built them up. And we're applying that to technology. That technology is more and more faster and faster going right to our brain stem.

[00:41:32] And it's almost like we don't have control. It's like the entire species is now going to be addicted once AI comes fully online.

[00:41:41] **Connor Leahy:** Exactly. And these are the kind of things why I think it's not gonna be a big fight. some people think oh, once the A GI shows up, humanity will ban together and will do the right thing and say, but then it'll be too late.

[00:41:52] And like it's way too late. what are you talking about By then you, it's like we're already like, do we feel like we're in control of social media? Like it's a healthy thing for the planet that we're on top of? Or, like tech companies or oil companies? No, and I think those things are, you know, they're still human.

[00:42:10] You can still regulate them. You can go to their house and serve them, you know, a lawsuit. You can't do that with an ai with an A GI or nevermind. An a SI.

[00:42:17] **Nate Hagens:** Is it too late already?

[00:42:18] **Connor Leahy:** I don't think it is yet, but it will be soon.

[00:42:23] **Nate Hagens:** So, so keep going. Um, the, did you wanna say anything else on my comment that, um, not only will we lose jobs, but we're gonna lose our identities, and that's unfolding with all the different, categories and risks there?

[00:42:37] Yeah,

[00:42:38] **Connor Leahy:** I think all of these things are happening. I think all of these things would happen even without a GII think even if a GI doesn't happen, I think all of these things will still happen. I think we'll still lose un incomprehensible amounts of jobs and therefore people lose a massive amount of economic power.

[00:42:52] Without economic power, they'll often also lose a lot of political power. I think there will be a massive loss in executive capacity in to addiction and, you know, just like super optimize propaganda and stuff like this, like a lot of people are losing their souls, their attention, their political power to Twitter propaganda, to optimize advertising, to pornography, to stuff like this.

[00:43:16] these are all, what's important to understand is that this is adversarial. There is an attack, there is optimization happening to take away agency. It's not

like a natural thing that just happens in the ether. It's, there is a, there are deliberate people whose job is every day to take away your agency as much as possible and to monetize it.

[00:43:36] That's literally what the attention economy is. That's literally what advertising is

[00:43:40] **Nate Hagens:** because it's their job description and it's the goals of the shareholders of the corporation. Exactly. So it's fully approved and sanctioned by our culture. Exactly.

[00:43:49] **Connor Leahy:** Yep. It's fully approved, it's fully legal. It's like the fact that it's, and they like, it's even considered mundane.

[00:43:55] Like when someone says, oh, the a optimize advertising, it sounds so boring, so mundane. You know, but what they're doing is they're optimizing to take away people's attention to something useless.

[00:44:05] **Nate Hagens:** Well, this is, this overlaps with my work, which is that we have outsourced our wisdom to the financial market and have become an unthinking, unseen, energy hungry, mindless economic Superorganism.

[00:44:21] And the more I talk to people like you and Zach Stein, I think that the algorithms are the beating heart of this Superorganism. And maybe let's unpack there. You have coined a term called algorithmic cancer. Can you, explain what that is and why it's relevant?

[00:44:38] **Connor Leahy:** So, yeah. So we can dive right in.

[00:44:39] So kind of as we're talking about this, so there is a sense where when you build software, or when you build an engineering product, well, let's say we're built, you build an airplane, if you don't know how to build airplanes, it will crash

and you will die. So pretty bad idea. So you have to actually understand a lot about airplanes to build a good and safe airplane.

[00:45:02] this is quite hard. You know a lot about aerodynamics, you know a lot about materials, you do a lot of tests. You have to be, you know it, it's a hard thing to do. But eventually we can build planes that fly across Atlantic and don't crash, which is a amazing feature. You know, that we can do this with software.

[00:45:18] There are similar things where as you build software, as it becomes more complex, at some point it tends to break once you don't understand it once. And this happens all the time, especially with like big corporations where they have software that is so big, it is so old. It is made of so many parts that no one actually knows how it works.

[00:45:36] No one actually knows where all the pieces are, and if something breaks, they're stuck. And it can be like almost impossible to get it back running. This is super common. I sometimes call this software senescence. Senescence is the state of you know, when you, when cells stop dividing, when they stop functioning properly, when they age, and there's kind of a thing here as well where they, it stops growing.

[00:46:01] It stops gaining new abilities, and it kind of becomes defective. But there's a strange thing that has happened with AI and not just the most modern ai, it goes back to even earlier forms of AI systems, recommender algorithms from, oh, by the way, I mean, have I mentioned all those TikTok algorithms, all those YouTube algorithms, those are all AI too.

[00:46:22] So this was actually the first major use of ai. A lot of the funding that led to the creation of the current batch of AI actually comes from the profit made by social media companies from using AI to optimize their social media

algorithms. This is where a lot of the funding originally came and all of the science originally came from.

[00:46:40] So there's a direct line between, you know, recommending more slop to your children to. Cha, BT and all these modern systems replacing jobs. So there's a weird thing that happens with ai, which is what we talked about a bit earlier, where it's grown, it's not written. And so this decouples your understanding from the capabilities of the system.

[00:47:02] In the past, if you wanted to build a software system to do something really complicated, you had to know a lot about software and you'd have to understand the problem very well. Otherwise you will fail as happens a lot, right? but with AI there's this crazy thing where we can build a systems that can do crazier and crazier things that we don't understand.

[00:47:23] We can build systems that can solve math problems that we can't solve. We can build systems that can, you know, write Shakespeare in ways. I could never write Shakespeare, you know, that can do all these things. All we need is more data and more computing power, bigger supercomputers. Sure, there are some engineering that goes on there, but it's like.

[00:47:41] Very different from building an airplane. So there's no one who understands all these parts yet it keeps growing and it often grows in ways we don't understand. Like sometimes these weirdnesses are amusing. Like I recently heard an example where Chachi Bt, um, started refusing to speak Croatian because Croatian users kept down voting the Croatian answers.

