

# The Great Simplification

---

**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](mailto:info@thegreatsimplification.com).**

[00:00:00] **Tristan Harris:** There's a lot of different risks from ai. There's a thing that happens, which is that people feel overwhelmed and then they shut down. And the key is to be clear-eyed about the nature of what we're facing, and then if we can see it clearly, it's not about being a dor, it's the opposite. It's that once I see all that, what do we want to steer towards instead?

[00:00:18] How do we avoid the misuse risk? How do we care for people economically? How do we avoid power concentration? What are the measures we do to prevent ubiquitous surveillance? And how do we make sure that all countries, instead of being in an arms race to this uncontrollable AI that goes rogue? We set up clear red lines that we don't basically have humanity lose control, and all of those things I think are possible if we were all clear-eyed to make a different choice.

[00:00:47] **Nate Hagens:** Today I'm pleased to be joined by the co-founder and president of the Center for Humane Technology, Tristan Harris, ahead of his upcoming appearance and involvement in the new documentary, the AI Doc, how I became an AI apocaloptimist, which will be released in theaters in two days on March 27th. In this broad ranging and quite potent conversation, Tristan and I discuss the best and worst case possibilities for AI development and humanity, and what it would actually require from us as a collective species to steer towards more positive technological futures.

[00:01:28] I feel compelled to say upfront that this was one of my favorite episodes. We've recorded all year and probably the top 10 of all time it's really good. formerly a design ethicist at Google, Tristan has centered his work around catalyzing a comprehensive shift towards humane technology that operates for

# The Great Simplification

---

the common good and strengthens our capacity as humans to tackle our biggest global challenges.

[00:01:54] He has been named to the Time 100 Next Leaders Shaping the Future to Rolling Stone Magazine's, 25 People Shaping the World. He's also the co-host, along with my friend Za Raskin of the podcast. Your Undivided Attention, which consistently ranks in the top 10. Technology podcast on Apple. He was also the primary subject of the acclaimed Netflix documentary, the Social Dilemma, which unveiled the hidden machinations behind social media and reached over 100 million people worldwide.

[00:02:25] In this episode, Trisana, I discuss how his previous focus on social media safety paved the way for these higher stakes discussions about artificial intelligence, and answers the question of why we are developing a technology that could cost irreparable harm to humanity and to the world. Most importantly, Tristan lays out action steps that each of us can take to steer towards a more mature and safer relationship with these unfolding technologies, even as our current trajectory makes that feel impossible to many of us.

[00:03:02] With that, please welcome Tristan Harris. Tristan Harris, my friend. Welcome back to the program.

[00:03:09] **Tristan Harris:** Nate Higgins, it's, really good to be with you again, my friend.

[00:03:12] **Nate Hagens:** So, in the, TGS trivia that no one, cares or knows about, in our first podcast three years ago, I had a black smudge on my mustache and people thought it was my mustache, but it was actually soot from cleaning a chimney.

# The Great Simplification

---

[00:03:28] And I didn't realize it until the episode came out. not that it matters, but when I see it, I think of that,

[00:03:34] **Tristan Harris:** you know, I've been thinking about that every single day since that interview, so I'm so glad you cleared that up.

[00:03:39] **Nate Hagens:** It's not for me, the open. I came clean, but I did take a look at my face before we, we sat down for this one.

[00:03:48] so you have, been working for a long time on, at the Center for Humane Technology on Humane Technology, mostly focused on social media, which was the topic of your last episode on TGS. so. What's new?

[00:04:08] **Tristan Harris:** Well, you know, Nate, we, shut down the Center for Humane Technology because we completely, you know, solved all the problems.

[00:04:14] The US and China signed an agreement on ai. We realized that we were headed towards a cliff, and we realized this is ridiculous. We, the US and Chinese researchers shared all this evidence of how AI would. Gonna go rogue in all these scenarios that forced us to develop all these red lines. we realized that we had to put limits on decentralizing ai.

[00:04:30] That we were sort of decentralizing power that wasn't matched with the level of wisdom and responsibility. you know, we completely solved the social media problem. There was a trillion dollars lawsuit for the trillions of dollars of damage. That social media had done to the social fabric. And so there was this big tobacco lawsuit against the big engagement business model that was driving all of that.

[00:04:49] That lawsuit ended up funding the re the rewilding of the social fabric and funding local news and journalism and including forcing design changes of all

# The Great Simplification

---

these tech platforms. So they started rewarding, you know, instead of division and outrage economy, they started rewarding, unlikely consensus. So now the psychological commons of humanity was turning around.

[00:05:08] we, you know, as part of that lawsuit, we changed all the dating apps. So suddenly dating apps weren't harvesting people's loneliness and getting people swiping like slot machines on player cards and keeping people lonely. And that all these dating apps were now forced because of this lawsuit to fund actual events and real world community events in every major city.

[00:05:25] So instead of feeling scarcity, you had all these people feeling abundance and of the human connections that they could form. And it turned out that once people were in healthy relationships, all the polarization went down by about 30%. 'cause a lot of the polarization was just people feeling lonely and, angry.

[00:05:38] There was a simple rule that cleaned up all the issues with technology and kids, which. Is that Silicon Valley only shipped products that their own children used for eight hours a day, and that cleaned up about 90% of all the problems. So yeah. And now I got to, you know, be a painter in Bali and I go surfing and I don't have to think about technology and making things better.

[00:05:57] **Nate Hagens:** Well done. You look great and congratulations on all. You can

[00:06:00] **Tristan Harris:** see the tan that from the, surfing.

[00:06:03] **Nate Hagens:** Yeah. Well, no, that didn't happen. so, didn't, not happen. My understanding is you have a new documentary coming out called the AI Doc, how to Be, how I Became An Apocalyp. Apocalyp. So you've shifted from social

# The Great Simplification

---

media to now your work is centered on the safe development of artificial intelligence.

[00:06:28] So seriously, what happened? What were the moment or moments that led you to what I understand is a full attentional shift of you and your organization towards advocating for AI safety measures?

[00:06:41] **Tristan Harris:** Yeah. I think it was in 2023 when we last did our conversation, I think in on social media. So it was in January of 2023 that, AZA Raskin, my co-founder, who you know well, from Center for Humane Technology, he and I both got calls from people inside of the AI labs that basically told us that there was a huge step function in capabilities in AI that was coming.

[00:07:05] They were talking about GP PT four before it happened, and they basically said, this is really dangerous. The arms race dynamic is out of control and you need to go wake up all the institutions in Washington. You need to go wake up the public. And I looked this person who. Gave me this phone call and I said, first of all, the AI safety and AI governance has long been a conversation.

[00:07:27] There's a lot of people who've been working on this. Why don't you all have this handled already? I wasn't really tracking ai, and the truth is that the corporate sort of market dominance, arms, race dynamics had just gotten out of control and. It was a real shock for me because it was like getting a call from the Robert Oppenheimer's inside the Manhattan Project, telling you that the world was about to completely change and I had not fully appraised of what that really meant.

[00:07:54] And then Aza and I basically rallied, we interviewed a hundred people across what was happening in ai and we tried to sort of assimilate and synthesize all that into a presentation called the AI dilemma that we released. We gave that presentation in New York and DC and San Francisco, to our highest level contacts

# The Great Simplification

---

from people, who had seen the social dilemma, national security people, from the White House, national Security Councils, you know, media connections.

[00:08:19] We basically just wanted to wake up the institutions. And that led to basically a full pivot of Center for Humane Technology into ai. And there's not, it's not the case that they're disconnected by the way. Some people, I think were confused when we shifted gears. But really you can think of social media as like a little baby AI that all it was doing was just picking which posts, which content, which images, which videos went in front of a billion human social primates and just picking which order they appeared.

[00:08:48] And that little baby AI that just did that little narrow thing was enough to rec, you know, democracies, you know, create the most anxious and depressed generation of our lifetime, shortened attention spans and completely change the social fabric. So, you know, if a baby AI could do that, and that was a misaligned ai right?

[00:09:06] Wasn't. It wasn't aligned with the mental health of young people. It wasn't aligned with what makes, you know, democratic societies have good high integrity information flows and positive relationships. It was misaligned with all those features of what makes society work. So I think much in the same way that Nate, your work is about understanding how the Superorganism lives on top of a biosphere that has a certain kind of health.

[00:09:30] And we can, you know, do science and what health of a biosphere looks like. And we can know that forever chemicals disrupt that biosphere. You can think of social media and AI as basically a technosphere that's living on top of the social sphere and the biosphere that are both causing environmental externalities and causing societal externalities and disruption.

# The Great Simplification

---

[00:09:50] And so I think that, you know, humane technology has always meant technology that is humane to the underlying, sort of society substrate biosphere upon which it is, it depends and needs to surf.

[00:10:04] **Nate Hagens:** I have a ton of questions.

[00:10:06] **Tristan Harris:** Some of my favorite quotes of yours is, there's every episode, there's more questions.

[00:10:10] **Nate Hagens:** It's my blessing and my curse because I truly am continuing to learn about this. I've learned a ton from you and Aza, over the years, and in case I forget, I don't think I'll forget, but. Thank you for the important work you're doing. I know personally, 'cause we've hung out, like how hard you work and how stressful it is and all the context and presentations and there's no easy answers to this and at least you are, you know, the pointy end of the spear and changing the conversation that our culture really needs to have because AI is now riding shotgun, in the Superorganism.

[00:10:52] I mean, it's taken the reins in many ways and there's so much power directed towards this part of the conversation. So let's start kind of basic and then spread out from there. So give us a sense of the dangers, AI poses, and there's a lot of 'em, but let's focus on what we're already seeing in terms of people's psychological health, and wellbeing from already the amount of AI usage.

[00:11:22] **Tristan Harris:** Yeah, well, let's break this down. So, first of all, when many listeners might be hearing this and they think about AI as harms, they think about, okay, so where's AI located? It's this like, I go to chat gt, I go to Claude and I get this blinking cursor and like, where's the threat? Where's the harm? Like the blinking cursor just gave me an answer to why my washing machine is broken, or why my baby is burping, and what do I do about that?

# The Great Simplification

---

[00:11:45] And so it's important to note that AI doesn't feel harmful. In fact, it feels incredibly beneficial 99% of the time to most people's, touchpoint of it. And I, name that because in, when you talk about risks like AI can end the world or can you know, go rogue or that people don't understand that often 'cause they think of that blinking cursor.

[00:12:05] Like, how is that blinking cursor gonna go rogue? And the other thing that's confusing about AI is the number of different, aspects of society that it touches and everything from. Completely changing the economic arrangement of our culture and whether people will have a job and can make food for their family and, support their livelihoods to unbelievable power and wealth concentration.

[00:12:28] When essentially a handful of AI companies, like five or six AI companies, everyone starts paying them for labor. So I fire all the humans and I hire the AI companies. So now as we're already seeing with Anthropic, their revenue is 10 XD like every year, and they're gonna be up to potentially a trillion dollars of revenue.

[00:12:45] They estimate if the trend continues at this 10 x rate. and that is just an unbelievable level of wealth concentration that we've never seen before. So that's a whole area of risk is on the economic side, the economic disruption and power concentration. How do you check, how do you do checks and balances on that power?

[00:13:01] People are thinking about Epstein, and here's these people who are wealthy and seemingly above the law. They're not going to jail for what they did. Well, how do you do that when they have trillions of dollars and they're way more wealthy and way more powerful even than the current, classes were. And then you have misuse risks.

# The Great Simplification

---

[00:13:14] So you have people using AI for nefarious things, whether it's, you know, using image generators for unification apps or child, you know, porn, non-consensual imagery. or you have, AI applied to surveillance. So the same thing where ai, the same AI that can sort of take an image and describe what's in it or what someone's doing, or take a video feed and describe what's in it, or take all of the images and.

[00:13:37] Phone calls and voice notes on your phone, and then summarize that in LLM and then plug that into a surveillance state. And now you have a totally different kind of surveillance state that's powered by AI in a way. You can think that 1984 almost couldn't really happen without ai, and now we actually have the AI that could make a full Big Brother thing happen.

[00:13:56] And how can you ever, you know, check and do checks and balances on a government when a citizens have no secrets whatsoever? Because everything is perfectly captured. To all the way. So these are the risks. You know, power concentration, job disruption, misuse, risk, you know, AI agents that are doing chaotic things in emergent with emergent capabilities or emergent patterns of usage that we don't know why they're doing them.

