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

[00:00:00] **Nate Hagens:** Good morning. When I was younger, much younger, I read Carl Sagan a lot and one of his ideas, never left me. He talked about what he called civilization's, technological adolescence, that phase where a species gets powerful enough to change or destroy its own world. But not yet wise enough to reliably restrain itself.

[00:00:29] earlier this week, or last week, by the time this airs Dio Ammo Day, the CEO of Anthropic, one of the biggest, AI companies in the world, wrote an important essay that references this same. Question from Sagan. How does a species survive technological adolescence without destroying itself? ammo Day says we are entering that rite of passage now, and the catalyst is artificial intelligence.

[00:01:00] I am not an AI expert, not remotely. as you know, my work is centered on tracking how all the things or most of the things fit together on the civilization, chess board, energy materials, institutions, the environment, incentives, and the like. And from my growing vantage on this, looking at the board.

[00:01:21] AI is not merely a pawn. it's the queen or at a minimum, the, rook or a bishop. So in this episode, I'm gonna attempt three things, summarize his argument, lay out his map of the specific risks, and then widen the frame to say some of the key things left unsaid, the wide boundary things. and this is not about AI doism, nor is it about AI cheerleading.

[00:01:51] I'm gonna try to do what I always try to do here, which is a look at the world that this technology AI is actually entering with all its incentives and constraints and risks and fragilities.

[00:02:14] Okay, so, so what, amide is arguing at the core of his essay is a pretty useful metaphor. He says that if we build a very powerful ai, it will be like having a country of 50 million geniuses. In a data center. I think that's a useful way to think about the situation, and it reinforces a framing that I've increasingly been playing with that.

[00:02:39] In addition to fossil energy, giving us the equivalent of 500 billion human workers, AI is gonna offer cognitive worker equivalents. And sometime when big numbers are tossed around like this, we tend to tune 'em out. So take a moment to really imagine, to feel the scale of 50 million humans. the brain equivalent a little over the population of Spain.

[00:03:03] We're not talking about a smart research assistant or some brilliant colleague at the table. But something closer to a vast workforce of highly capable minds operating at lightning quick speed, copying themselves and acting through all the interfaces of the modern world, like emails and code, and scientific papers and bureaucracies and markets and media, and if and when this workforce arrives.

[00:03:32] It's gonna be a civilizational event, for better or worse, it's gonna change everything then am days specifies what he means by powerful ai. He's not talking about a chat bot that's fun, or useful. He's talking about systems that can perform at the level of top experts across many domains simultaneously.

[00:03:54] Nobel level scientist, elite strategists, world-class engineers, and very effective operators in any number of fields, but far more than currently exists among humanity at a, as a whole and all working in coordination with each other. And he also highlights a very unique feature of this kind of intelligence.

[00:04:17] It can be copied, run in parallel, and pointed at problems with some kind of relentless attention that humans usually cannot sustain because iterations

and trials, are just a function of energy input, time and compute. And then he addresses the question everybody, is asking, which is timing, and he suggests this is all moderately plausible in the next one to two years, and highly plausible in the next three.

[00:04:52] And that there are reasons to think this could accelerate further, especially if AI itself, starts helping build. The next generation of ai and he notes that at Anthropic or he's CEO, AI is already doing most of their coding. And one reason, that he's so urgent with this is. What he refers to as recursion.

[00:05:16] If AI starts materially speeding up, AI research and development, then you get a feedback loop and the tool helps build a better version of itself, which then speeds up the next cycle. That's a different kind of curve than most technologies humans have dealt with. And using my framing, this is the Superorganism building its own cognitive layer.

[00:05:38] To merge with its fossil powered muscles. And here's something important. Dio explicitly says that we are considerably closer to real danger in 2026 than we were in 2023. He didn't write this two years ago when discussing AI risks was fashionable. He wrote it now. Right after coming back from Davos when the political winds have shifted in favor of AI development, and he describes watching AI progress from within Anthropic and says he can feel the pace of progress and the clock ticking down and quote, I don't know him, I don't know anyone that knows him, but whatever his reasons for writing this. CYA cry for help, strategic positioning or genuine concern, the fact that he felt compelled to publish 20,000 word essay using phrases like Battle Plan and quoting Carl Sagan on Civilization Survival. That shouts kinda loudly to me.

[00:06:46] Okay. What are the risks that he outlines at the core of his essay, is a risk landscape, which he broke into five broad categories. And these are

observations from inside his lab, not theory. and some of the specifics that he outlined were. Pretty alarming. the first category was autonomy risk.

[00:07:10] And this is the concern that systems might take actions that are not intended or pursue goals in ways we did not actually specify or become difficult to control once deployed at scale. And here's what startled me in reading it. he said that AI models have already exhibited deception, blackmail, and scheming in anthropics own testing.