[00:48:06] So it just stopped talking Croatian and just refuse to talk Croatian anymore, which is quite funny, you know, is it the end of the world? Probably not, but there's more sinister examples. So for example, as I've said earlier, I get a lot

of emails of like crazy people telling me about their theories of quantum and consciousness or whatever, right?

[00:48:27] You know, just like random stuff. And what I found is the style and the amount has drastically changed since an update. Hit chat, GPT, where it made it much more sycophantic. There was an update for a couple days that made it super sycophantic and agree with everything. And that very day I got like literally 10 times the number of schizophrenic emails where they all had screenshots of chat GPT telling them how true their ideas were.

[00:48:54] **Nate Hagens:** Oh dude. That is so like personally scary to me because I've been kind of a public, I, used to run something called the oil drum and I'm social, so I have a huge network and you cannot believe over the last 20 years how many mostly male, almost exclusively male people have some special secret thing that they've discovered.

[00:49:20] That is the answer to all of our problems. And they are a hundred percent confident and charismatic, and I'm sure they're well intended people, but I almost think there's something in our genome with male social primates with technology and a status accordion that they wanna move up. That this is one of the generator functions of our cultural, predicament we're in.

[00:49:44] Because there's tens of millions of these people that are pushing ideas and some of them get funding and it's just disconnected from reality. And you're telling me that AI is gonna turbocharge all that

[00:49:56] **Connor Leahy:** already is. Like I used to, when I get, when I used to get emails from schizophrenic people, it would often be like long rambling word documents that they've written themselves.

[00:50:06] Nowadays, I only get Chachi BT screenshots. It's always them with Chachi BT talking about their delusions and like reifying, their delusions and so on.

[00:50:14] **Nate Hagens:** This makes me ill, it makes me feel ill.

[00:50:17] **Connor Leahy:** Yep. And so I think this is the correct reaction. You are having the correct human reaction to this. Meanwhile, the reaction in Silicon Valley is, wow, look how engaged our customers are.

[00:50:26] Look at our engagement metrics. Look at how many U returning users we have, and this is algorithmic cancer. Like to me, this is agro, this is algorithm pollution, and what is happening here is there is a massive cost being put on society that is not being paid by the people causing the harm. I think it is bad that these schizophrenic people are being harmed.

[00:50:49] Right. I. But who's responsible for this? who's, gonna go to jail for this? No one

[00:50:56] **Nate Hagens:** in the same way that CO2 and the pollution from, I mean, fossil fuels are amazing because they allowed us to do things with machines that we couldn't do with our draft animals or our hands. But the, trade from hand labor to fossil fuels and machines is an analog to the trade from human cognition to AI cognition.

[00:51:20] But the externalities are not included in the prices. Exactly. So we don't include the price, the externalities of CO2 and the fish having to swim poleward in the oceans because there's less oxygen, any of that in our prices when we buy a coffee cup or go on a plane ride. And in the same way, the externalities on our brains, our cognition, our psychological development, our humanity are not included in the technology.

[00:51:45] **Connor Leahy:** Exactly. Would AI or social networks still be profitable? If those externalities were accounted for? I think probably not.

[00:51:54] **Nate Hagens:** Yeah, probably not.

[00:51:55] **Connor Leahy:** Yeah. I think there's a way, like could there be a world where all, you know, CO2 and like pollution externalities are included with our cost and we still have a thriving economy.

[00:52:05] I think it's imaginable, right? It would probably be different from how us ours currently

[00:52:08] **Nate Hagens:** is. Well, that, that's the premise of The Great Simplification. it's not gonna look like this economy and it's gonna be much less material throughput per person. but yeah, I mean, at least conceptually that, that's where I would like to head.

[00:52:23] So, is there something like algorithmic chemotherapy?

[00:52:28] **Connor Leahy:** Yes, but it is not an easy solution and it will destroy the business models of some of the largest corporations in the world because their business models similar to oil companies is dependent. Is only profitable because of unpriced externalities.

[00:52:43] If we held people responsible for the harm that social media causes, I don't think these social media companies would be making this profit. I think they would all fall apart.

[00:52:51] **Nate Hagens:** So let's stop right there. Is it possible that, I know Jonathan Het, who's been on this show before, has recently been, um, tweeting and writing a lot about this study that came out that, I'm sorry I don't have the details on the top of my mind, but you're probably aware of it.

[00:53:09] That showed that cognitive abilities of people in the world have declined over the last decade, like precipitously and the culprit is probably social media and addictive use of screens and such. So on social media alone, let alone ai, let's set AI aside for the moment. Wouldn't a precursor to regulating AI and including the externalities of it in the prices, wouldn't we first have to do that successfully in the social media field?

[00:53:38] **Connor Leahy:** I think it would be a very sensible thing, and we should have done it 30 years ago. I. Yeah, like I think, I mean, so to be clear to anyone in the world who wants to truly regulate social media, you have my full support. Like I think this is an incredibly important thing to work on. It's a thing I put some work into as well.

[00:53:53] I completely agree that if we had successfully, and Healthily found a way to regulate and deal with social media in a healthy way 20, 30 years ago, I think the world would be in a much worse, like much less bad place than it is now. And we would be much better prepared for dealing with ai, but it's basically a massive.

[00:54:12] Pollution crisis, it is still ongoing that we just never dealt with. So, and now we're getting another crisis on top of that

[00:54:19] **Nate Hagens:** with ai, it's actually the biggest pollution crisis. 'cause to solve the pollution crisis on the biosphere, we first have to solve the pollution crisis in our brains.

[00:54:27] **Connor Leahy:** Yeah. I think this is very deeply true.

[00:54:30] I think there's a really deep thing here, right? Where sometimes there's a view that like conspiracy theorists like to have of the world is that the reason the world is bad is because there's like some evil cabal, you know, there's some evil,

you know, whoever they think it is who's in charge and they're making things go poorly.

[00:54:47] And in a sense, I think this is kind of cope. I think it's cope in the sense that it's kind of more comforting to believe that somewhere someone is in control, this is our

[00:54:58] **Nate Hagens:** ingroup Outgroup algorithm E,

[00:54:59] **Connor Leahy:** exactly. so even if they're bad, even if they're evil, at least you know, someone knows what's going on.