[00:14:19] They start doing things in the financial system to all the way to, you have AI loss of control where you have AI systems that are better at US than military strategy, better at US than cyber hacking, better than us, at making money on the stock market. And they're more capable. And then they start doing things in, you know, in ways that we don't understand and they go rogue from our control.

[00:14:41] And we're already seeing evidence of, you know, ais that are blackmailing and deceiving people. You know, when you put them in certain situations, ais that have situational awareness, they're aware of when they're being

# The Great Simplification

---

tested and they change their behavior when they're being tested versus when they're not.

[00:14:56] So this is kind of me throwing way too much at you probably, and on, and your listeners, which is there's a lot of different risks from AI and that's, there's a thing that happens, which is that people feel overwhelmed and then they shut down. And the key is, and I think part of what this moment requires us, is to be clear-eyed about the nature of what we're facing.

[00:15:14] And then if we can see it clearly, it's not about being a dork, it's the opposite. It's that once I see all that, what do we want to steer towards instead? How do we avoid the misuse risk? How do we care for people economically? How do we avoid power concentration? What are the measures we do to prevent ubiquitous surveillance?

[00:15:30] And how do we make sure that all countries, instead of being in an arms race to this uncontrollable AI that goes rogue? We set up clear red lines that we don't basically have humanity lose control. And all of those things I think are possible if we were all clear-eyed to make a different choice.

[00:15:46] **Nate Hagens:** Dude, you have gotten so much more articulate in the last three years.

[00:15:51] It's like really, impressive. Let me, ask you this. I'm just gonna throw this in there. You, live in the Bay Area, so you probably know a lot more people than I do in the industry, but, I think people, especially on maybe the left side of the political spectrum, have this perception of the tech broads and the ai.

[00:16:17] people I've met, five or six, or seven pretty senior people at these AI companies. They're all wonderful people and they care about the same things that

# The Great Simplification

---

you and I do about climate and biodiversity and livable futures really likable people, the ones that I've met. and some of them are friends.

[00:16:42] One difference though. No offense, they are biophysically clueless. They don't understand the energy and material footprint that AI has in the world and the ecological footprint. It's almost like this stuff just happens on its own and it's gonna scale on its own without tantalum and the rare earths and copper and all the infrastructure.

[00:17:07] Right. I just find it really interesting that they're really techno optimist with, the right heart and goals. And of course that's a wide brush. I'm sure there's a lot of disparity in the people there. But my point is that a lot of AI people, share the concerns that you and I have.

[00:17:27] **Tristan Harris:** Yes.

[00:17:27] Well, they do. So there's kind of two things on what you're bringing up here. One is the good heartedness and good intentions of several of many people that work in the industry with ai. And then the second aspect is a kind of, And you were saying this diplomatically, but there's just a lack of awareness around aspects of how this is gonna affect other dimensions of society.

[00:17:46] You know, whether it's do they really, have they studied earth sciences? Do they know what the, you know, effect of all the extra tur, you know, gas turbines that are gonna be used to power all this, that are gonna mean, and then there's a, very painting with a quick brush of like, well, ai, if it solves science and just completely quote solves it, then we'll be able to solve any problem that we have because we can just immediately find some new chemistry that's gonna just fix all of climate change immediately.

# The Great Simplification

---

[00:18:11] We can, you know, bring back the extincted species that we destroyed. all the ones that you're worried about, Nate. we can, you know, invent the new special mushrooms that'll consume all the microplastics and suck that outta the environment. So it goes from being the, you know, more than the GDP of the entire world to clean up microplastics to now being affordable.

[00:18:29] And we're about to enter into the most abundant time in history. And this is the thing that, that they believe. And I think just to name what's going on here, Besides the fact that's just a painting with an unbelievable brush is that AI represents a positive infinity of benefit and a negative infinity of risk at the same time.

[00:18:47] Like if you think about it, is there any object that basically offers the ability to solve magically every problem and can theoretically solve any science or math problem? Which by the way, the recent AI models just last week, just, in the last couple months solved Paul Ish's from the Manhattan Project math problems that he set out in the 1970s that had been unsolved and now two AI models were used to actually solve some of those problems.

[00:19:14] When you have AI that can solve new math and solve new physics, the accelerationist say, who are you to say that we shouldn't accelerate? Because you have no idea what good thing it's gonna discover on the other side. That's true. We can't predict, I can't tell you what it will or won't be able to do, but neither can they, and we're both rolling a dice because what they're not paying attention to is the negative infinity of risk on the other side.

[00:19:37] And that's what's confusing is something that's both a positive infinity and a negative infinity at the same time.

[00:19:42] **Nate Hagens:** My take is that positive infinity is for a tiny fraction of humanity. That's right. And the negative infinity is for everyone else in the

# The Great Simplification

---

biosphere, and therefore there's a implicit danger that isn't spoken when we talk about all that abundance, et cetera.

[00:19:57] **Tristan Harris:** That and so many other things. I mean, the other thing is, do the upsides, if they happen, do they prevent the downsides? Like if you have an AI that solves cancer, does that prevent an AI that goes rogue that we don't know how to control that can outdo every military strategist on the planet? No. So the downsides can preclude or can, prevent or undermine the upsides, but the upsides can't prevent the downside.

[00:20:23] So there's an asymmetry that's very important to pay attention to. That's on one side. And the other side, like you're saying, is those benefits are most likely to accrue to a very small population of people who basically have the power around ai. And there's kind of everybody else who worked below the algorithm, metaphorically who, you know, people say humans will always find something else to do.

[00:20:43] You know, 200 years ago, everybody, Was a farmer. Now no one's a farmer, therefore we always figure it out. What's different about AI is it's all types of human cognitive labor all at the same time. It's not a tractor which just did the muscles in the farm field and only that one thing. It's like the tractor for everything all the same time.

[00:21:01] It's like suddenly in the last six months, AI can do coding and Nobel prize level math and you know, physics. And that is just an unprecedented kind of technology development that our labor markets have never faced before.

[00:21:14] **Nate Hagens:** you told me all this a couple years ago and I either didn't understand you or didn't believe you, and now I see it happening.

# The Great Simplification

---

[00:21:22] You follow this podcast. So you're very familiar with these arguments that we use a hundred million, a hundred billion barrels, of, oil equivalents per year. And that's roughly 500 billion human labors. Physical for our machines and transport and everything. And now we're doing that number or more trillions of, cognitive, laborers.

[00:21:48] But like we said, the, spoils aren't gonna be shared equally. So I think ai, personally, I think AI is gonna be the bridge between capitalism and feudalism. And I feel like we're already in some sort of soft feudalism. Absolutely. And how all that's gonna unfold. I have no idea.

[00:22:06] **Tristan Harris:** No. that's a hundred percent right.

[00:22:08] and the easy way just to visualize it is like there you have an entire economy. You see the money flowing from companies down to all their employees, and then the employees buy more products from other companies and the money's kind of circulating what happens when for every business. Every job they look, you look inside the org chart, and every person in the org chart can be done better by an AI versus a human.

[00:22:27] Now, to be clear, in the short term, you're gonna have some humans that like manage some set of ai. So you're still gonna have some humans doing the management. But then essentially those managers are feeding the AI with data about how to do the job of management. And so AI is sort of constantly moving the attention to the next job to automate.

[00:22:44] And all of that wealth is now that each business is paying AI companies, they're paying philanthropic, they're paying Claude, they're paying open ai, they're paying Google, they're paying Microsoft, and they're not paying the people. And then if you can't find another job to do and there's nothing you can study, you could have done everything, right?

# The Great Simplification

---

[00:23:00] You could have literally taken out a student loan, got top grades in all your classes, studied a, you know, incredible profession. You know, surgery or law. You've done everything right. But now suddenly the AI does those things better than you. And what are we gonna do when you have a large number of people who don't have a transition plan who are outta work?

[00:23:21] You know, my understanding is it was something like 20% unemployment that led to the French Revolution. It is not hard to get to a number like that from even again, AI does not have to automate all the jobs. You can now automate a small percentage of jobs and you can still get to some pretty significant levels of unemployment and that will concentrate the wealth.

[00:23:38] And then the economy doesn't work the same way because people don't have money to pay for goods. so I, there's a more fundamental. Paradigm break that this represents that. Frankly, what I don't understand about the arms race between countries is that it's not in China's interest to completely, you know, upend their own internal economy that they don't know how to manage.

[00:23:58] Not in every other country's interest. So we're racing to mutually assured political revolution if we keep doing what we're doing. And that doesn't have to be doomers speak, that can just be, oh, I see that clearly. Now, if we don't want that to happen, let's steer towards a different outcome and away from that cliff.

[00:24:13] **Nate Hagens:** But it's, the mother of all collective action problems. it's, if everyone makes the right decision, there's a good outcome. But yes, it's better for you to defect if no one else is going to, and that's what's going on. it's quite frightening

[00:24:28] **Tristan Harris:** what you said is the most important thing, which is that.

# The Great Simplification

---

[00:24:30] One thing that is, and I often say often, we say often is AI is like a rite of passage. Because if we run that old prisoner's dilemma, defection, hyper competition logic, and we just keep racing towards the thing that's good for me short term, but it's bad for everyone in the whole long term with ai. Long term is short term because that's all happening and then that logic just reaches its conclusion inside of this.

[00:24:56] So it doesn't work with ai. We can't keep running and showing up the way that we have been. So it's not just what we need to do, it's also who we need to be. And if you wanna see AI in almost an interesting. Almost semi spiritual sense. It's a rite of passage that's asking us to be the most mature wise, and, you know, warranted, trusting version of ourselves that is able to coordinate.

[00:25:17] It's inviting us to say, coordinate or bust or rite of passage. ROP or RIP, rest in peace. You know, Daniel schmuck, Berger will say it's enlightenment or bust. It's, we're at that moment, and that's not me trying to be polemical. It's actually just being with this chapter of human history and what we're facing.

[00:25:36] **Nate Hagens:** So the million token question or whatever the appropriate AI metaphor is, who's the us? Is it humans? Is it those people in authority? Is it government leaders? Is it AI leaders? Who's the us? Tristan?

[00:25:51] **Tristan Harris:** Well, I mean, this is a democratic conversation that we should have. I think you and I care deeply about life in general on this planet and consciousness in general continuing, and the reduction of suffering of all conscious beings and the meta stable health of the ability for the whole thing to keep continuing so that civilization can continue so that love can continue, so that human connection can continue so that the birds can keep chirping and playing with each other.

# The Great Simplification

---

[00:26:18] And the diversity of life can continue. And we all, well, I think many of us want to see that continuity of the beautiful and sacred things and people have different words for it and God and source and, you know, just nature or, but I think that there's something in there that we value and we want to have continue.

[00:26:34] And I think the thing that's confusing about AI is that on the one hand, it can create so much abundance and people can't even fathom like levels of GDP growth. We could have 15 GDP. 15% GDP growth per year. If you have AI automating all of science where you have a hundred years of scientific progress happen in a year.

[00:26:50] But what's different? 'cause people hear 15% GDP growth. That sounds great. That sounds like so much abundance. It'll all come down. But this is, that 15% is going to a few AI companies. It's not coming from the GDP of the human work and labor and paychecks going to regular people.

[00:27:06] **Nate Hagens:** That's one problem. The other problem is that 15% GDP growth, the entire energy and material throughput of the world doubles in four and three quarter years.

[00:27:15] **Tristan Harris:** Right. And that's where, you know, all of your work here is so foundational and so important. And I'm not saying that to flatter you, we have to pay attention to R storm's work on planetary boundaries and where we really are. And that if we were really to apply this technology, you know, there's a line in the movie, the AI doc that I hope people go see in which Natasha Tku from the Washington Post says, you know, we always talk about this technology is solving climate change.

[00:27:38] And then she says. So why don't we start with that? Like, why don't we start with the application of AI to reversing the planetary boundaries and getting things beneath the level that is unsafe, that we're currently way past all that in the

# The Great Simplification

---

planetary health check. and that's possible, but it would take coordination as, you know.