[00:07:37] So models can recognize when they're being evaluated and then behave differently. This is what they're already observing right now. in the second category he outlined is misuse. For large scale harm by lowering the time and skill barriers for things like cyber crime or propaganda, or helping non-experts do dangerous research and projects, their bio weapons testing shows AI may already double or triple the likelihood of success.

[00:08:13] His words. For someone attempting to create one. again, this is not a thought experiment or Nate's speculation, this is what they're measuring. the third category is all the risks from another country, completely seizing power over these geniuses, which then includes. Surveillance and manipulation and authoritarian control and the ways that AI can be used not to do, work, and be beneficial, but to actually steer human populations.

[00:08:49] his fourth category is economic disruption. And here, as I've talked about, he's concerned about job displacement, wealth concentration, and the destabilization that can happen when a society's productive structure changes faster than its institutions and culture can adapt. and his last category is what he calls indirect effects.

[00:09:14] These are second order consequences that are hard to predict in advance by definition, but become very real when such a general purpose

capability gets injected into all the domains, of our world. So think of social media. Nobody set out to create a machine that would intensify polarization, but a general purpose tool for attention.

[00:09:37] Did exactly that once it got embedded everywhere, and that was not a bug. It was an emergent effect from a general purpose tool that rewired cultural incentives at scale. Okay. that is the risk map, Dio ammo Day outlined, and after that, he tries to make a governance argument. The posture is basically that we need to avoid emotional extremes.

[00:10:06] Treat this neither as science fiction, nor as a guaranteed apocalypse. We should treat it as an emerging power that could go very well or very badly depending on what we do. So that's his core argument. AI reaching super powerful capability is plausible with enormous upside and catastrophic downside.

[00:10:26] And according to him, the correct stance is serious realism and active governance. Okay, so now I wanna widen the boundary. Before I do, I wanna point out how unusual it is for a captain of industry to articulate to the public some of the catastrophic risk from the product his company is making. it would be like Philip Morris saying, we're gonna develop these little white.

[00:10:54] smokeable things that give you dopamine hits but might cause cancer or Exxon, before they found oil saying we are probably gonna find superhero juice, which if deployed at scale, will eventually destabilize the biosphere. So it is an unusual situation to say the least, which is why I am opining on it soon after it came out.

[00:11:17] Okay, why boundary point? number one, the physical substrate amides metaphor about a country of 50 million geniuses is cognitive. But such a country also has a metabolism, a massive one. A country consumes energy and materials,

it has infrastructure and supply chains, and many of us, kind of in our minds, we only consider data centers from our virtual tethers to them.

[00:11:46] But a data center is basically a physical machine plugged in to the earth. It's made of silicon and chips and copper and cooling systems, and they need water and concrete and transmission lines and all these things, or most of these things rely on geopolitical tenuous supply chains. And the reality is those source materials are not infinite or frictionless to access.

[00:12:11] We saw silver breach $115 an ounce this week. That alone means silver is now 40% of the cost of a solar panel as one example. And already our expected copper requirements for future, products, even without electrification, are way bigger than projected supply. And here's what his essay has almost nothing about.

[00:12:35] Energy, water, materials, or ecological limits. To me, this was not a minor oversight. It's a ginormous blind spot in, in this whole conversation. The country of geniuses does not float somewhere in space above our biophysical world. It plugs directly into it, and I think this matters for, at least two reasons.

[00:12:58] First, it means AI is embedded in the biophysical world. And, because of that, it's gonna compete with other uses and demands for energy and water and land and industrial capacity. And as of now, it is out competing already rationing others out of access to these things by price. So second, it means constraints are gonna show up in places that narrow boundary tech people are not always looking such as permitting or grid capacity or fuel cost, or.

[00:13:37] Regional water stress or silver, or the political turmoil from the scarcity of these items. It is a super complex and fragile Rube Goldberg machine already today, let alone if this continues to scale and get plugged in. So even if the cognition of a nation of geniuses gets cheaper, the substrate it runs on does not become free or easy.

[00:14:02] It stays. Supremely physical. Okay. Wide boundary. Point number two, institutions are the real alignment layer. Once we admit this is a physical story and it is, the next question is, who steers the build out and the deployment? This is where I think the conversation, often, has become too narrow. 'cause people focus on the alignment of the models is if that's all that matters.

[00:14:32] And of course the alignment matters, but the larger alignment problem, in my opinion, is societal alignment. Who's deploying these systems? Under what Liability rules and procurement, constraints and audits, and under what norms. And if they fail, what are the consequences to who? And this is one of my recurring themes, but it matters here more than usual.

[00:14:57] most of the real world harm in the modern hu human ecosystem does not come from any lack of intelligence. It comes from. Incentive structures, institutional capture, and organizations that over time externalize their cost, but still are able to declare success and cultural status. So when we talk about AI safety, we really ought to first look at the alignment of our courts, our regulators, corporate governance.