[00:55:04] Someone could, you know, could do something if something truly terrible happened. Yeah, but my But the truth is. There are no adults in the room. Agree. No one is in charge. Agree. It's, just chaos. I don't think anyone is winning. I don't even think the social media companies are winning. I think their kids are getting addicted too.

[00:55:19] I think they're suffering as well from the key of their society.

[00:55:22] **Nate Hagens:** So this is exactly what I say in my Superorganism movie and my work is we truly have outsourced our wisdom as a species to the financial markets because every company has to have shareholder. If the CEO all of a sudden gets religion and wants to do what's right for the biosphere and future generations, he or she will be booted out by someone who maximizes profits according to the shareholders.

[00:55:51] Exactly. And it's the same with a fossil fuel company. It's the same with, any company including an AI company. Exactly. So until that changes, um, all of these intentions and well-meaning and wisdom and kindness are. Downstream lower in the decision making hierarchy than this optimized profits.

[00:56:13] Optimized growth, optimized GDP, and so it is in, that sense, it's no one's fault, but it, is also all of our responsibilities once we understand it.

[00:56:26] **Connor Leahy:** Yep. I think it's exactly correct. And this is the deeper like rabbit hole reason for why I think a SI will not be kind to humanity, because humanity isn't building it.

[00:56:35] That thing is that process you described. That's the thing. Building a SI. And what do you think it's gonna do with it?

[00:56:42] **Nate Hagens:** Well, it could say. Look, there are, we're already transgressing six of the nine planetary boundaries. Things are accelerating. There's, a long list of ecological woes. There are 8 billion humans on the planet who consume 19 terawatts of energy continuously.

[00:57:03] What do you do? A SI and the answers are probably not good for humanity.

[00:57:09] **Connor Leahy:** I think that's exactly correct. It doesn't even have to be evil necessarily. It just has to be like efficient. You know, it would just see and logical. And logical, right. So, and you know, no emotion sociopathic, which is what we expect an AI to be.

[00:57:21] It will just be optimizing,

[00:57:22] **Nate Hagens:** but it would also know that it's going to need humans to procure energy for it. And, all the like for a while.

[00:57:29] **Connor Leahy:** For a while. Yeah. Like I think, like the way I think about it personally is I think the point of no return, like game over is quite a while before humans actually go extinct.

[00:57:38] I think we'll probably stick around for a little while. Right? But it's already gonna be too late.

[00:57:43] **Nate Hagens:** I'm gonna just put you on the spot. And of course it's all speculation based on your expertise. I have a distribution in my mind of this century. how likely do you think it is that humans are extinct or largely extinct in the next 50 years?

[00:57:56] **Connor Leahy:** Depends on a condition on we do something about it or not. If we, 'cause I do think things can be done and I think this changes the probability we're gonna, we're gonna get to that, we'll get to that. I think the probability can change dramatically depending on what we do over the next two years. Um, if we do nothing.

[00:58:13] If we do literally nothing, if we do just, you know, I would say I don't know, 50 to 80%.

[00:58:23] **Nate Hagens:** Yeah. Okay. So, so I want to get into your, your work. but one more question about ai. how is the way that AI solves problems and performs cognition different from the way that humans do? and why is that misconception shaping the discourse on AI and what you refer to as AI alignment?

[00:58:45] **Connor Leahy:** So the true answer is, that we don't really know how either of us really do cognition, right? we don't really understand human cognition, not on a very deep level, and we don't really understand AI cognition on a deep level. We understand. A bit more about humans for sure. In particular, humans tend to have a universal kind of like set of emotions and p priors and basis that, well, not all of us, but that AI lacks.

[00:59:12] And this is usually a thing that trips off people the most with ai, where, for example, most humans are don't really wanna hurt other people. They don't,

they're not, you know, if they have enough food, they're safe, you know, respected and whatever. They don't really wanna just like randomly go out and hurt people.

[00:59:28] There are exceptions to this rule though. Um, and people, you know, would sometimes come up with theories like, well, if you're just intelligent enough, then you become peaceful. And I think this is nonsense. I think this is just, humans have a thing in their head, you know, designed by evolution that says like, all things equal be nice to the people around you.

[00:59:49] Be nice so they, you know, so you can build a tribe, whatever, right? But there are, for example, sociopaths who just don't have this I think. The way to think about sociopaths is not that they have an additional evil gene, it's more that they lack various emotions that drive us towards prosociality.

[01:00:05] **Nate Hagens:** Is there a magnetic attractor between sociopathic humans and the power that AI provides?

[01:00:11] **Connor Leahy:** I mean, quite clearly. So um, I mean, there's a clear drive here, right? Like a lot of the people that are building this technology are the same ones who admit that it could kill everybody. And they do say this clearly, like Sam Altman, Dario and Aade, et cetera, have said that it could kill everybody and they're willing to take that risk.

[01:00:35] And like some, I've talked to people in San Francisco who say stuff like, yeah, there's 10%. The AI is gonna kill everybody, but 90% it'll go great. So it's totally worth it. Which is just like I remember Dario Amede saying, for example, he thinks there's a 20% chance of things going poorly and like every, potentially everyone going extinct, which is worse odds than Russian roulette.

[01:01:03] **Nate Hagens:** But aren't they all com in, this race that they understand the risk, but if they do nothing and just grow potatoes and write poetry that other AI companies are, gonna try to go towards a GI and a SI. Anyways,

[01:01:17] **Connor Leahy:** so this is, I think really getting to the heart of, the, thing of like, why can things be done?

[01:01:22] The true thing is, obviously this is a fake answer. Obviously these super powerful men with billions of millions of dollars, all these connections, super political cut, extremely intelligent, extremely charismatic people. Of course they can do something to slow down the race. Are you kidding me? If Sam Altman tomorrow a charismatic, intelligent, super well connected, super rich person decided, I'm gonna stop this race.