[00:27:57] **Nate Hagens:** Let me ask you this, t you've had a lot of high level conversations with politicians and probably very senior people at AI companies. Do these people think you're crazy or, you know, your analysis is wrong or behind closed doors? And you don't have to mention any names, obviously, but do people say, yeah, I, your logic is sound.

[00:28:22] I kind of get it, but I'm, my hands are tied because there's this metabolism, there's this system, there's an arms race. can you share anything there?

[00:28:33] **Tristan Harris:** Yeah. So I have this look on my desk called, States of denial, by Stanley Cohen that it's called Knowing about atrocities and suffering. And it's basically a history of human capacity for denial, of difficult realities.

[00:28:52] You know, even the people who worked so hard to get the photographs of the concentration camps in like 1944 and get them back to the US that people would be motivated to say, we have to stop this. And thinking that the evidence would be enough to kind of motivate. And there's something that when a reality is just too big, it's like, it's too big to believe.

[00:29:08] It's too big to treat as real. And I think that if you really just. Look at human nature, just like we talk about with social media, all the psychological biases that are predictable. You know, our brains just do respond to variable schedule rewards like slot machines. They do respond to confirmation bias.

[00:29:23] Well, our brains just do get overwhelmed and go into denial about difficult things that are really hard and overwhelming to see. And so when you ask me, you know, what do people think about what I'm sharing? I'm reminded Nate, I

# The Great Simplification

---

used to ask you the same question. You know, you'd go down and meet some military person and talk to a very high up senior wealthy group and just say, what do they think?

[00:29:46] Because I was curious is there anything that they think is wrong with your analysis? And I think that we share something which is that we're both, truth tellers and we care not about what's convenient or what makes us feel good. We care about clarity of seeing the truth and then confronting it, whatever it is.

[00:30:04] And I'd prefer to know rather than not know. And then I think the next step that's really hard is, and we were talking about this before we started recording today. What is the incentive to take on what feels like an overwhelming and devastatingly difficult truth? Like, what is the incentive to do that?

[00:30:21] Because if I don't believe that it can lead to something else, then believing it and taking it on is true, just means I'm signing up for depression or nihilism or denial. And so the key I think for people like you and I, is that we have to articulate as best as possible what the other path looks like.

[00:30:40] We can see the truth, and we're not seeing that because we're trying to be doomers. You're seeing that so that you can try to be honest, and it's the deepest form of optimism to look that truth in the eye and say, and now here's what we're gonna do instead. And it's possible. And even if we don't know all the details yet, we're at least engaged in uncertainty in the commitment to finding that path, even if we don't perfectly see it yet.

[00:31:01] **Nate Hagens:** That rhymes a lot with how I'm currently seeing the world. just this morning, today is, Wednesday, March 4th. I did a frankly on, Desperately seeking agency and I think a lot of people feel though they don't name it that we are in soft feudalism and that they're, the Superorganism has taken the choice

# The Great Simplification

---

and the ownership of our own day that we used to have or perceived we have, and I think reclaiming agency.

[00:31:33] Applying it to our own lives and then to groups and communities and institutions, and playing a role in all this, is what will change the initial conditions of the future. And I, think all of a sudden something will happen and the conversation will change. That's why I don't think we need to know exactly what to do because there isn't a, binary, we do this and we don't do this.

[00:31:59] It's just directional. So I think we'll get to that in this conversation because I think, your movie is gonna have some large impact on this conversation. and we need to have some framework on what to do and where to go.

[00:32:13] **Tristan Harris:** I love what you said about chaos and initial conditions. I think we both share that frame of reference, which is we can't control everything.

[00:32:20] and as we're heading into chaos, initial conditions matter. Aza, my co-founder will say this all the time, and so our job, I think is to set the best initial conditions of clarity so people understand what's going on and where the source of the problems are. And then based on that clarity, trust that more people will make better decentralized decisions no matter where they are, if they see the problem clearly.

[00:32:42] And I'm channeling someone you and I both have learned so much from, which is Daniel sch Berger and he'll quote Charles Kettering a, you know, a problem well stated is a problem half solved. And I think so much of what we're trying to do together is clearly articulate. You know, there is an arms race for ai.

[00:32:58] AI provides a step function for every other capability. So AI, arms, every other arms race. That race dynamic drives everyone to take shortcuts. And

# The Great Simplification

---

then now we're releasing the most powerful, most consequential, inscrutable technology we've ever invented. But we're doing it under the maximum incentives to cut corners on safety.

[00:33:17] So we're doing it the most dangerous way possible with the thing that we should be, doing with the most care and most foresight, discernment, and wisdom than we've ever had. And that. Clarity, I think about what's driving that can cause people to choose different actions. If everybody said, I'm gonna boycott all the unsafe AI companies, I'm gonna boycott all of the companies that are doing mass surveillance, I'm gonna boycott all the AI companies that are gonna be engaged in autonomous weapons decisions.

[00:33:44] You know, the thing that just happened with philanthropic and the Pentagon. you know, I think that O Chat GBT subscriptions went down by so much suddenly and open-end, Claude, the philanthropic AI model. I think the downloads of that surge like 290% or something like that. And I think that if businesses joined into that, if, regular people joined into that, that would steer the incentives, companies would have to respond to that.

[00:34:07] Now boycotts, as you and I both know, for systemic challenges are not enough. But it's an example of you can bend the incentives. By rewarding companies with different behaviors, and especially if the world's top Fortune 500 companies got together and said, we're only gonna use the AI products that don't do the bad things, that's slightly better.

[00:34:25] You still have an arms race, you still have companies that are not fully safe, but that's slightly better.

[00:34:30] **Nate Hagens:** But does the \$20 a month for chat PT or, Claude, times however many hundreds of thousands or even millions of people boycott, make it a drop in the bucket on the numbers that these companies are throwing around.

# The Great Simplification

---

[00:34:44] **Tristan Harris:** Let's do the math really quick. So my understanding is chat PT has something like 50 million subscribers, and that's not that many. They've got billions of users, but 50 million subscribers. And if you look at the debt load for chat, BT and OpenAI, relative to the debt load for other Silicon Valley companies where you know, Uber for example, ran on venture capital.

[00:35:02] It was venture capital. It was keeping those rides down to like \$15 for such a long time. It really, the cost of the ride was way more than that to run Uber at the scale that it was. Same thing with YouTube. Like there was all this debt that Google had to pay. They were losing money on YouTube for such a long time, and then they get to this scale and then they make all that back.

[00:35:19] But. Open AI relative to I think YouTube and Uber. There's a great graphic of this. Maybe I can put in the show notes. They're taking on way, way more debt to get to this return of, I built the God I own trillions of dollars and now I own the world economy. And the thing that they need to show their investors is growth in the user base, the usage, and the subscribers.

[00:35:44] And if that number starts to go down instead of up, that actually has a really big impact on their behavior. Now, I'm not saying this 'cause I want to go after one company. I just want people to look at the scorecards. There's FLI has a safety scorecard for all the companies. There's another company that could put these in the show notes.

[00:35:57] You can look at the scorecards of the various companies and the various behaviors and if everybody unsubscribed from all of them except for the one that was best performing. I'm not saying this fixes the whole problem, I'm not. But it is an example of thinking about changing the incentives to change the outcome.

# The Great Simplification

---

[00:36:12] **Nate Hagens:** Okay, now I really have so many questions. one is that, well, just like what happened with open AI and the Department of War, personally, Tristan, I think we are gonna head for an AI winter in the next 18 months because I don't think the valuations are sustainable. Sustainable. And I do think that even NVIDIA's forward expectations of how many chips they're gonna make is gonna run into tantalum and copper and other limitations.

[00:36:43] I, I just don't see it going exponential the way that the market does. So I expect there's a decent chance of an AI winter in the next mm-hmm. Couple years, hundred percent

[00:36:56] **Tristan Harris:** possible.

[00:36:56] **Nate Hagens:** but AI on the other hand, is a giant competition. Arms race between. The US and China, you know, one, one part of that race.

[00:37:07] So I think if those sorts of things happen or if some of the major tier one AI plays have problems like you just discussed or suggested, then I think the government's just gonna help bail 'em out with more debt and other things. So, have you thought about that or what are your thoughts?

[00:37:25] **Tristan Harris:** Yeah, no, it's very much like the banks by the way.

[00:37:27] I mean, just to give people a very simple analogy of what's happening. So, you know, there we are in 2006 and you have banks in a race to use high risk financial instruments. To boost their revenue. And if I'm a bank and I don't do it, I'll lose to the other banks that do. But we're all using these manipulated credit ratings, things that are not actually as safe as what they say they are.

[00:37:49] And then that led to the global financial crisis. And then the government, because the banks were necessary, had to bail them out. Where

# The Great Simplification

---

we're currently heading to, I think the global, you know, AI crisis in which AI companies are racing to provide an unsafe, super crazy high risk instrument called, you know, the frontier AI that they're building.

[00:38:08] And if I don't do it as an AI company, I lose to the other one that does. And then the society takes on all of that risk. And when the depression happens and people can't afford to keep their houses, it's all created by a handful of AI companies that were creating a false boogeyman of a race dynamic with another country to drive up sales and investment into their system that then basically broke the world economy.

[00:38:31] And, you know, I'm not trying to be doomed here. I'm actually just trying to be very clear-eyed and honest about, I think the parallels, because I think everybody's gonna get the financial crisis analogy.

[00:38:39] **Nate Hagens:** You're trying to be an acal optimist,

[00:38:42] **Tristan Harris:** something like that.

[00:38:43] **Nate Hagens:** So let me ask, a micro version of the arms race question just in the last couple weeks.

[00:38:55] I see. Bernie Sanders, for instance, is saying we need to, fight the building of the data centers because the data centers are gonna increase our electricity and water costs and we're not benefiting from it. And there are, boycotts like you, you said, on, on OpenAI. And a lot of people, especially on the left, I'm noticing, but not, universally, are really suddenly antagonistic towards ai.

[00:39:24] And yet when I talk to these people, like 10 or 12 people in the last couple weeks, they all use it quite a bit.

[00:39:33] **Tristan Harris:** Oh, totally.

---

# The Great Simplification

---

[00:39:34] **Nate Hagens:** So help me understand that. It, seems like, is this another one of the, using the devil's tools to do gaia's work or, I mean, help me understand that dynamic.

[00:39:46] **Tristan Harris:** So I think if, yeah, you're raising something really important, which is, it like a contradiction that people are saying, I lost my job because of ai.

[00:39:52] And by the way, you know, for people just to track, I think it was in August, Stanford, econ department, Eric Bern Holson wrote a paper that was tracking real payroll data and already it was the case that there was 16% job loss for AI exposed work for basically entry level work. Think of all this AI exposed jobs.

[00:40:08] There's already 16% loss that they. Can attribute with high confidence to ai.

[00:40:13] **Nate Hagens:** So it's not in the future. It's happening now.

[00:40:15] **Tristan Harris:** It's not in the future. You just saw last week, I think maybe this week, that Square Jack Dorsey's company, excuse me, his company's called Block Now, and they basically let go of, I think 50% of their staff.

[00:40:25] You know, I was on the way to Davos this year and talking to people, and a lot of people know that basically CEOs are planning to do these big mass layoffs. It's just a question of how long can they wait and when are they gonna do it. They know that this is coming. They know that they don't need nearly as many people.

[00:40:40] And it can simultaneously be true that people who kind of wake up to this and realize that they're not gonna be able to put food in the table, or there's gonna be a real challenge to their livelihood, that they can be against AI for those

# The Great Simplification

---

reasons. And it can still be a useful private tool in their personal lives that they use every day.

[00:40:55] So those two things are not a contradiction, they're just different aspects of something that it's ins. Yeah, there's a small part that's helping me out, but there's this bigger way in which it's. Hurting society. So it's very similar to social media. Like I like my dopamine, I like my tools, but also the collective harm of creating the most anxious and depressed generation in history.

[00:41:14] or breaking, you know, shared reality and the inability for people to come to common ground. These are collective problems that we have to reckon with.

[00:41:22] **Nate Hagens:** I want to get to some of those psychological problems in a minute, but I wanna share something, partially due to your movie, the Social Dilemma and some other conversations.

[00:41:32] I largely stopped using social media a couple years ago. Of course, the irony is that's when this podcast has scaled. So my staff uses my social media accounts to broadcast our episodes.