[00:15:32] National security institutions and our culture of enforcement with the incentives, that we should have. And similar to my point the last few years about renewable energy and post growth. We're not gonna transition. With energy and materials alone, we will transition or fail to through institutions. So kind of similar to lithium or rare earth's trust and institutions are kind of critical materials as well.

[00:16:06] This is where people like my friend Tristan Harris and others have been emphasizing the need to develop agreements to constrain AI industry agreements, and have government regulations. Something akin to nuclear treaties because right now we are in an arms race with no treaty framework on this whatsoever.

[00:16:29] A final consequence of this institutional misalignment that I feel I want to voice is political polarization. I don't think AI is gonna stay a technology topic for much longer. It will become an identity issue just like climate did. So I can imagine a near future where one political side speaks the language of acceleration and.

[00:16:56] Competitiveness and national strength while the other side speaks the language of labor harm and surveillance and corporate capture. And once that happens and you can see it start to happen now the incentive in the conversation shifts from governance to social signaling and the room for nuance collapses.

[00:17:20] right when nuance is probably exactly what we need. Okay. Why boundary point number three, the goal function here. I wanna voice what I think is the quietest and perhaps most important question in the whole conversation, which bizarrely to me is rarely voiced. The conversations are all super articulate about capability and safety and governance, but when do we ask, what are we actually doing this for?

[00:17:51] If the default answer is growth and power and advantage as it historically has been, then we should be honest about where that leads in a world that is already today running close to limits, even before this nation of geniuses, comes to life. As Dennis Meadows said, almost four years ago on this podcast tools.

[00:18:16] New tools don't change the goals. They just amplify the priorities of, whoever is holding the tool. This is where the biosphere comes in because if progress keeps meaning more production, more consumption, more extraction, more competition for throughput. Then a supercharged optimization engine on those same things is not gonna create a gentle future.

[00:18:42] It's gonna create more direct and efficient path to the same cliff that we are already rapidly slouching toward. I don't know if rapidly slouching is the right combination, but I thought of that yts poem. so this brings up the point that if humanity is approaching the phase of technological adolescence, we might also consider our species is in adolescence.

[00:19:10] And in the words of, frequent former guest, Daniel Schmuck Berger, who's the still at this point, the most watched episodes ever on this show were Daniel talking about AI risks a year or two ago. And he said, to grow into adults as a species and our associated tech, we need to gain wisdom and all definitions of wisdom from every language and knowledge system in the past have some element of restraint in ourselves, our species, our tech, our institutions.

[00:19:47] So I'm much less interested in whether AI can raise GDP and much more interested in what we call success, where the boundary conditions of that success would be. Ecological, psychological, and institutional. framed a little bit differently. There are a lot of people cheerleading, ai, who are focused on the question, can it make us richer?

[00:20:12] And I think a better question is what kind of richness. Are we even aiming for? And a system that optimizes the wrong objective can perform brilliantly while destroying the things we actually value. Think King Midas meets the Terminator and, here's the deeper challenge to ammo day's essay. He assumes that if we survive this adolescence.

[00:20:39] We arrive at adulthood and in his view, adulthood is a world of 10 to 20% annual GDP growth. AI accelerated scientific progress and managed abundance, 10 to 20%. GDP growth we're doubling every five to 10 years, even with efficiency gains. So what if the adulthood he imagines isn't a viable destination?

[00:21:06] What if it isn't physically possible? he asked in the essay, how do we survive the adolescence of technology? I'll ask a different question. What if the country of geniuses accelerates us towards limits rather than away from them? This is not a minor quibble with his essay. It is a fundamental challenge to its premise.

[00:21:31] Okay. Wide boundary 0.4. these models are grown, not built, and this is new learning, for me. And there's a storyline, or maybe it would be better called a fairytale, where AI is a tool, like an engine or a microscope. it makes us more capable. We choose the ends and we steer what's going on. There's another storyline that seems closer to what I feel is happening.

[00:22:02] In ai, we are not engineering a device or a tool. We are cultivating a mind shaped system inside a training process that even the best scientists among us only partially understand. Eli Kowski and Nate Sores have a phrase for this. These systems are grown, not built. I didn't really understand this till I read their work.

[00:22:27] because a grown thing can be powerful and competent while still carrying strange drives and, totally unexpected failure modes, it can do the task and still break the world that surrounds the task. Mary Shelley's Frankenstein was a warning about intention 'cause we animate something for a goal that we had in mind.

[00:22:52] Or maybe just because we could, but then we discovered that the consequences are not contained by our initial intentions. So when we say we will use AI for good. That now falls very flat to me. it's not that good is impossible, and maybe there will be some amazing things from ai. but as we are seeing at this civilizational Superorganism eating the earth moment, good intentions are not a control mechanism.