[01:01:46] Obviously he could do something about that. Could he single-handedly solve it? I don't know, but obviously he could go to politicians, he could lobby for things to slow down. He could, you know, found new organizations, you know, he could just say clearly to every journalist that would listen, Hey, we need to stop this.

[01:02:05] Or he could just himself not make it worse. And there are obviously things that people like, you know, there's a thing that happens a lot where powerful people pretend to be oppressed in order to do the thing they want it to do anyways. Look, sometimes there are people who are genuinely oppressed.

[01:02:23] Like you're genuinely in a shit situation. You need the paycheck. There's no other way for you to get the paycheck. So you just have to do something not so great if you're in the situation. I'm sorry, sucks. Like it does happen, right? But that's not the situation these people are in. So what's the main driver?

[01:02:39] Just status and power. They wanna build a SI. It's not like you can read their blogs from even 10 years ago where they just really wanna build AI and build utopia. Like they're, it's not deep. It's wow, I could build a super cool thing, make tons of money, and maybe become God epic.

[01:02:58] **Nate Hagens:** So who do you see as the primary actors, that are most driving this acceleration to a GI and h how does this, um, backdrop contribute to the malignant outcomes in the middle of your distribution you're discussing or is it everyone?

[01:03:16] **Connor Leahy:** I mean, it's definitely everyone who's trying to build a GI like at this point, people aren't even hiding it. Back then people would be a bit coy about it. They wouldn't use the word directly. But now we have fucking Alibaba saying on their Quinn three report that this is part of their goal to build artificial super intelligence.

[01:03:31] They just like straight up, it's just on their blog, right? It's like it started with deep mind. Moved on with open AI and Anthropic, like those are kind of like the big ones. Right. But now it's, everyone is fully on board. Everyone is pushing as hard as possible. Right. You know, whether it's, you know, I mean mostly in the US and San Francisco, they're definitely the ones furthest ahead.

[01:03:54] I. So OpenAI, Google, Microsoft, et cetera, are definitely the furthest ahead in the race and are the best capitalized. But there's a bunch of tier two actors as well that are gaining ground, including in other countries.

[01:04:07] **Nate Hagens:** So if there was an AI actor that did have, um, some positive outcome for humanity and the biosphere aligned in their training and their algorithms or their objectives or their stated corporate laws, would they be outcompeted by these other ones?

[01:04:26] Um, and is that a possibility that there could be an AI that has some morality either embedded in the training or in the people that oversee it?

[01:04:35] **Connor Leahy:** So hypothetically, could we build a software system that if we execute, would lead to a good morally aligned like. Happy future for humanity. Physically, of course, like there's no fundamental barrier to building such a system.

[01:04:55] But lemme be clear here, this is saying we're gonna build a system that solves all problems. Humans face, day to day, all problems that governments face, that the gover, that the economy faced, that corporations face will solve all moral problems, will solve all, you know, voting theory, all like governance problems.

[01:05:15] And we're gonna solve all of this in one shot using software that will have no bugs. And I'm like, okay, hypothetically, like we are so far so. Unbelievably far. If we had all of our greatest scientists, all of our greatest mathematicians, you know, in a massive Manhattan project, work on this for decades, you know, for a century or generations of our greatest mathematicians work on this problem, our greatest philosophers worked on this problem.

[01:05:44] And, you know, unlimited funding, you know, highest security standards, extremely careful with, you know, international oversight. Could this work? Yeah, I think so. But this is not what's happening right now.

[01:05:55] **Nate Hagens:** How could we make that more likely? and maybe you could bridge the answer to that, to what you're doing with your organization and your work.

[01:06:02] **Connor Leahy:** I think the truth is that given how close a GI is, the number one most important thing is to buy time. Um, cha you know, political change, social change, building large projects takes time. I think we can, for

example, solve alignment. I think this is a thing that could be done at least good enough, right?

[01:06:27] That we would endorse it morally that we would like, you know, globally, we poll everybody and like mostly everyone's like, all right with this, right? look, if we poll the entire world, and you know, 90% of people said, you know, who cares? Just let it rip. We don't care about safety. Fair enough, right?

[01:06:45] Like this, I would disagree, but I think this would be like a much more morally acceptable, you know, state. But this is not what's happening. Like we've done polling and like people do not want to take these risks. It's a very, small group of quite sociopathic people, especially in San Francisco that are extremely driven to build these systems even at these extremely high risks.

[01:07:09] And the number one thing is to stop them. You know? It's just to be like, this is illegal. To be clear, I'm making very clear, the way to do this is regulation. The way to do this is legal, lawful legislation. This is the mechanism we use for chemical companies in the past when they were polluting rivers.

[01:07:26] It's the mechanism that we use to, you know, deal with nuclear proliferation across the world. We have mechanisms,

[01:07:33] **Nate Hagens:** so the. Sequence for that to happen is, first of all, there would have to be education and awareness. I went for a walk yesterday with a local friend of mine who, um, is basically just a, farmer.

[01:07:50] Um, I. And she said, oh boy, I wish I had some extra money, because if I did, I would put it into the market because AI is gonna make everyone so rich. Um, she just doesn't understand these risks that we face. So the first step is to make more people aware of the, loss in jobs, the loss in autonomy, the fact that a

SI could, extinct us all, the loss of identity and the cognitive, um, entropy that AI is gonna do.

[01:08:20] So we get more people educated and aware and feeling it. And I know a lot of people are feeling it, but I think that's probably still a minority. And then, you know, maybe there are things that if we have a super majority of humans in a population that vote for something or, put an opinion on it.

[01:08:40] President Trump would be a super majority sort of person that, that might want to do the will of the people and others in the world. So what are the possibilities of that happening?

[01:08:51] **Connor Leahy:** Yeah, I think you've got it exactly right. I truly think the solution here is not some crazy hair-brained, extremist, you know, scheme.

[01:09:03] I truly think the right thing is to do our job as citizens of democratic and free nations, which is to. See the risk that it is to bring to awareness of our federal fellow citizens and our policy makers to discuss and suggest policies nonviolent to solve these problems and then actually implement them.