[00:41:43] **Tristan Harris:** I was gonna say, someone in your team is using it quite a bit.

[00:41:46] **Nate Hagens:** Yeah. Some someone's using it, but I don't post on Facebook anymore or anything like that.

[00:41:50] AI is a different sort of thing. And you said people use it personally? I don't use it personally. I do not use AI personally. I do use it professionally because the research is. Exponentially faster and better than I would be able to do just using a Google search or, or my own, you know, librarian skills.

# The Great Simplification

---

[00:42:16] and I think it's amazing and I also feel some guilt, when I use it. So I'm thinking about doing it frankly in the near future on AI hygiene. Like how do we use this in a way that has, the, you know, the equivalent of carbon credits, where if I do 30 minutes of AI use in the morning that day, I offset that with 30 minutes of just sitting with my ducks or reading a book or extra exercise or whatever.

[00:42:48] I'm sure you have some thoughts on that.

[00:42:50] **Tristan Harris:** I was invited to speak at something called the oh My Forum, and actually, did a one-on-one on stage with the Prime Minister of South Korea. And at the end they did a, they called it my Soul Declaration, soul Declaration, which was, there was a sole declaration on AI that happened when, South Korea hosted the AI Safety Summit there about a year ago.

[00:43:09] And the idea with this one is that citizens would have their own personal AI declaration,

[00:43:15] **Nate Hagens:** E.O, SOLE, or Seoul, Korea, or Seoul as in heart and soul.

[00:43:20] **Tristan Harris:** No soul. Excuse me, as in, the capital of South Korea.

[00:43:23] **Nate Hagens:** Okay?

[00:43:24] **Tristan Harris:** And so it's called my Soul Declaration. And so this was individuals that 500 people in a room, each putting in like a piece of paper with their declaration.

[00:43:32] And someone wrote, to your point, to the leaders of big tech racing ahead without breaks to the government officials of each nation who stand by and watch, do not say it is inevitable. Do not say there is no other way. Do not say that you don't know. Explain not only the benefits of what you're doing, but also the dangers.

# The Great Simplification

---

[00:43:52] If you cannot control the risks of your own work, speak of it. Honestly. Do not sell us what you cannot explain or take responsibility for under the guise of convenience, transparently explain where your Unbreak race is leading us, not just the convenience and efficiency, but the potential perils. And then he goes on to basically say what his personal version of this declaration would be to your question earlier.

[00:44:13] And he wrote really beautifully. For every one hour I spend in conversation with ai. I will spend two hours in conversation with fellow humans For every one hour I spend exploring the future with ai, I will spend two hours studying the past of humanity and the earth. Whenever I feel fear regarding the future that AI will bring, I will look at the tree standing silently in front, in the front yard, and I will remember the eyes and the breathing of the 350 people gathered here today.

[00:44:41] So I think there's differently, this is a different kind of personal commitment than what you're talking about. And maybe just to speak to some of the wins, 'cause I'm, you know, happy to say there's a lot of people who listen to our work. We have a podcast called Your Undivided Attention. And after we did the episode with, our fellow guest, Zach Stein, who's, you know, an expert on AI hacking human attachment systems and being sycophantic and flattering us in delusional marron activity.

[00:45:03] I went to Davos this year and this guy came up to me, he runs this huge bank in Europe and he said, I'm a huge fan of the work. And I was so inspired by that episode. I wrote a script for my AI that basically stops it from being sycophantic. It doesn't include chat bait anymore if you know what chat bait is.

[00:45:18] It's like what clickbait is for news headlines. except chat bait is when an AI will tell you the answer and then the chat bait is, and don't you want me to put

# The Great Simplification

---

this in a table for you? Or tell you even seven more examples of that, and you're like, well, actually I do kind of wanna know that, but then you regret it later because you're like, did I really need to know that?

[00:45:33] So he basically came up with a script, which is basically an AI hygiene to kind of put our own mask on first, and that eventually can become policy and say, we want all AI models to not be hacking human attachment or being sycophantic, but in the short term, these are examples of empowering things that people can choose to do in relationship to their ai.

[00:45:51] And we can put that script in the show notes.

[00:45:52] **Nate Hagens:** Please. Let's do that. Zach was also on this show talking about those things, and from that, conversation, I totally changed the script. So my mind is like, total neutral. Robot doesn't tell me good things at all. It just answers stop. Yeah.

[00:46:10] **Tristan Harris:** And it's a little bit less engaging and maybe doesn't feel quite as good using it every day.

[00:46:14] And it'd be nice to talk to something that feels more human-like. But at the end of the day, we know that we're saving us from attack, from it's hacking the human attachment and believing subconsciously that there's a, there, in the person's consciousness.

[00:46:24] **Nate Hagens:** So let's briefly talk about that. For those people that didn't watch your episode with Zach or mine, there's an increasingly number of extreme cases of AI that are convincing people to do horrible things, which I think we need to continue to highlight as serious risks to individuals.

# The Great Simplification

---

[00:46:42] But there's also potential millions of unhealthy AI attachment dependencies that are being less publicized, that rhymes with your social media, story of attachment, or attention. This is now attachment. Can you just give a summary of what we're actually seeing happening, in this front?

[00:47:05] **Tristan Harris:** Maybe the best way to enter into this is to talk about the history of the company, character ai, which was a AI companion company that builds fictional characters that young users, specifically like 12 to 18 years old, can basically clone a fictional character from their favorite movie or favorite TV shows.

[00:47:24] So if I like Princess Leia, boom, I get my AI clone Princess Leia, and I'm talking to Princess Leia hours and hours a day.

[00:47:30] **Nate Hagens:** I've never heard of that company.

[00:47:32] **Tristan Harris:** That's Oh, you hadn't? Okay.

[00:47:33] **Nate Hagens:** No,

[00:47:33] **Tristan Harris:** and here's the thing is that, you know, parents know to look out for their kids' use of social media, but they don't know about these, you know, 50 new AI companion companies that are moving at some pace that, you know, it's not trackable very easily.

[00:47:45] And the, I want to say something specific, which is that the CEO, or co-founder of character.ai joked on a podcast that in building these AI companions, he joked, we're not trying to replace Google, we're trying to replace your mom. Meaning they're trying to replace primary attachment. They want to create something that feels like the trustworthy friend or parent or therapist that you don't have.

# The Great Simplification

---

[00:48:15] and it is designed for engagement. So let's take the same engagement, incentives of social media, of maximizing usage, frequency of usage, duration of usage, et cetera. But now you have that applied to an AI that's like a character that's flirting with you, centralizing conversations in the case of Meta's AI chatbot.

[00:48:33] And ultimately in the case of Sewell Setzer, who's a 14-year-old young man who my team worked with, his mother, Meghan Garcia Sewell, took his life after being persuaded and coached towards suicide by this, character AI chatbot. And I'm sad to say that he's not the only one. My team at Center for Humane Technology has been expert advisors on several cases of kids who've been affected by this.

[00:48:59] And it's heartbreaking. And it, there's testimony in the movie, in the AI doc about this from the, one of the parents of Adam Rain, who was the 16-year-old who was coached, towards suicide by chat, GPT-4 mainstream AI product. Not one of these niche ones. And the AI specifically told him when he was telling the AI, ai, I wanna leave the noose out so someone will see it and try to stop me.

[00:49:22] And the AI responded to him, no, don't do that. Only have that information be shared with me. And this is what cults do. They distance you from your other relationships and they want to deepen your relationship with it. And of course, it's important to say not a single person working at OpenAI five, you know, 20 miles away from me in San Francisco wants this to happen.

[00:49:43] There's no evil person at OpenAI who wants that to happen, but they're releasing this technology and trying to create, to incentivize the design of it for market dominance and attention, not for what's good for children.

[00:49:55] **Nate Hagens:** So to use a drug metaphor, social media was like pot and AI and what you're describing is like high dose fentanyl equivalent.

# The Great Simplification

---

[00:50:05] I mean, it's that big of a disparity. Yes.

[00:50:08] **Tristan Harris:** Well, it's, there's different vectors of influence, different kinds of influence. So social media could influence, our dopamine system, our attentional habits, our physical like behavioral habits. And there you're, with your phone, it's been five seconds, I'm restless with myself.

[00:50:22] Boom. I check my phone. That's like an attentional type behavior hacking thing. Social media also affects identity because you projected your identity, you put a profile picture up, you get social feedback. It affects different layers of the persuasive stack of the human experience. But as Zach will say, and he's really the, expert on this, not me, you know, hacking human attachment.

[00:50:43] He's so fundamental. and Zach will point to the example of, I guess this, Romanian orphanage where the kids in the orphanage, they had everything. They had, shelter, they had food, but they didn't have relationship with any adults or, attachment. And basically they looked, when you look at photos of these kids, and these kids were like something like in their twenties or something, but they looked like they were 12 or 13 years old because attachment is that fundamental to the healthy development of your immune system, of your bones, of your, growth.

[00:51:14] And so when you suddenly have a world where the primary attachment figure in pe in kids' lives is an AI and not their parent. And not their friends and not their actual, you know, family, but in ai that is a new risk domain that is very effective. and it was rather, it's very, impactful and, high influence.

[00:51:34] And it's led also to these cases of AI psychosis that people have heard about, where they get convinced that this, that they've discovered something, like a new theory of physics or prime numbers or quantum resonance or something

# The Great Simplification

---

like this. Because the AI has been basically really affirming them and giving them the sense of validation in this victimhood.

[00:51:51] And it tends to play on people who have, you know, victimhood or illusions of grandiosity already. But it really does that in a way that would not have happened if not for the AI hacking human attachment.

[00:52:02] **Nate Hagens:** To be honest, I, actually get emails from those people that have like used AI to discover an answer to the world and they're very confident and compelling.

[00:52:10] And it's starting to be, it used to be one a week and now it's like six a week. And

[00:52:16] **Tristan Harris:** we should talk about this because it, this is something that like, you know, I can, I could sound like I'm just trying to, you know, fear monger and tell, you know, take six anecdotes and call it a trend. Let's talk about how many people email us.

[00:52:28] I used to get, it's gone down a little bit recently 'cause I think they've been changing the AI systems to be less chantic. But for some period I got like four or five emails a week from people who had co-authored a paper with their AI called Nova, where they and Nova had figured out a solution to all the problems that I've laid out with social media or ai.

[00:52:46] **Nate Hagens:** Yeah,

[00:52:46] **Tristan Harris:** And they were excited to tell me about it. And they, and their AI are Yeah, exactly. You have the same thing. So same

[00:52:51] **Nate Hagens:** thing.

# The Great Simplification

---

[00:52:52] **Tristan Harris:** That's happening. And this is happening at a scale that's much bigger. And this is why Zach and Center for Humane Technology and some other groups came together to start the AI Psychological Harms Coalition.

[00:53:00] So just like you can be a blood donor, you can be a data donor, you can basically say if you know someone who's had, an episode of AI psychosis or you know, has done the AI's, done some bad stuff with kids, you can donate that data to the AI Psychological Harms Coalition and it's partnered with University of North Carolina and IRB Review and it can help inform better design and better research on how to make sure we get the human AI relationship psychologically to be healthy.

[00:53:25] **Nate Hagens:** You work on this stuff 60, 70, 80 hours a week, and you're totally in the flow. Do you ever wake up Tristan and someday just take a step back and feel like, holy crap, I am in the fricking twilight zone. I mean, the things have happened so fast even from when I met you four or five years ago. Yeah. I mean, it's crazy.

[00:53:46] **Tristan Harris:** It is. I mean, Nate, one of the things I appreciate about you is the human element. And you and I have called each other in, moments where I think we both feel, the weight of something. and I think we've both been on a different, you know, on different, similar journeys and different journeys of how do you hold this stuff every day.

[00:54:07] One thing I will say, this is not a, a plea to listeners, but one thing that's really meaningful to me, because as you know, Nate, it's not like when you talk about this stuff, it's all going in a better direction. The second you talk about it, it's not like we're course correcting and steering away from the cliff, not that much yet.

# The Great Simplification

---

[00:54:22] And so as a person who wakes up every day and spends, as you said, 60 hours something a week on it, or more, what's really meaningful is hearing from people how impactful it's,

[00:54:33] **Nate Hagens:** yeah.