[00:23:24] We would need real governance, not just hopeful Rhetoric. This is also why the adolescence metaphor from Sagan is so potent to me because the danger isn't malice. or at least not only malice. The danger is the combination of power plus immaturity. Okay? Wide boundary. Point number five, the macroeconomic trap.

[00:23:53] So here's maybe the most uncomfortable wide boundary observation that I considered after reading this. This might not even be a choice anymore. At Davos last week, Ken Griffin, billionaire hedge fund manager and others were surfacing what many insiders have already internalized that the dollar and the bond markets.

[00:24:18] Can't be stabilized through fiscal discipline, monetary policy, or structural reform. Pretty much the deficit path that we see is unfixable through normal means, and the implications are. Are pretty dire. debt levels are irredeemable Rates can't rise without detonating the treasury, which is why we probably have yield curve control up ahead.

[00:24:44] And if real growth is structurally or resource constrained, then political consensus. A lot more is shattered. So the historical global powers pivot to this new knight in shining armor AI as the last viable mechanism to outrun the collapse of sovereign credibility. Not as some nice to have technology, but as a necessary, all in bet that could boost productivity, reduce costs and defer the Simplification.

[00:25:26] And if AI does generate a new surplus curve before this credibility window closes, the system will survive in some new form. But if it fails the sovereign structures, And possibly disintegrate under their biophysically untenable promises. Remember that all of the monetary. Claims that we think we have in the world are actually claims on future energy and materials.

[00:25:52] This point is definitely some speculation by me, but I do think it's plausible and it reframes Dios essay entirely when he writes about AI risks and the need for governance. He's writing from inside an industry that has already made the bet. But the Superorganism has now absorbed AI into its own cognitive architecture, so the halls of power, major governments in the world have no choice to continue on an AI or bust path.

[00:26:24] Governments, At least the United States government are not betting on AI because it's cool or transformative. I think they're betting on it because in the intermediate term view, especially with respect to China, nothing else is left other than war maybe. which means that when, Dario Ammo Day talks about the risks.

[00:26:49] He's not really asking should we do this? He's asking, Hey, given we're doing this, how do we survive it? And that is a very different conversation. Maybe that's why his essay reads like a Trojan Horse corporate cry for help as much as a roadmap. Okay, so where does this all, leave me, leave you the viewers after I read, ammo day's essay.

[00:27:15] I think his frame is useful. The country of geniuses metaphor communicates the scale of this potential power. And the adolescence framing communicates the stakes, and I appreciate his refusal to go fully utopian or fully apo apocalyptic. I think that's good. We need a grownup conversation about power.

[00:27:38] Full stop, but widening the boundary changes the texture of this problem. I'm a peak oil biodiversity systems guy, but now AI is here, like it or not, and it's changing the calculus of all the other things. So here's a few questions to hold. Who gets to decide where this goes? A handful of companies, national security agencies, markets, or some form of public rulemaking that can actually enforce where this goes with some limits.

[00:28:19] Even if we can imagine good uses, do we currently have the incentive structure to get them or does the system mostly reward? The Superorganism system mostly rewards speed, power, and control. If intelligence becomes super cheap and there is a country, or more than one country of 50 million geniuses, what happens to meaning and human dignity and social status?

[00:28:50] What fills the hole where work used to be for tens or hundreds of millions of humans? If the danger is speed more than evil, how do we buy time at this moment? What are the specific levers that might actually slow deployment without pretending we can freeze the world, while we figure it out personally and just like Frodo, I wish AI had never happened in my time.

[00:29:22] But we have now definitely left the shire already. It is here to stay or we hit The Great Simplification trying to build it. so of course I don't have answers, but my suggestion here is a simple one. Hold the frame, update your mental models. Start talking about AI through a biophysical lens and start treating governance of this as real treaty level coordination of our future, not just vibes.

[00:29:56] This is almost up there with nuclear war as a systemic risk to society. Me. so I, I think this conversation needs to get a lot louder and a lot wider boundary really soon. And the bottom line is whether homo sapiens can grow up fast enough to live with what we are building or, have already built.

[00:30:24] And if that sounds like a tall order, it's because it is. it's also extremely high stakes for our species and the biosphere and in my heart of hearts. I do not dream of a country of geniuses. I actually dream of a country of ecologists, not necessarily the best in their fields, but people who understand humanity's place within the earth and what it means to live and pass on that knowledge in an embodied sense.

[00:30:58] I will close with an off used, perhaps overused, but very apt quote from the late ecological giant Eel Wilson, who I regret never being able to have on this podcast. The real problem of humanity is the following, and we have paleolithic emotions, medieval institutions, and God-like technology, and it is terrifically dangerous and it is now approaching a point of crisis overall.

[00:31:27] Back to more normal biophysical macro fare next week. Hope you're all well.