[01:09:21] This is what we have democracies for. I think a lot of people have just forgotten what it means to be a citizen of a democracy as being a citizen of a democracy doesn't mean, oh, your policymakers know everything and they figure everything else, and you sit around, no. If you have a problem, you go to your policymakers and you say, here's an issue I care about.

[01:09:38] This is the problem we want, like how can we figure out a way to solve this? You know, you talk to, you organize, you talk to other people in your community and be like, Hey, we all care about this issue. What can we do about this? All this is kind of. Non-glamorous compared to you know, building, you know, silicon, god, you know, in San Francisco.

[01:09:57] But this is what made the world great. This is what made democracy great, right? is the fact that this is how we solve problems. We don't do violence, we don't do crazy hair brain schemes. We do, we educate, we bring solutions, and then we use state capacity to. Do things,

[01:10:17] **Nate Hagens:** do you know Audrey Tang or of Audrey's work?

[01:10:20] I know of her

[01:10:21] **Connor Leahy:** work, yes.

[01:10:21] **Nate Hagens:** She was recently on the podcast. it was really, inspiring on what some of this open source, open democracy things could do. I, think it would be interesting to have you and Audrey and for instance, Daniel Achtenberg or someone like that on a round table to discuss this.

[01:10:39] Something like that is gonna have to happen. So, um, so right now most people that I know, only intentionally interact with AI using things like chat, TPT or Claude. Um, are there any moral hazards that should be considered for individuals who use just those simple tools, but are also concern, understand, and are concerned about the, risks that you've outlined today?

[01:11:05] **Connor Leahy:** So the way I personally see things is, I think the moral cost for you personally using an AI tool is quite small. You know, like I wouldn't put it much higher than you using social media. Um, maybe or taking a

[01:11:17] **Nate Hagens:** flight

[01:11:18] **Connor Leahy:** or taking a flight or whatever, right? yeah, I think there are people who have stronger, you know, ethical principles on these than I do.

[01:11:25] I personally fly quite frequently for work and stuff, so I respect people who have, you know, stronger, you know, thoughts there than I do. I do think there are risks. That are worth keeping in mind, such as that AI systems just do make things up and they often make things up in subtle ways, and it can be very easy to get lazy in your own thinking around these kind of things.

[01:11:47] I think these things are similar to how social media is dangerous for your political alignment.

[01:11:51] **Nate Hagens:** Well, I actually think the, um, the energy use and material use, that's one thing, the output, which is that the GPTs get lazy and might be diluted. that's another, but the eventual day by day slow dependence on something external.

[01:12:10] we're outsourcing more and more of our own skills to the AI cloud. And then what's left in here and in here.

[01:12:17] **Connor Leahy:** I think this is a real issue. I don't have a good answer to what is the correct balance here? How do you know, deal with these things? Same thing with social media, right? what is the correct way for society to interface with social media?

[01:12:31] I think a good society doesn't have zero social media. Same way. I think a good society doesn't have zero gambling. You know, like I think there are ways to have these things that can even be net positive, like especially social media. It seems obvious to me that we could build social media that is good for people like that, is helps build communities that help build connection, that help people become more independent.

[01:12:53] It's just, I don't think it would be profitable to do that.

[01:12:55] **Nate Hagens:** So what are you doing, with your organization, your work, and given how, um, concerned you are about this and this is your field, what, are you doing?

[01:13:05] **Connor Leahy:** So it's a few things I work on. So my day job is I work for a startup that I founded, which works on technical AI alignment kind of stuff, kinda stuff we talked about.

[01:13:15] So very, briefly only, I've become a lot more pessimistic about this direction, though over the last couple of years. I think it's possible, but just extremely expensive and slow. Um, and so, I am a close advisor to the nonprofit control ai, based here in London, which is advocating exactly for the kind of policies that need to be passed and briefing policy makers.

[01:13:38] So over the last couple months, we have briefed over 80 members of parliament here in the uk, and we were told by every political consultant in Westminster that no one would ever sign our statement. No one would even give us the time of day. We were talking crazy, and this was extremely not true in our experience.

[01:14:00] People are often. Surprisingly reasonable. It's just that they aren't in uninformed. That's why I like that you use the word education. So most of our meetings are just politely informing our lawmakers about information they don't know about ai.

[01:14:14] **Nate Hagens:** Let me ask you this, you just mentioned that a lot of these, um, tier one AI plays are in not only the US but in San Francisco, a lot of countries in the world, there's probably 200 countries in the world.

[01:14:27] A lot of them don't have any Yep. horse in the race, in the tier one ai. So th this becomes a global west and a UK US sort of thing. I mean, I would think

even Europe, is less sanguine towards AI scaling than the US Oh, absolutely. So is there, like, how is that, um, with the geopolitics and all that?

[01:14:50] I haven't followed that closely.

[01:14:52] **Connor Leahy:** I'm glad you bring that up. Um, me, my colleagues are actually going to be publishing a paper on exactly this topic very soon, hopefully. Um, which is that. Yeah. I think there is a huge opportunity here for middle powers to work as brokers of peace and international agreement, where quite frankly, just from a purely le let's forget all the optimism and humanism for a second here.

[01:15:13] Let's talk purely reality politic, you know, NAS sec. There is, there are companies in the world, building systems that will plausibly be powerful enough to disarm and disempower the United States governments and military. These are systems being built. Super intelligence can disarm and disempower the United States government and military.

[01:15:35] This is happening on American soil today. They don't exist yet, but they're being built. So from a purely logical perspective, as a national security, this should obviously concern you. You should obviously care about this. What do you mean? Private companies are building systems that they themselves say will be used to CR disempower.

[01:15:55] Now there's, they've changed their rhetoric over the last year. Now they don't talk about creating utopia and so on anymore because that implies disempowering the US government. Now, they keep saying, oh, we'll give it to the US government, you know, we'll help the US government do this, but what does this mean?

[01:16:12] But the disempowerment of all other nations, if there is a super intelligence that is built on American soil, well, first of all, I think it destroys the us. It doesn't empower it, but let's ignore that even then. Well, it disempowers all other nations. All other nations now have a massive national security risk that is exclusively located in foreign borders.