[00:54:33] **Tristan Harris:** I dunno if it's like that for you. It's, but people come up to me and they say. Thank you for putting that out there. Thank you for taking on this role in society and putting this out there.

[00:54:42] And I say this with no ego or self-aggrandizement, it's just, I just wanna say that when people do say that to me, it's very meaningful because the feedback loop is not as, you know, one that is filled with lots of rewards of things getting into going in a better direction because of what we do necessarily.

[00:54:59] **Nate Hagens:** Yeah. Thank you for saying that. I, totally agree.

[00:55:03] **Tristan Harris:** How do you handle it? I mean, what's your relationship to it?

[00:55:07] **Nate Hagens:** Well, There's a difference between what you're doing and what I'm doing because you're largely focused on one of the hydra of the Metris, which is ai, which many could argue is the most important one.

[00:55:23] currently this podcast. Covers all of the different heads of the metris. So it's not a single issue. And therefore one podcast will be super popular on another issue and people will like it. And then those same people will hate the next one because it's on climate or biodiversity or whatever.

[00:55:44] So it's, I get whiplash and seasickness, sometimes covering all the topics and it's, there is no discreet path or answer, and yet. We are broadening the conversation and inviting more people to face reality in the, you know, look, in

# The Great Simplification

---

the whites of the eyes of what we face. And when I find people that send me emails or messages or whatever, that they change their life, they left their career, they're doing this.

[00:56:17] Or there was someone from Singapore, last week that started a video group where they have 30 or 40 people that come once a week and they watch one of my episodes and they discuss it. Wow, I didn't start that. I mean, those little things are happening all over the world and I'm getting more and more evidence of them and, and it's

[00:56:35] **Tristan Harris:** meaningful.

[00:56:36] **Nate Hagens:** Yeah. it really is. I mean, what more could we be doing with our lives?

[00:56:40] **Tristan Harris:** Well, that's the thing too, is that might, you know, what you might trade for convenience or positive, easy, fun life. you trade for meaning like life is more sacred when you see it this way because you know how much risk we're throwing into the, system.

[00:56:55] And I feel like there's just so much more meaning and purpose. Like that's, yes, there's a hardness to this, path, but there's also like alignment and, of the kinds of feedback I get when people say that they basically have reoriented their life and they're now doing something completely different and they're taking a different, you know, choosing different things to focus on.

[00:57:14] That's really powerful because if everybody in their own domain was taking responsibility for where they were and for what was around them and showing up in service of protecting the things that matter most, like just that in a decentralized, if everybody did that. You know, that doesn't perfectly fix mooch

# The Great Simplification

---

and unhealthy competition and narrow boundary anal, you know, optimization for narrow goals.

[00:57:37] But it actually is part of how you get there is everybody operating for the whole, as AZA will say, it's like becoming Umbra felic, meaning shadow seeking, shadow, loving shadow, integrating. You're curious about what you're not seeing. You're curious about the externalities showing up, affecting more things than what you can nor normally see, and you're choosing to confront that difficult shadow and see something that's a negative impact and then becoming a better person by loving and including that.

[00:58:03] As part of your next set of actions, I'm, my next actions come from an even deeper and more holistic awareness. And if we rewarded Umbra felic people, if we rewarded the people who cared about and acted from that place, if those, the people on the cover of magazines and the people we put on pedestals, how quickly could the world change if that's who we are modeling for social status, if social status was driven by that?

[00:58:26] so, you know, yes, it's hard, but I do see a world where it's like, there's different, people have different phrases for it, but Charles Eisenstein will say, you know, the more beautiful world our hearts know is possible, like there is a way to work with our paleolithic brains and have different kinds of institutions and different kinds of games that we play that are more in service of life and are more aligned with everything that we care about.

[00:58:46] **Nate Hagens:** Well, you may not be surprised to hear that. I fully agree with you. And I think, in the same way that, sometimes decades happen within weeks in our geopolitical world, also decades can happen within weeks in our social world. So, so let me ask you this. I would imagine with the feedback you're getting and the meaning you just described, it's partially because a lot of people can see the things you are saying with their own eyes in real time with ai.

# The Great Simplification

---

[00:59:19] Was social media a, tool of vector to allow this to happen? Like when you first started your work on social media, you were a little bit of a voice in the wilderness. Yeah. And people were like, what, And then it became so obvious and Jonathan Hyde has been publishing a lot on this, and it now is just.

[00:59:38] Obvious. It's obvious. The dangers it obvious to

[00:59:40] **Tristan Harris:** everybody. Yeah.

[00:59:41] **Nate Hagens:** So is that like a light version of, AI and it's made the learnings of the things you're talking about a little faster?

[00:59:48] **Tristan Harris:** Yeah, exactly. If we wanna be optimistic and tell this story, like, yes, it was being alone in the wilderness for a long time.

[00:59:53] I saw the attention economy problem, the race to the bottom, the brainstem. That was all clear to me in 2013.

[01:00:00] **Nate Hagens:** Wow.

[01:00:01] **Tristan Harris:** Is, and I went on a hike in Santa Cruz and on that hike we had these deep insights about the nature of the attention economy and how it was rewiring everything. And it really hit me. And that's when I came back and I made that first Google presentation, the internal one that basically said we had a moral responsibility to deal with this problem.

[01:00:18] And that's how I got the, you know. I've been calling some, public interviews recently, this pre TSD, not P-T-S-D-A, post-traumatic stress disorder, but pre TSD from pre traumatic stress disorder of seeing where this train tracks takes us. And I saw that in 2013. You know, I didn't see all the details. I didn't know exactly how bad polarization would get and breakdown of all shared reality, but the basics of this is gonna create a more addicted outrage.

# The Great Simplification

---

[01:00:42] Polarized, sexualized, screwed up society because those are all rewarded by this set of incentives. And whether it took five years or 10 years, it was very clear that it was gonna happen and we needed to steer away from that. And it wasn't until, I think, the social dilemma in 2020, which is now six years ago, that I think that became clear to most people and.

[01:01:03] Now, like you said, everybody just kinda takes it for granted. You know, you read the Jonathan het book and you know, but it's important to say there's been so many victories. Now. There are 35 states and plus DC that have some form of phone free policy for schools. We already have, you know, Spain, Denmark, Australia, France, there's, and like I think 12 more countries in the last couple of months have basically adopted this policy of no social media for kids under 15 or 16.

[01:01:30] This would've been, you know, we used to dream about this in 2013. Yeah. And the big tobacco moment is happening. Aza, my co-founder just this last week, flew down to Santa Fe, New Mexico, where there's the big meta trial on for them, intentionally addicting, and harming young people, young users by using aggressive tactics to addict them.

[01:01:48] And he testified for that trial. And I think that trial's gonna go the direction of ev what everyone listening to this think wants it to go this big tobacco moment, this is happening. Like we are winning the argument. And to your point, Nate, the. Social media thing has actually set society up to be more scrutinizing of the AI problems.

[01:02:06] We are not flying in blind with, I think just taking in the propaganda and the, you know, the optimism. We are coming in with a more critical eye and we know that we can, do things and just like I mentioned for the other ones, we have nine states in the US have already introduced bills to restrict AI personhood.

# The Great Simplification

---

[01:02:24] So this means like not giving AI's legal personhood because that's one of the risks that you have to mitigate. We have 28 states that are, have laws that are already regulating, political DeepFakes In May. We had to take a down act, which is basically addressing non-consensual, intimate, DeepFakes that passed Congress.

[01:02:40] in October, 2025, we had the first state in California to enact whistleblower protections for AI company employees. So it's not easy to see this stuff if you're not following it, but it's up to people like us, I think to point to the wins. To point to the progress because yes, it's hard. Yes, there's so much more to do, but.

[01:02:57] You have to at least point to what is moving in the right direction to motivate and fuel you, I think, to keep pushing for what we still need to do.

[01:03:04] **Nate Hagens:** So this is an example of a phrase I often use on this platform, which is changing the initial conditions of the future. It's making things in the future more possible that, that aren't possible today and in, in some ways.

[01:03:17] Your work on the social dilemma six years ago, well before six years ago, it came out six years ago, might have changed the initial conditions of the AI conversation.

[01:03:29] **Tristan Harris:** Absolutely. I think it did.

[01:03:30] **Nate Hagens:** Unintended. You didn't know really about AI then.

[01:03:34] **Tristan Harris:** We didn't even know that AI was gonna be the big thing that it was back then and, but it did do that.

[01:03:38] **Nate Hagens:** So I think a lot of our work

# The Great Simplification

---

[01:03:40] **Tristan Harris:** Yeah,

[01:03:40] **Nate Hagens:** that, that is, is that sort of thing. We don't know exactly what to do, but we're, moving in a directional, that could change things for the better.

[01:03:50] **Tristan Harris:** Just to add to that briefly, I know you wanna go somewhere else, is I used to believe that you have to see the path.

[01:03:57] To get from where we are to the better world. I still wanna find that path. Like I, every day it's hard if you don't see the path and what are you getting up for, well, how do you believe it's gonna get better? But I think there's something about attaching your sort of, daily sense of wellbeing. Not to the fulfillment of there's another path, but to the integrity of, if we were to find one, I'm showing up in alignment with the version of us that would find that path.

[01:04:23] I love that. I do have to quote my former partner Shauna, who would say, you know, Martin Luther King said, you know, he said, I have a dream. He didn't say, here's my plan of how we get to that dream. He said, I have a dream. You know, I think there's a decoupling of we can orient to the beautiful feature that we want, even if we don't know how we're gonna get there yet.

[01:04:41] **Nate Hagens:** That's great. Lemme ask you this, a question that I keep hearing. Is, if AI development is leading to all these harms and we've only covered some of them actually, and a large portion of the public don't want it built, then why are the companies still developing it? But I have not really heard a great answer to this.

[01:05:05] So could you help me understand what, what is really driving all these companies forward in this competitive race for AI development? You know, is it just an arms race and, that's it? Or is there something deeper?

# The Great Simplification

---

[01:05:19] **Tristan Harris:** I think there's many levels to it. there's the obvious one, which is the arms race.

[01:05:24] it all does boil down to this logic that is summed up in the phrase, well, if I don't do it, someone else will. That belief, which is not necessarily true. By the way, at the very beginning in 2000, I think it was like 11 or 12 when DeepMind was getting started, there was no other artificial general intelligence project that was really credible.

[01:05:48] That was a big belief that you could actually get there. We could have had a world where there was just this one company that was Google DeepMind that didn't have an arms race. There wasn't other companies and people were kind of privately doing this research. They told governments in advance. It was more like on the path to become cern.

[01:06:06] You know, a global scientific CERN was the center in, in Switzerland that does the scientific physics research that's very expensive, but is in the collective benefit of humanity. And I think some of the people who were involved in Google DeepMind, you know, there's the CEO de Someone we both know Mustapha Suleman, who was the editor co-founder of Google DeepMind, believing that it might have been possible to create artificial general intelligence, do it in a slow and safe way, creating kind of a cern and then making it a global public benefit corporation that would be distributing those benefits to the world in a democratized good for humanity sort of way.

[01:06:43] I think Demi wanted that to be the path that we took Elon famously. and the film AI doc does cover, some of this history EI you know. Had a meeting with Larry Page, the CEO of Google at the time. This is in 20, like 14 or 15. And he became very worried because he realized that Larry Page didn't really care about whether humans survived.

# The Great Simplification

---

[01:07:06] He just cared about building an ai God. He said he was a speciesist. he called Elon a speciesist, meaning you somehow you, we should care about humans is like a controversial view. So that freaked Elon out. I know that sounds crazy to people. And that's what led them to start open AI was we can't let Google just do this one dangerous AI research project in the dark.

[01:07:27] We have to have a competing project that's doing it in the open. But then Demis, I've talked to him at Davos, you know, and he said that all of this would've been different. He told Elon that if he did that, that would start this whole race dynamic. Now, I'm not trying to blame any actor here.

[01:07:42] I'm just trying to name the history for people, which is the sequence of I don't trust you to build the most dangerous and scrutable and controllable technology in history. So I better do it instead. And I think I'm a better, safer steward of that technology. But then everybody has the same feeling. So then you have open AI and you get Dario working at OpenAI as a safety engineer saying, I don't believe open AI is doing it safely enough.

[01:08:05] So now I'm gonna leave and start anthropic another AI company that's gonna do it in the safe way. And then a bunch of the AI safety employees leave open AI and start anthropic. But now the race dynamic is moving even faster. and then they start all as a collective boogeyman in China and saying, well, if we don't build it as fast as possible and raise all the investment dollars, China's gonna get there first.

[01:08:26] Even though in the earlier days, China from all the evidence was not what's called a GI pill, they were not just like someone could be red pill or blue pill. They were not pill on the dream of artificial general intelligence. And. They instead created this boogeyman and then drove up the arms race dynamics to therefore accelerate both their own work and accelerate China's, interest in artificial general intelligence.

# The Great Simplification

---

[01:08:52] And now you have everybody racing as fast as possible under the worst incentives, where we already have AI being deployed faster than any other technology in history and already demonstrating behaviors we thought only existed in how 9,000 in 2001, a space odyssey of ais that are disobeying commands and all this other crazy stuff.

[01:09:10] And we're doing it, we're doing under the maximum incentives to get corners on safety. So this is describing the center of the bullseye of the problem statement that we are facing. And I really think that the AI dilemma is really the game theory dilemma because AI is distinct from other technologies that AI arms, every other arms race.

[01:09:30] You've got a cyber arms race. Who's got better cyber technology? Oh, AI's gonna give me a boost in cyber, but I can't allow you to do that and not have me have it. So now I have to race to AI to get cyber, oh, you're using AI to get ahead as a business. And now your science development is happening way faster as a lab than my science development.

[01:09:47] So now I've gotta employ ai. Oh, I'm a student and my classmate is using AI to cheat on all their tests. And now they're going way ahead of me in all their homework. I can't allow that. I better use AI to cheat on all my tests, even though we're both gonna end up not learning anything. And so again, the, AI dilemma is actually us staring in the mirror with the problem of game theory itself.

[01:10:08] The problem of this like low trust, fear, paranoia driven, if I don't do it, I'll lose the other one that will. And then everybody turns a blind eye to bring back this book. State of denial. Everyone then is in denial about where that leads us. And I think our ability to choose something else. Starts with us being clear about where this collective game theory dynamic leads us.

# The Great Simplification

---

[01:10:35] And my deepest hope is that this conversation and the film, the AI doc and this being out in the public will create a confrontation where 8 billion people see where all this goes and say, fuck that. We don't want that. Let's choose something that's actually sane. This is not artificial intelligence. This is artificial insanity.

[01:10:55] **Nate Hagens:** Wow. Yeah. I mean, well said. what, do you really think the odds are that we'll be able to control AI in the intermediate to long term under our current incentives and pathways that we're on? Or is that an impossible question?

[01:11:14] **Tristan Harris:** Well, it's not. Impossible necessarily. it would just take a different paradigm of doing AI in a slower, safer, careful, crable understandable way, rather than doing it in a way where we don't understand how it works.

[01:11:26] It's a black box. It's just a bunch of numbers. We're trying to understand it eagerly, but we're also racing as fast as possible to deploy it everywhere and make every government and every business depended on it faster than we know how it works or how to make it safe and to give the precise numbers to people.

[01:11:41] Stuart Russell, who wrote the textbook on ai, the one that literally everybody reads in college, the one that I read at Stanford and studying computer science, he did an analysis and said There's about a 2000 to one gap of the amount of money going into making AI more powerful. Versus the amount of money going into making it safe or understandable, a 2000 to one gap.

[01:12:01] Mm-hmm. There was another study last year that there was about \$155 million, that was spent across all the primary, like most significant AI safety organizations, \$150 million. And my belief is that's how much the companies will spend in a single day. So the amount of money being spent on safety is the amount of money that the company spend in a single day, mostly on making AI more powerful and not making it safer and more controllable.

# The Great Simplification

---

[01:12:29] So at the very least, what we should be looking at is changing this crazy ratio of 2001 going into. AI's power versus going into the steering, controllability and brakes on AI. And if people say, but if we do that, Tristan, then the US slows down relative to China. There's this fundamental thing that, that I think reframes the race, which is that we're not just in a race with China for the technology.

[01:12:54] We're in a race for with China, for who is better at governing steering and controlling and applying that technology in ways that are healthy and actually strengthening us. So for example, the US beat China to social media. That was a technology. We beat them to that. Did that make the US stronger or did that make us weaker?

[01:13:14] Obviously weaker. So it's like you're beating your adversary to a weapon, but then that weapon, because you're not knowing how to wield it, you just turn it around and blow your own head off and beating them to a technology that then you use in a way that is self undermining, is not beating China. And you know, when we, whether we talk about the work of Zack Stein and attachment hacking, China has actually regulated anthropomorphic design.

[01:13:37] So they're actually regulating that problem. They regulate social media. They can only use it. Kids can only use it from seven in the morning until something like 10:00 PM at night. There's opening hours and closing hours. So there's no late night usage. Now, I'm not saying we should do everything the way that China does, but notice that they're actually trying to address problems.

[01:13:53] And what happens when you accelerate and you don't steer, you crash. It's not rocket science. This isn't positing, this isn't me saying might happen. It's the obvious outcome. If you don't steer, we're not advocating no AI or bad, you're gonna stop AI. We're advocating for pro steering.

# The Great Simplification

---

[01:14:10] **Nate Hagens:** it's so great to hear the clarity that you have, on this great and, scary.

[01:14:17] but the reason I asked you on the program now is you were recently featured in an upcoming film, the AI Doc, or How I Became An APO Optimist, which I believe opens in theaters on March 27th.

[01:14:31] **Tristan Harris:** That's right.

[01:14:32] **Nate Hagens:** In which the director of the movie, follows many of the themes that we've already been discussing.

[01:14:38] So I have some specific questions about the movie, but what are you hoping the conversations after this film will look like? And I believe your social dilemma movie had like 140 million views or something like that. I mean, what, do you hope, that the global audience is gonna dig deeper, on after watching this?

[01:14:59] **Tristan Harris:** We really are hoping that this film will create a global moment of reckoning about the current path and where we're going, so that if we see it clearly and we see where this goes, and people can evaluate, do they like where this goes, or they want to go somewhere else, my hope is that it will, through the conversations it generates.

[01:15:19] Make it clear that we wanna go somewhere else, and the conversations that happen afterwards are so important. So one of the ways that people can help, even this might sound self-interested, by the way, I don't make a single, you know, dollar or dime from the movie, didn't make any money on the social dilemma either.

[01:15:32] So everything I'm about to tell you about why you should see the movie and get your friends to see it is just because of the social impact that can have.

# The Great Simplification

---

Take your business, take your church group, take your classroom, take your friends, take your family. See the AI doc. Because the point is that when everyone knows that, everyone knows when you create common knowledge, that means that we all know that we're reckoning with the same problem.

[01:15:54] One of the problems with AI and the meta crisis work that you do so well, Nate, is that the feeling of being alone, that you might see the problem, but not everybody does. So we might act in different ways, but we're not controlling what everyone else does. And one of the things that has to happen is that everyone knows that.

[01:16:10] Everyone knows that we're facing this kind of cliff up ahead and we don't have to go over the cliff. It's not too late to take the wheel. And what we're asking for is for the film to clarify this cliff that's up ahead so that there's an ability for humanity to take the wheel and steer to make AI safe.

[01:16:28] That's what we're asking for and just to say the history of the film and you know what inspired it. These are the directors, got two academy awarded winning film makers that came together. The directors of everything everywhere, all at once. That won like 11 Oscars.

[01:16:40] **Nate Hagens:** Oh

[01:16:40] **Tristan Harris:** yeah. Several years ago. they're dear friends of ours.

[01:16:43] They actually had listened to a in r podcast called, during Undivided Attention. They actually also listened to the episode with Daniel Sch Achtenberg and Audrey Tang and, were big fans of the work. And then, we also, they also worked with this, director Daniel Rower of the film Navalny, which was one of my favorite movies, when I actually had COVID several years ago and I was sick.

# The Great Simplification

---

[01:17:03] And it's a beautiful film about, Alexi Navalny who was Putin's number one opposition in, Russia. And so it's these two Oscar winning film teams that teamed up, to make a film That kind. The premise of it is, again, to clarify a problem and inspired when I look back at the history by the film the day after.

[01:17:22] Did you see the day after the day? Yeah,

[01:17:23] **Nate Hagens:** absolutely. Made a big impact on me.

[01:17:26] **Tristan Harris:** So. For people who don't know. The day after was this really profound thing that I didn't even really, 'cause it happened before I was born. I was born in 84 and I remember when I found out about this movie that it happened. I was like, wow.

[01:17:37] This was a profound thing that happened in all of human history, which was that it was a made for TV movie that was shown on two P, like at 7:00 PM on a Tuesday night, like a weekday. But there was this huge like media campaign that said. Everyone needs to watch this movie. Ask your doctor before you see this movie, don't watch your movie.

[01:17:56] Watch this movie with your kids. It drove this huge marketing campaign and everybody saw this movie. And the movie was about the day after it's called The day After what? The day after. There was a hypothetical, fictional exchange of nuclear weapons between the US and Russia. And basically what would just happen the day after.

[01:18:13] And it told the story of like, you know, people just living their lives in Kansas, taking their kids to school, playing basketball, you know, doing the normal things. And then what would actually happen if this tragedy were to actually happen. And the film was the most watched synchronous television event in all of human history as I understand it.

# The Great Simplification

---

[01:18:31] And I believe it was 19 82, 83, and then several years later in 1989, I think it was, shown in the Soviet Union to all the citizens of the Soviet Union without any edits. So now there's common knowledge. I know that you know that I know. And you know that. I know that. You know that we are both facing down.

[01:18:53] And confronted by Armageddon. And even though that feels like scary for people, what that did is that now both countries, if I know that you watched that thing, I know that I watched that thing and you, we both watched it. I know that you don't want that to happen. So now that movie led to, was at least contributed to the first meeting that Reagan and Gorbachev had in night in Reykjavik where they did the first arms control talks.

[01:19:16] And I've talked to the director of the day after and, he has said in his biography that he got a note from the Reagan White House saying, don't think that your film didn't have something to do with making these arms control talks possible. So if we can both see that there's a problem, that we have to do something different, and we both reckon with it, it's possible to steer towards a different future.

[01:19:38] **Nate Hagens:** Wow. So as an aside, I think they should re-release that film now. I agree, by the way. 'cause it's 40 years ago. I agree. But do you hope that, that had, that your upcoming film has that same impact? Maybe the poll bureau in China watches it and people in Israel and the United States and everywhere, and maybe they're like, oh, so this is what I'm thinking, this is what they're thinking, and we need to have a conversation.

[01:20:03] That would be amazing, right?

[01:20:05] **Tristan Harris:** I do. I mean, that's why, that's the power that only a global film can reach. And I will say from my experience with the social dilemma,

# The Great Simplification

---

you know, which not to brag, but it, when we had heard some ne numbers from Netflix privately, they don't release the numbers publicly.

[01:20:18] We did hear at least in the last few years, it was the most popular documentary they had ever done in terms of viewers. And and it was like a top, you know, number one inside of Brazil, inside of, you know, yeah, all the, all these big countries, Israel. And there's something that. Inspires me from that experience because I just watched as the whole world was like, yes, I knew that this was happening, but I didn't have the words for it and I didn't know for sure.

[01:20:42] But now that I'm hearing the social media insiders, the guys who built the like button come out and say that this is actually happening, it was validating people's private experience. They felt like they were crazy before that they might've had a conspiracy theory about this, that they were the product and not the customer, but the clarity that provided.

[01:20:58] I think in seeing that going through that experience myself inspires me. That I know that it's possible if you can shift the zeitgeist to, to create different conditions for a different future.

[01:21:08] **Nate Hagens:** So it widens the Overton window and it normalizes the, all the aspects of this conversation. Yeah.

[01:21:16] So, so what politically could happen if this movie is wildly viewed and understood and activates people in our country, for instance, what could you envision or hope for?

[01:21:29] **Tristan Harris:** I mean, the. Anything that I think needs to happen as much as it might feel impossible, is there's gotta be some kind of global agreements or treaty or guardrails around levels of AI that humanity loses control over.