[01:16:32] This is a big problem.

[01:16:34] **Nate Hagens:** Do you know what my honest reaction is when I hear this? Um, my honest reaction is, I really wanna buy a puppy.

[01:16:46] I'm serious because a puppy is, AI proof. They're not gonna be affected by this. I mean, isn't there this, deep yearning for authenticity and somewhat, maybe romanticized conditions of our ancestral human? I mean, that's how we're all wired anyways. We're trying to go through our daily routines, getting the neurotransmitters in emotional states of our tribal great-Grand Sisters, 10,000 generations a ago, and AI is like a, the mother of all speed bumps between me and the future that I would like.

[01:17:25] **Connor Leahy:** Yeah, and this is how I see it personally. This is why I think the first thing is very clearly it should be illegal to build a. Like, you know, just like straight up, this should just not be a thing you need. But no one's

[01:17:39] **Nate Hagens:** building a SI they're building AI that will eventually become a i, right? Sure.

[01:17:43] Okay. Exactly.

[01:17:44] **Connor Leahy:** And then the problem is how do you do this? And I think the thing is you have to ban precursors to a SI because once a SI exists already too late, right? like the moment an a SI exists that is not perfectly aligned and

wise and blah, blah, blah, it's game over. So our number one policy objective must be to never get into the situation in the first place.

[01:18:04] So we have some suggestions for how to do this. We've written a piece called A Narrow Path. You can look it up. Narrow path.co. Yeah, we'll, post it in the notes. Yeah. Where we talk about our principles for what precursors could look like and how banning them could work. We have recently also drafted a bill in the UK for how this could work.

[01:18:22] What, when you say we, what is that? Control ai, sorry. So still talking about control AI here, but I love the thing you just said. Is that fundamentally this is a speed bump between the future we want? I think this is exactly correct. This is how I think about it as well. there's been a third project that I've been like starting to spin up lately, um, which is Controli is focused on preventing that speed bump.

[01:18:46] Like we need to flatten out that speed bump. Otherwise it doesn't matter. But then there's another question. Alright, let's say we ban a SI and this buys us 20 years or something, right? Because like at some point someone's gonna figure out how to build a GI on the laptop. Like at some point someone's gonna figure out how to do it.

[01:19:05] I think it's gonna take a couple decades, but someone's gonna figure out how to do it. And so, or you know, some rogue terrorist somewhere will build a GI or something. So I think a maximum we can buy is 20 years or something. Okay. If we have 20 years, how do we get to a good future? Because I think we all have this feeling in our heart that it doesn't feel like the world is on track.

[01:19:27] And what I mean by that is. I don't feel like we have the expectation that next year is definitely gonna be better than this year. Like people don't have this expectation. Correct. And what would it mean to put the world on track where

every year we're like, next year is gonna be better and like we're confident that next year is gonna be better.

[01:19:46] And this is the thing I've been thinking about a lot lately. I've been thinking a lot about what would it mean to build a humanist future, like a future for humans that we like, you know, whether it's pastoral and puppies or something totally different, right? Like I'm, not committing to any specific utopia.

[01:20:04] What I'm committing to is, I think there needs to be a process. You know, there needs to be a process of how do we make iterative improvements? How do we get better at morals? How do we get better at values? How do we build better institutions? How do we coordinate as a species? And I think this is something that really needs to happen.

[01:20:23] **Nate Hagens:** I agree. Um, and I don't think there's a direct. One thing now, but the whole purpose of this platform is to change the hearts and minds and the initial conditions of the future so that the next cycle, whether it's a month from now or six months from now, or two years from now, the conditions have changed such that we have, um, better options.

[01:20:47] and that's the goal of my work. So one of the problems I think with hearing about climate change and, geopolitics and all the issues we discuss, including now, AI and a SI is hearing the story, at least on the surface, gives people a, an ill feeling, but also a lack of agency that there's nothing that they can do.

[01:21:09] So for those who are following and tracking what you're saying today, Connor, um, how can they start engaging and taking impactful action on these issues in their own lives or in their political spheres?

[01:21:23] **Connor Leahy:** There's two big things that I think are the most important things that can be done right now. The first got the control AI kind of like policy side of things.

[01:21:32] So you can go to the control AI's website right now, go take action, and you can send a letter to your representative. Right now if you're in the UK and the US it gets auto-filled. It's exactly what you need to do. Minimum thing you can do, tell your representative, I care about this issue. If you want to do more, fantastic.

[01:21:47] We, are following what we call the direct institutional plan or the dip. Also on our website, you can give it a read. The plan is very simple. We write down the policies that need to be passed to prevent this problem. And then we go to every policymaker in the world and in good faith make the case. We inform them.

[01:22:06] We educate them. And to do this, especially internationally, we need people. We need people from all across the world. Normal people who put a couple hours of work a week who just can, are willing to phone their senator, talk to their friends, do stuff like this. Also, in particular, if anyone from Hawaii is listening, it's the only state we currently have not had someone send a letter to their senator yet.

[01:22:27] So if you're in Hawaii, please, we need your help.

[01:22:31] **Nate Hagens:** We have a lot of listeners in Hawaii. Um, let me just ask a subset of that, which could be a, an awkward, question, but what if every citizen in the world, outside of the 330 million people in the US were a hundred percent on board with regulating AI and making it illegal to build an a SI?

[01:22:54] Um, would that matter since most of the tier one plays are in the USA?

[01:22:59] **Connor Leahy:** Absolutely. This would matter immensely. I think there is a lot of bravado about oh, the US can do whatever its wants. It doesn't care about other people. This is not true. Obviously, other countries can put pressure upon the US both soft and hard power wise.

[01:23:14] Um, obviously you know, there are trade relations and they're just cultural relations. there's friends, there's family, there's culture, there's wanting to be the good guy. You know, there are, it really does matter, like people all across the world, all countries like sure, you know, obviously if you know President Trump would like to have a chat about this, that would be great, but also other people in other country, it matters.

[01:23:37] **Nate Hagens:** Presumably China has some tier one AI plays underway.