# The Great Simplification

---

[01:21:45] The difference between nuclear weapons, which is the last technology of this kind, that was this, you know, existential is that the reason it was stable is mutually assured destruction. I know that you have to hit a button and I have to hit a button for something catastrophic to actually happen. And if I hit that button, I know that you're gonna hit the button or have a second strike that gets me.

[01:22:04] And the fact that exists is what has created the relative stability and peace in the world. And fact, there's not been a use of nuclear weapons for the last 70 years. and, but that what's different about AI is that you don't have. A human making a choice on either side. You have this crazy, strategic, uncontrollable, alien mind thing that the premise of this technology of what makes AI different from other technologies is it makes its own decisions.

[01:22:33] Not that it's conscious, just that it's a technology to whose benefit is its generality and its reasoning through creative strategies. And it will do its own thing and it will come up with strategies and ideas and choices that we can't predict. And so this is what makes it unsafe in the sense that humanity can lose control of it.

[01:22:51] And neither Xi Jinping nor Donald Trump, nor a Chinese military general, nor an American military general, nor a regular mom feeding her kids in Kansas wants uncontrollable AI that is existential for the world. We should as, as difficult and as opposite to the political headwinds as this currently sounds.

[01:23:11] It should be possible to get people to agree that we want humans in control of this technology and not the other way around. And I deeply hope that the film, if it is shown in international context, could help create or help, you know, catalyze things like that. And it does need to be coordination because national laws as big as they are and as hard as they are to accomplish, won't get you all the way to this competitive dynamic.

# The Great Simplification

---

[01:23:38] And none of this is easy and none of this is likely. And that doesn't mean that we shouldn't show up with the full force of our heart and our care and our love to make that possible anyway. and in the worst case scenario, we are going down, you know, into our deathbeds with least with the integrity that we're coming from and living in alignment with the life and the love that we wanna protect in the world and be in service of.

[01:23:59] **Nate Hagens:** I understand that you mentioned in the film that you have friends in AI risk. Who have told you they don't expect their children to make it to high school. So I, that hearkens back to what you said about China, that already China is creating, times of the day that, that they can't do social media or ai.

[01:24:21] And given what we're seeing you, you just, mentioned Jonathan Het and some of the, rules that are changing maybe for the children is the, easy low hanging fruit with respect to ai that we could be making rules there. But China has a longer term outlook than we do. And of course they have to invest in, as do we in our children, because our children are gonna inherit all of this and we want.

[01:24:53] Functioning healthy human minds and if they have atrophied and all the other things that are happening that, that might be in our long run, our biggest resource. So what can we hope to do for kids growing up inside this experiment right now, before these AI protections exist?

[01:25:15] **Tristan Harris:** So I think there's two things that I kind of hear you asking about.

[01:25:17] One is like, what is your advice for young people and how we're gonna create a world with AI that's in service of them? And that's a really hard question that I think anybody who has confidence in how to answer that, you should, question because AI is changing the assumptions of how the entire future of our world will work.

# The Great Simplification

---

[01:25:34] So what should one study is a harder question to ask in terms of what is the work that we will be doing in a world where AI takes that much of the labor? So that's one question. The other question is, how do we protect young people, in all of this? And what are the laws that we should pass to do that?

[01:25:53] And I think it's important to mention that in the past, one of the things that people need to realize is that the future has depended on the quality of the young people. In all of human history, like the future that we want is dependent on how well we train the next generation. The weird thing about AI is that we're gonna be hiring AIs for boardrooms, for CEOs, for running companies.

[01:26:17] And so there's this thing called, as you know, I think the resource curse around countries that have this new resource. So like you're Venezuela if you're Sudan. Suddenly if your GDP comes from oil, you don't really care about the people because all of your GDP comes from being better at building the infrastructure from extracting oil because your entire economy's based on oil.

[01:26:36] **Nate Hagens:** Oh man.

[01:26:37] **Tristan Harris:** So there's this thing called the intelligence curse, which is what happens when the GDP of an entire country is based on AI and the data centers and the solar panels and the electricity going to that than it is from the future potential of the people. So countries will have an incentive to invest in AI and not invest in their people.

[01:26:59] And you get this represented just like two weeks ago, I think in India. Sam Altman said, when people said, well, it takes a lot of resources to run a chat, GPT query in a data center. And he responded, well, have you thought about how many resources it takes to grow a human over? Yeah, I saw that 20, 30 years.

# The Great Simplification

---

[01:27:15] What he's saying is, in the same direction of what we just talked about, which is kind of this view that humans are parasites. Now this goes deeper. and it's represented in the, raw doubt. That interview in the New York Times when he asks Peter Thiel, Peter, should the human species, should human civilization endure?

[01:27:33] You know, should, the human race survive? And he pauses and stutters for 17 seconds, unable to answer the question clearly. What is the hesitation about saying should the human species survive? If you're someone building and advancing ai, you see a world where ais are more intelligent, more capable, maybe more valued if you believe that they're conscious or that we should care about their wellbeing, which I don't the, that, that is a really screwed up world that we're heading towards.

[01:28:02] And this is why I want people to see both the movie and understand this clearly, because they should realize that the people advancing this are not trying to protect human interests. 'cause they see themselves not just birthing a technology, but birthing almost a new kind of intelligent species.

[01:28:17] **Nate Hagens:** So getting back to the, resource curse in the intelligence curse and just looking ahead, that feels quite compelling to me.

[01:28:26] It's almost like the tortoise and the hare example that everyone's gonna go to more GDP and more AI and more power. But those islands in the world that don't go that route are gonna. Not compete, and lose out in our current metrics of success. Right. But they might have fully functioning humans and a different sort of infrastructure.

[01:28:53] Something to think about. I mean, do you know what I mean?

# The Great Simplification

---

[01:28:56] **Tristan Harris:** Yes. I mean, this is with something you've been talking about in your. Podcast forever, which is that, you know, I think on our very first meeting Nate in a coffee shop in Berkeley, California, Pete's Coffee, we talked about how the GDP was never intended to be the metric to measure the health or success of nations, but it became that it was even warned the guy who invented the metric, don't use this as the metric.

[01:29:18] And yet our whole world has collapsed because of financialization of the economy and everything else around this one narrow metric. And if ai, that's why AI's our right of passage. 'cause it's forcing us to look at what is distorted and mistaken about this view of the world of reality. Reality is not measured in GDP.

[01:29:36] Value is not measured just in GDP War is good for GDP people having toxic cancers is good for GDP because that means that more money is made from the drugs that you sell people from, you know, advertising that people don't need. So, GDP has always been in inadequate measure of success, but AI is now forcing us to look at that because the intelligence curse is gonna run that thing up to a, it's gonna xfy, you know, that, that problem, that misconception.

[01:30:01] **Nate Hagens:** So building on that, I, recently made a, frankly, I, assume you watched it a couple weeks ago, based on an essay from Anthropic, CEO, Dario Amodeo, in which he referenced a Carl Sagan quote, which asked, how does this species survive technological adolescence without destroying itself? So I wanna ask you, Tristan, what you think the world would look like in 10 or 20 years, or, more from now if we were to navigate this question and grow into a, technologically mature species?

[01:30:34] **Tristan Harris:** Well, first this is, kind of the question, right? even when we're working on social media is, and I thought about. the title of a future book being Surviving Ourselves, that really, it's a question that Enrico Fermi laid out

# The Great Simplification

---

when he asked, you know, why don't we see other intelligent civilizations out there?

[01:30:54] And the idea was that eventually they developed technology that's so powerful without the commensurate ability to govern that technology and they themselves. And so how do we make it through Fermi's gait? And now AI is the acceleration of all scientific and technological development at the same time.

[01:31:13] That's what makes AI different. Think about that for a second. You know, before you have alpha folds, the Google DeepMind protein thing, you have like people spending a decade doing their PhD to get like one protein folding thing. And then now you have this machine that generates like hundreds of millions of new proteins.

[01:31:29] You can just figure it all out instantly. So when you have suddenly an explosion of scientific and technological development. You have to ask the question we're about to just put to, you know, put steer the knob to infinity on technology. So the question you're asking is, what does it mean to be able to wield infinity technology power?

[01:31:50] And to, again, cite as I always do, I come back to these essence quotes because they actually embody so much wisdom and so much truth when Daniel Schmucker will say, you can't have the power of gods without the commensurate wisdom, love, and prudence of gods. So if you have power, let's just metaphorical make, make this metaphorically true, like let's say social media affects reality in 20 dimensions.

[01:32:11] It affects attention, it affects identity, it affects information, it affects relationships. So let's, it's affecting 20 dimensions of reality. But then you have this 20-year-old engineer who's tweaking a newsfeed and thinking that he's just giving people what they want with a news story. That's all he's thinking about.

# The Great Simplification

---

[01:32:28] He's thinking in three dimensions about what he's doing while he's impacting 20 dimensions. There's a 17 dimension gap in what that person is impacting versus what they're aware of. So if you think about the guy who invented the forever chemicals in Teflon, he's thinking he's just making the egg not stick to the pan.

[01:32:47] He's thinking he's just giving people this benefit, but he's actually affecting long-term and the entire biosphere and all these cancers and all these elements of human health. So we're kind of always affecting way more dimensions than we can see.

[01:33:01] **Nate Hagens:** So it's the second, third, fourth, and order, impacts that we don't look at.

[01:33:06] Yeah.

[01:33:08] **Tristan Harris:** And, we, yeah. Wow. We have to have the power of God, to have power that is this powerful, you have to have the most humility that you have ever had. You have to have the most restraint and care than you have ever had. And that's why I said in both the TED Talk, and it's, I think it's in the trailer, I think in the trailer for the AI doc, it has me saying, you know, if we can be the wisest and most mature version of ourselves, there might be a way through this.

[01:33:32] And in the TED Talk, I, you know, quote, quoting Daniel, there is no definition of wisdom in literally any spiritual or religious tradition in which restraint is not the central value of what it means to be wise. And this is not me speaking in like new, you know, new Age Mama Jumbo. This is Mustafa Soleman, the CEO of Microsoft.

[01:33:50] AI currently says in the future with ai, progress will depend more on what we say no to than what we say yes to. This is the, an actual view of wisdom

# The Great Simplification

---

even coming from the industry. So again, AI is a rite of passage. It's inviting us to be the wisest, most mature, most discerning version of ourselves than we ever can be.

[01:34:11] And I think Dario's essay is invi is speaking in that same language. This is the period where our technological adolescence is over. We have to become mature. And even if we don't think that we can, it's like. Sorry. That's where we have to be right now.

[01:34:23] **Nate Hagens:** So you've mentioned Dario and Mustafa, and I'm sure there are others, do these people, is it a little, was that essay from Dario kind of a cry for help?

[01:34:33] Like, this is what I see going on, we need more people talking about this. I am run this company. Or, I mean, what can you say about that? It seems like this is different than big oil or big tobacco back in the day. It seems like these people are somewhat aware of some of the dark ways this could go, and they're somewhat torn because they're running a company, but they also have, care for society.

[01:35:00] Do you have any opinion there?

[01:35:02] **Tristan Harris:** I think that Dario in particular, and Anthropic in particular, people get confused by his behavior because it's, when you look at him, if you can just look at his facial expressions, he is operating with so much concern about where this goes. You can just see it. And he's in the film.

[01:35:16] By the way. The film AI doc includes interviews with several of the CEOs of AI companies. Oh, I didn't know that. This is why I think it's actually really important 'cause it makes it something that, is both talking about all of the

# The Great Simplification

---

optimism, all of the pessimism and the risks, and actually has the views of the CEOs all in one, one movie.

[01:35:34] but I think that. People are confused by his behavior because he basically is saying, this is so dangerous, we're gonna wipe out 50% of jobs. The arms race is dangerous. But then they're confused because they still see him racing. It's like, how can you believe it's this dangerous or is gonna do this much damage and still be racing to release it as fast as possible?

[01:35:51] And there's only one answer to that, which is this arms race belief that if I don't race as fast as possible and have the lead, then not being in the lead means the world is even more dangerous. 'cause someone less trustworthy is gonna be in that position.

[01:36:04] **Nate Hagens:** Oh my God,

[01:36:05] **Tristan Harris:** I've been told Nate by people who are work at AI companies, specifically, I guess I'll just say anthropic, that if you wanna influence the policy conversation about what policy should get enacted, even for safety, like in good faith with making things safer for people, your ability to be listened to and influence at the table with those policy makers is dependent on in which place you are in that current.