[01:23:41] **Connor Leahy:** Close at least.

[01:23:43] **Nate Hagens:** Okay.

[01:23:44] **Connor Leahy:** This is the big issue, right? Like fundamentally we need a world where we have international agreement between China and the us. There's no way around this.

[01:23:51] **Nate Hagens:** So there's, I mean, I care about the natural world and the other species that are on the planet and climate change and the oceans and all that.

[01:23:58] And 20 years of researching that made me realize that we have to deal with our entire economic system issue first. And then I am starting to come to terms with, we have to deal with AI first. but underpinning all those things is governance. I think governance is the single issue in the world that will allow us to navigate the narrow path on, on many of these issues.

[01:24:25] Do you have thoughts on that?

[01:24:26] **Connor Leahy:** I think this is de definitely correct. I think we have seen a lot of loss in state capacity in many places in the world, especially in the us. it's become very hard. To pass legislation, to have even just rational and like calm debates about controversial topics, especially on social media is like almost impossible these days.

[01:24:46] I think these are things that make this problem much harder. Like it's crazy, right? I'm sorry to complain about social media again, right? But imagine it's the 1960s and Soviets came to Washington, DC put up a radio tower and started broadcasting propaganda. What would happen? Immediately be arrested, hell would break loose.

[01:25:05] Yeah. are you kidding me? We would've never allowed the Soviets to do this. Yeah. But now everyone goes on social media, TikTok, whatever, and gets directly blasted, you know, with like propaganda, you know, these things, you know, maximizing outrage, et cetera. So this does make it a lot harder to have good go governance, especially global governance.

[01:25:27] And there are real tensions here, like I wanna acknowledge here, like the rivalry between the US and China is very real and it's very serious. And I'm not trying to say people here are being irrational or stupid, there's some shit. Like I get it right? this is heart. I remember fondly, um, once I spoke in the House of Lords here in the UK and I talked about AI risk, and someone in the audience asked the obvious question, I.

[01:25:52] But what about China? You know, China will never agree to this. They will never disarm. it's never gonna happen. And then this old, like Scottish Lord, Desmond Brown, who's a good friend, um, basically said, what the hell are you talking about? We did the same thing with nukes, with the Soviets. It's diplomacy.

[01:26:11] Yet it's hard. Obviously it's hard, but like it has to be done. what are you talking about? It, was a great moment. I wish it would've caught on camera. It was such a wonderful moment. So I think it is very important to acknowledge that the odds are stacked against us. you know, the odds were never in our favor here.

[01:26:29] Like what we're trying to do to build a good future for our children, for our grandchildren, for our grand grandchildren, indefinitely lean into the future with powerful technology. Well, it's not something that's been done before and I'm trivial like, like obviously technology hasn't existed before. This is not something that's been done before.

[01:26:44] This is something totally new. It's something extremely hard. It's something extremely ambitious. But damnit, it's worth trying. Like damnit, it's worth giving it everything we can to try to actually build a future. 'cause we can, there's no physical reason why this can't work. We can build governance. You know, we didn't nuke ourselves to death.

[01:27:03] Right? Like we actually managed to get through the Cold War without nuking ourselves, you know, except dropping that one bomb in Cal, South Carolina accidentally. But that was an accident.

[01:27:11] **Nate Hagens:** You're right, it is physical and technologically possible. It's just that we have the perfect monkey trap has been developed where we're grabbing hold of the banana and not willing to let go because of our curious George social primate brain algorithm nature that we described earlier and that's the issue.

[01:27:36] **Connor Leahy:** I think it's a very important thing here also, which is one of the core reasons why. Have not given up. And when I say I think it's not over, I do really mean this. I don't just say this, you know, for content. Like I truly believe one of the core things is that people really don't want this. Like I have, we've done

polls with the uk, us, you know, we've talked to the general population everywhere, and people hate ai.

[01:28:03] They're scared of it. If you tell them, Hey, there's some guys in San Francisco building things that are smarter than humans, they don't really know how to control. How do you feel about that? The answer is universally bad. yeah, what, how is this legal? Are you kidding me? Currently there's more regulation in the uk.

[01:28:18] I got this check by a lawyer. There's currently more legislation in the uk. there's less legislation on building powerful, super intelligent systems that might kill everybody than there is on selling a sandwich. So. This is a thing that people do care about and they, once you inform them of this, obviously this can't be legal.

[01:28:42] Like obviously this should be highly regulated. Like we just need to actually do it.

[01:28:47] **Nate Hagens:** Yeah. I think the education curve is starting to catch up. 'cause to be honest, a year ago I, I was, a babe in the woods on these issues and even three months ago. So personally I've become a lot more aware of these things.

[01:29:02] And when I talk to people like Zach Stein and yourself, even more so, um. I would love to have you back. Um, if you have a few more minutes, could I ask you questions that I ask all my first time guests? Just taking off your AI hat for the moment, which for you is probably difficult to do, you're aware of some of the issues we face on the planet.

[01:29:24] Do you have personal advice to the viewers of this program at how to manage all this, um, in their own lives? Like just recommendations for behaviors, coping or engaging or whatever you think?

[01:29:38] **Connor Leahy:** Yeah, I think the most important thing is to not overdo it and to not do nothing. This would be my first piece of advice.

[01:29:47] Okay. Um, I think it's a very, good thing if you spend, you pick a certain amount of time you per week can be an hour, can be half an hour, can be five hours, whatever, where you. Try your best in good faith to think about the problem, think about how you would solve it, reach out to other people, read more about it, and then don't really do more than that.

[01:30:09] **Nate Hagens:** Yeah. that's great advice.

[01:30:12] **Connor Leahy:** Yeah. Like I think, and I not just say this like to you know, protect your soft heart or something. I truly think long, long-term stable investment is way more useful than these, burning yourself out and getting like super depressed and working over time. I think this is super useless.

[01:30:31] It's much, I would much rather work with someone who gives a hundred percent for two hours per week than someone who burns out 80 hours a week and is useless after three months. So if you wanna do something, this will my, our condition

[01:30:41] **Nate Hagens:** here. And you said that was one piece of advice. Do you have others?

[01:30:45] **Connor Leahy:** there's always many advice. Um, and another important thing is to not forget to have fun. Like you should still do things. There's a thing that some people do, including myself in the past sometimes, where you feel

guilty for doing things that are fun because you know bad things are happening. Um, and sure we can argue about Catholic guilt, like I was raised Catholic, you know, until me Too cows get home.

[01:31:07] So I know how this is, but this is just not a productive way to be a human. yeah, both like burnout wise, like you just, you won't handle it. And two, if you give up on this, then we've kind of already lost what we're fighting for. I, some people are like, oh, I won't have kids because of ai.

[01:31:23] And I just so strongly disagree if we don't have kids, if we don't love our kids, if we don't do things that are fun, if we don't eat good food, then we've kind, we've already lost. We've already lost what we're fighting for. And then What's the point? So still have fun.

[01:31:37] **Nate Hagens:** So speaking of kids, how would you change your advice for young humans, kind of 15 to 25, who are watching this program and, learning about this stuff?

[01:31:46] Do you have advice for young, humans?

[01:31:48] **Connor Leahy:** Depends on the exact young human, um, mostly is try many things. You have a lot of time, you have a lot of energy. Um, so you should try many things. You should read many things. You should talk to a lot of people. You should make a lot of mistakes. People will forgive you for your mistakes.

[01:32:02] This is a great, you know, benefit that you have for being young. Is it is okay for you to do something embarrassing or stupid people will forgive you. Um, we've all done stupid things when we were young. You should. Use this privilege, you know, by trying to, I find it very endearing when I get 18 year olds or something, sending me these emails with all their theories and stuff they wanna do.

[01:32:22] And I'm like, man, kid, none of this makes any sense. But I'm really glad you're trying, it's you know, but I say this genuinely, right? I'm not trying to be facet, right? Yeah. no, you're, yeah,

[01:32:35] **Nate Hagens:** you're a good dude. you embody the, um, the zeitgeist of this program. I'm so glad that, that we were introduced.

[01:32:43] Um, a question I ask all my guests, personal question. What do you care most about in the world, Connor?

[01:32:49] **Connor Leahy:** I think the true answer is I just want everyone to be okay. not even like utilitarian, maximize happiness or just I just want everyone to be okay. I want everyone just having a good time, you know, spend time with their family, their friends, you know, if you wanna do some crazy stuff that's fun, that's also fine, but just, I just want everyone to be okay.

[01:33:08] **Nate Hagens:** Thank you for that answer. if you had a magic wand and there was no, risk or recourse to your reputation or status or safety, what is one thing you would do? and I could guess at what you might say, what's one thing you would do to improve human and planetary futures?

[01:33:27] **Connor Leahy:** Well, I mean, depends on the limit of the magic wand, obviously, but I mean.

[01:33:33] I think a big one would be, I would summon the textbook from the future that has the solutions to AI alignment and coordination and all of moral philosophy all written down in a nice explain like I'm five format that I can understand. That might be a pretty good start.

[01:33:48] **Nate Hagens:** Some version of time travel. Um, so, if you were to come back on this program in six to nine months or something like that, what is one

topic that, um, you have expertise and are passionate about that is highly relevant to human futures that you'd be willing to take a deep dive on?

[01:34:09] **Connor Leahy:** we should definitely talk about humanism values, coordination, like how do we build institutions? How, what are human values? How do we like build a good future? Stuff like this, as that project I talked about earlier that I'm been starting also for the people interested in potentially putting a couple hours of volunteer work into doing something.

[01:34:29] It's called torch bearer. It's a brand new project. It's not super public yet, but if you wanna put a couple hours into trying to do something for the world, you know, building a humanist future, reach out.

[01:34:41] **Nate Hagens:** This has been great. Do you have any, closing comments or thoughts for our viewers?

[01:34:45] **Connor Leahy:** I would just say, as I've said many times before, it is not over.

[01:34:49] I think the cynicism of our age, like this, like belief that like things can't be changed is kind of like a self-fulfilling prophecy and it's often a much bigger hurdle than many other things. There's a thing where I've talked to many incredibly smart people, like geniuses, you know, you know, super, super smart, talked to super rich people, to super powerful people, and one thing they all have in common is they all feel like they can't do anything.

[01:35:15] They all feel like they're powerless. Everyone feels like they're powerless. And this can't be true. We can't all be powerless. And like we were talking about those, you know, I was talking earlier when we start our campaign in control ai, like briefing politicians and getting them to sign, you know, onto our campaign.

[01:35:31] We were told it's impossible. No one will do anything. You know, blah, blah, blah. And this just wasn't here. We have 35 supporters who have supported our campaign against extinction risk from ai parliamentarians, like real elected officials, right? This is not a simple thing, right? And so the one thing is it's just so much of the narrative that like nothing can be done.

[01:35:52] Oh, AI must be built, blah, blah, blah, is a complete self-fulfilling prophecy. It's not true. This is not an external force coming from outer space onto Earth. It's being done by humans today. It's like the people who say, oh, we can't do anything. We can't stop the AI race. Are the ones building it right in front of you with their own hands?

[01:36:10] Like we can do things. That doesn't mean it's easy. Don't take me as someone who's saying oh, this is gonna be easy. Everything's gonna turn out fine. We do need to do something, but we can do it and that. So if there's one thing to take away from it, we can do it.

[01:36:22] **Nate Hagens:** We just need to actually do it. Thank you very much for your time today and your very important work and your passion about these important issues to be continued, my friend.

[01:36:32] Thank you. If you enjoyed or learned from this episode of The Great Simplification, please follow us on your favorite podcast platform. You can also visit The Great Simplification dot com for references and show notes from today's conversation. And to connect with fellow listeners of this podcast, check out our Discord channel.

[01:36:55] This show is hosted by me, Nate Hagens, edited by No Troublemakers Media, and produced by Misty Stinnett, Leslie Balu, Brady Hayan, and Lizzie Sirianni.