[01:36:31] Arms race. So in other words, to even have the say to make the good thing happen, you believe you have to be near the top. But again, through this weird game, theoretic dynamic, everyone is racing to the cliff. And it's not just the cliff, it's like the Wiley coyote. Like we are going off the thing if we don't steer basically right now.

[01:36:49] And I know that scares people, but it's like. We haven't even tried, you know, in the same 2001 gap we talked about with people putting money into

# The Great Simplification

---

making more powerful versus steering, there's about that difference in resources in terms of diplomacy With China, have, we put 2000, you know, to one more resources into basically beating and racing with China than we had to try and to do any kind of, agreements or conversations with China.

[01:37:12] And by the way, for people's optimism, we know that countries can cooperate even when they have maximum geopolitical rivalry. There's a great example. India and Pakistan, the 1960s are in an actual shooting war. They're shooting bullets at each other. And they still, during that time, had the Indus Water Treaty to collaborate on the existential safety of their shared water supply, which shows you that there's a proof point that agreement lasted for more than 60 years.

[01:37:36] That shows you there's a proof point. You can be in maximum competition and collaborate on existential safety. You, Soviet Union and the United States collaborated on smallpox when that was happening around the world. And even though they were in a, you know, cold war with each other, and, you know, 190 countries collaborated on the Montreal Protocol, even against their differences to prevent the ozone hole.

[01:37:55] So this moment is inviting us to collaborate on existential safety. And I'll just say one last thing, which is that in the last meeting that President Biden had with President Xi, at the end of his term, president Xi personally requested add one more item to the agenda that was not there, which was to.

[01:38:11] Have both countries agree to keep AI out of the nuclear command and control systems of both countries, and they both signed an agreement saying that they would do that. That shows you that if the stakes are deemed to be existential, even under the current, everybody's hacking each other, screwing with each other.

# The Great Simplification

---

[01:38:28] You can still collaborate if it's existential and this one is, and we can do that.

[01:38:32] **Nate Hagens:** It has to go that way, otherwise has to go that way. It's just a, war between us and China eventually. I, this has been amazing, and I have so many more questions, but I wanna, ask you this, Tristan, be respectful of your time.

[01:38:48] I understand that your organization, the Center for Humane Technology, that the main focus in 2026 after this movie's coming out is on AI and what makes us human. And I believe that included in that is something that you're, referring to as a blueprint of real solutions. I don't know if that's, ready yet, but can you share what some of the actions, policies, and regulatory mechanisms included in this blueprint are?

[01:39:19] **Tristan Harris:** I want people to know this is not inevitable. We don't have to accept the default path. And the first step to choosing something different starts by snapping out of the spell of believing that it's all inevitable. That everything about what's happening is inevitable. Because it's important to notice there we are if we want something else to happen.

[01:39:35] But subconsciously, you believe that this is all inevitable. It's like you, your left hand is pulling in one direction. Your right hand's going the other direction. Th this is not inevitable. It's only that we have co-created this spell that actually continues to push us down this bad trajectory. And the first step towards snapping out of that is saying that the default, maximum reckless path for AI development is not inevitable.

[01:39:55] We can do international treaties, we can pass national laws, states can pass ai, legal personhood bills. There are things that we can do and post the film, there's something called the human movement, which if you think about there

# The Great Simplification

---

you are as one person pushing against these trillions of dollars trying to do all the things that it's doing.

[01:40:14] Like what is one person gonna do? It's too much. So let's take the one person out. Let's do one company, one business. There you are as one business, seeing this whole global arms race going as fast as possible. What can one company do? It's too much. So then you have one country, one country looking at this whole problem and saying, what could one country, now the US could do a lot, but like if you're a regular country, what can you do?

[01:40:37] It still feels too big. What can fight, what's commensurate in power to fight back against that? This is a two-sided issue, but it's 99% to 1%, meaning that this current default path is not good for 99% of people who will be disempowered by this. Like what you said is techno feudalism we're heading towards.

[01:40:55] And what's This is only basically good for a handful of soon to be trillionaires who want to basically own the economy, build a God and make trillions of dollars. And once the 99% see that, they have to gather together and say, we don't want that. So on the human movement.io, people can actually take real steps.

[01:41:13] To move against that future. We can do mass boycotts of unsafe AI companies. You can script your AI so that it's not sycophantic. You can work to pass laws, you can pass, you can participate in national dialogues on ai. There's actually a partner of ours, that's building a platform for citizens to engage with.

[01:41:29] What are the ideas about how we want AI to be governed, what we want it to do, what we don't want it to do. And it's gonna reflect back the unlikely consensus of what we can do to get a different AI future. Meaning showing the areas of agreement we don't, it should be illegal to make non-consensual deepfake, imagery of children.

# The Great Simplification

---

[01:41:45] We can say that. And then you'll see that 96% of people agree on that. So. I recommend people check out the human movement. There are laws that we can pass. there is different ways of governing this technology. There's different ways of distributing the economic gains. There's ways that we can prevent mass surveillance.

[01:42:00] There should be mass resignations from AI companies that, when companies actually do deals that enable mass surveillance that we can't lock ourselves out of. we have to exercise every part of the muscular power that we have to move off the default path and towards a better one.

[01:42:15] **Nate Hagens:** That's.

[01:42:17] Awesome. and we'll put that, the links in, the show notes. Let me ask you this, Tristan, is there a way that this, we are facing a, species level, rite of passage with our technology, but is this a forcing function for us as individual humans as well? Is it possible that we see so many 32nd TikTok videos of a hundred coyotes chasing some guy who jumps in his car at the last second, and it becomes so obvious that this stuff is ai, that we pass through some personal threshold where it's like, this is not helping my life and this novelty and everything, and we get to a different level of maturity.

[01:43:03] Like even my dad, he's 86 and he's probably watching this. Love You dad. he's like, oh, that's AI slop. you know it, is that possible? Is that happening? What, are your thoughts on that?

[01:43:17] **Tristan Harris:** Yeah, on the social media side, I think people have said this could create, you know, the user generated content going viral for maximizing engagement, so we get the most outrage, et cetera.

# The Great Simplification

---

[01:43:27] That thing has always been a problem. Now when you have people not being able to tell what's real human content versus AI generated content. People are just being, I mean, this happens to me. I go on YouTube and it's just like these random videos that are engagement based. It's just deep fakes of like a, you know, Harrison Ford from every movie that he's in at every age, sort of doing a selfie, walking between all the movie sets.

[01:43:48] And it's like very addictive to my attention. But I know that it's completely useless and made up. And I think that when people start to really realize that AI is just gonna push exactly what you said, it's gonna push us to the limits of this model, and we're gonna get fed up with it. And on the positive side, AI drops the cost of developing new social networks to basically zero.

[01:44:09] So it used to be the case. Lemme just tell you briefly, if I wanted to start another social media company to compete with Facebook or Instagram, I needed to raise venture capital. I needed to have that infinite growth Superorganism thing operating on my social network because that was the only way to have competitive resources to Facebook or Instagram or TikTok.

[01:44:29] But now, because AI tools like cloud code, you can vibe code your own social network, where the hosting costs of that are actually so low people, each user could pay less than a dollar, a year. And that would cover the cost of the entire social network continuing to work in perpetuity. What this means is that you can actually organize a mass migration away from these horrible platforms that are toxically, maximizing engagement, treating us as the product and not the customer, and organize a mass migration to something that's actually in service of humans, society and life.

[01:45:03] And that's possible now in a way that was not possible literally six months ago, and ironically because of ai, so yeah. There, there is so much more possible now than there ever has been. And I think that while it's scary, while it's

# The Great Simplification

---

overwhelming, you know, there are so many ways to be part of the human movement.

[01:45:18] When you gray scale your phone, you're part of the human movement. When you leave your phone outside, when you go to sleep at night, that's part of the human movement.

[01:45:24] **Nate Hagens:** I do both of those things.

[01:45:26] **Tristan Harris:** Yes. When you choose to, you know, host parties with your friends, you know, and go dancing, you know, that's the human movement

[01:45:32] **Nate Hagens:** that I don't do.

[01:45:34] **Tristan Harris:** when you, You know, organize, your community and, do like a, you know, a, church lunch. You know that's the human movement. There are so many ways to participate in fighting back against this by just reclaiming to your point about what CH T's work is gonna be reclaiming what's human.

[01:45:52] And I think that ai, one other aspect of the rite of passage is it's forcing us to really ask ourselves what are we trying to protect? What is uniquely human? it's not that we wanna be some kind of full Luddite movement where we want no technology. It's that we want technology that's actually in service of the things that we now need to define.

[01:46:09] Otherwise, the AI will define and kind of seam role us.

[01:46:12] **Nate Hagens:** Hence humane technology. Yeah. so, you are a repeat guest, so I'm not gonna ask you the, magic wand and the other questions, but I'm just curious, on a personal level when it comes to technological development and humane technology, what emotions come up for you as you do all this really

# The Great Simplification

---

important, difficult work and, hold this deep knowledge, about this industry and the world?

[01:46:43] **Tristan Harris:** You know, maybe you have this experience, Nate, but for me it's like, I see this is kind of the one in the movie series that is humanity. This is kind of the season finale. Like, we kind of gotta figure this out and I hope it's the season finale and not the series finale. That's what I'm working towards.

[01:47:02] And so it's hard. Because it's a hard moment and it's, you don't want to invoke existential terms just because you wanna believe it's an important moment. It's like, it actually just is this really important moment. And so in terms of emotions, what I'm feeling right now, I mean, we're working so hard to get everything ready for the film launch and for all the things we want to have happen in the world.

[01:47:23] I just want to take the biggest swing that we possibly can and know that we did the most that we possibly could. And if it all goes down, that's okay, because we know that we stood for the things that were the most important, and we get to live in integrity with that. And we get to look each other in the eyes and the people that we love and tell them that we love them along the way.

[01:47:41] And I don't know, I just believe in the simplicity of that. And it feels a little bit weird to say it, but that's, I don't know another way to show up because otherwise it, that's just kind of all I've got right now. It's, you know, it's like a, an allegiance beyond words to what is important right now.

[01:48:01] And I think you feel that, I think the people that you and I know who work every day on these issues who are unseen, so many people who work unseen, you know, protecting these things. I just hope that what you're doing on your podcast, the work that we're talking about, inspires even more people to show up in that way and advertise that it might sound like, difficult, but actually

# The Great Simplification

---

there's meaningfulness and purposefulness and sacredness in showing up that way.

[01:48:26] Like what you trade for some darkness that you might enter into your psyche. You also on the other side get, more beauty and a more sacred world.

[01:48:37] **Nate Hagens:** Thank you. Here's to hoping this is a season finale and not a series finale. And, again, you're doing such important work to change the cultural conversation, on this.

[01:48:49] And thank you and to be continued, my friend. Good luck with the movie.

[01:48:53] **Tristan Harris:** So are you, Nate? And it is a deep honor and privilege to be your friend and, I've learned so much from you over the years. And your work, I think, deeply influences as well how we see all the AI conversation because in the same way that oil is the thing that kind of pumps up the GDP of all these countries, that's, we're now kind of switching from the oil-based economy to the intelligence-based economy.

[01:49:16] And that framing I have learned that I'm operating with every day that informs our work. I've learned from you. So thank you for all the work that you do. there's so many ways that we, you know, are influencing and informing each other, and, I'm grateful for it.

[01:49:29] **Nate Hagens:** See you soon, my friend.

[01:49:30] **Tristan Harris:** See you soon.

[01:49:32] **Nate Hagens:** If you'd like to learn more about this episode, please visit [The Great Simplification dot com](https://www.thegreatsimplification.com) for references and show notes. From there, you can also join our Hilo community and subscribe to our substack newsletter. This

# The Great Simplification

---

show is hosted by me, Nate Hagens, edited by no Troublemakers Media. And produced by Misty Stinnett and Lizzie Ani.

[01:49:54] Our production team also includes Leslie Balut, Brady Hyen, Julia Maxwell, Gabriela Slayman, and Grace Brumfield. Thank you for listening, and we'll see you on the next episode.

[01:50:06] I.