Nate Hagens (00:00:00):

Artificial intelligence is in the news. We hear about ChatGPT, making people more efficient, learning quicker. We hear about AI replacing artists and mid-level programmers. We see deep fakes and fake beautiful baby peacocks that are much cuter than real baby peacocks. And lots of people are debating about the benefits and risks of artificial intelligence. But today's guest is my colleague and friend, Daniel Schmachtenberger, who is back for a deep dive on how artificial intelligence accelerates the superorganism dynamic with respect to extraction, climate and many of the planetary boundary limits that we face.

(00:00:57):

I have not heard this angle on artificial intelligence before. I think it's really important to have this conversation and throughout this talk with Daniel, we talked about AI, but underpinning it all was what is intelligence and how has intelligence in groups in human history out-competed wisdom restraint of different cultures and different groups of humans? This is an intense, dense three and a half hour conversation and we weren't even done. We'll be back in the next month or so to have the follow on questions. It's probably one of the better conversations I've ever had on The Great Simplification, and I think it's really important to merge the environmental consequences of AI into our cultural discourse. Here's my friend Daniel Schmachtenberger.

(00:02:12):

Hello my friend.

Daniel Schmachtenberger (00:02:13):

Hey Nate. Good to be back with you.

Nate Hagens (00:02:15):

I prepare a lot for my podcasts. I read people's stuff, I prepare questions, I think about it, but with you, I'm like, I got a appointment with Daniel at 4:00 PM. I go for a bike ride, I go play with my chickens and I just show up and we have a conversation. So I'm hoping this'll work because this conversation actually is the culmination of how our relationship started a couple, three years ago. Remember we came to Washington, DC for a five or six day meeting where I wanted to discuss energy, money, technology and how this combined into a superorganism. And you were focused on existential risks and

particularly oncoming innovation in artificial intelligence and how that led to a lot of potential unknown destabilizing risks for society. And now we've educated each other after a couple years. And today rather than continue our bend versus break series, I thought we would merge these two lines of thought on artificial intelligence and the superorganism.

Daniel Schmachtenberger (00:03:37):

You and I have done five parts so far in this bend versus break series. Given all the things that are in the public attention on AI, we decided to do this one. I imagine some of the people will have heard that series and we can reference the concepts for anyone who hasn't. Do you want to give a quick recap on superorganism so we can relate it and maybe superorganism and metacrisis? And those are kind of frames we've established that we're going to be bringing into thinking about AI now?

Nate Hagens (00:04:10):

Sure. So humans are a social species and in the modern world, we self-organize as family units, as small businesses, as corporations, as nation states, as an entire global economic system around profits. Profits are our goal and profits lead to GDP or GWP globally. And what we need for that GDP is three things. We need energy, we need materials, and we need technology or in your terms, information. And we have outsourced the wisdom and the decision making of this entire system to the market. And the market is blind to the impacts of this growth. We represent this by money, and money is a claim on energy. And energy from fossil hydrocarbons is incredibly powerful, indistinguishable from magic effectively on human time scales. It's also not infinite. And as a society we are drawing down the bank account of fossil carbon and non-renewable inputs like cobalt and copper and neodymium and water aquifers and forests, millions of times faster than they were sequestered.

(00:05:54):

So there is a recognition that we're impacting the environment and all of the risk associated with this, you label the meta crisis or the poly crisis or the human predicament, but they're all tied together. The system fits together, human behavior, energy, materials, money, climate, the environment, governance, the economic system, et cetera. So right now, our entire economic imperative as nations and as a world is to grow the economy partially because that's what our institutions are set up to do,

partially because when we create money primarily from commercial banks, increasingly from central banks, when governments deficit spend, there is no biophysical tether and the interest is not created. So if the interest is not created, it creates a growth imperative for the whole system and we require growth. Now so far the market has dictated this growth, but suddenly there's a new kid on the block which is artificial intelligence created by prior intelligence, by humans. And that's what we're going to talk about today.

Daniel Schmachtenberger (00:07:18):

That's a good frame.

Nate Hagens (00:07:21):

I'm an educator. I've recently been a college professor. My whole role today is to inform and lightly inspire humans towards self-governance, better decisions, better pathways forward. And my trade deals with science and facts and systems ecology. And I'm afraid that AI will spell the end of what we know is true. And on both sides we won't know what's true and there will be things that people can grab on the internet that many of which are fake or influenced by artificial intelligence that destroys the social discourse. So that's one thing I'm worried about AI. The other is will AI accelerate climate change because it will make the superorganism, it would be like playing Super Mario or Donkey Kong or something like that and pressing the turbo button and it makes processes more efficient and it's just speeds things up, which means more carbon either directly or indirectly more efficiency. It feeds Jevons paradox.

(00:08:39):

Another one of my worries is a lot of jobs are going to disappear from AI and how does that factor into the superorganism? So it seems to me that AI both simultaneously makes the superorganism hungrier and more voracious, but also runs the risk of killing the host in several ways. So these are just some of my naive questions, but I think before we get into artificial intelligence, maybe we'll just start with intelligence and humans, I think previous conversation you and I had, that's what differentiates us from the rest of the biosphere is our ability to problem solve and use intelligence to grow the scale of our efforts that is coupled with energy and materials

always. So maybe you could just unpack how you see the historical role of intelligence before we get to artificial intelligence.

Daniel Schmachtenberger (00:09:54):

So I might start in a slightly different place, which is to actually start with a couple cases of AI that are obvious and then we'll go back to intelligence. The relationship between intelligence and the superorganism itself. Why human intelligence has made a superorganism that is different than animal and natural intelligence made in terms of the nature of ecosystems and then how artificial intelligence relates to those types of human intelligence, not just individually but collectively as you mentioned, mediated by markets or larger types of human collective intelligence systems. And then get to what has to guide direct bind intelligence that it is in service of something that is actually both sustainable and desirable. So let's talk about just artificial intelligence for a moment to give a couple examples because people have heard since, and the reason it's up so much since artificial intelligence was kind of innovated in the fifties and some could argue precursors before that.

(00:11:06):

The reason it is in the conversation so much currently is the deployment of large language models publicly. And we're starting with GPT-3 and the speed of the deployment of those relative to any other technologies. GPT-3 getting a hundred million users in, forget exact exactly what it was now six weeks or something, which was radically faster than TikTok's adoption curve, Facebook's, YouTube's, cell phones, anything which were already radically faster than the adoption curve of oil or the plow or anything else. So world changingly powerful technologies at a speed of deployment, which then led to other companies deploying similar things, which led to people building companies on top of them, which leads to irretractability.

(00:11:54):

And so the speed of what started to happen between the corporate races, the adoption curves and the dependencies is of course understandably changed the conversation and brought it into the center of mainstream conversation where it had been only in the domain of people paying attention to artificial intelligence or the risks or promises associated previously. So when people talk about AI risk or AI promise of which it has a lot of both, there's a few things about cognitive bias worth addressing here first, which is a topic you always address on why people come to

misunderstand the superorganism and get kind of choice making wrong, get sense making wrong.

Nate Hagens (00:12:42):

Thank you. That means you actually have watched-

Daniel Schmachtenberger (00:12:44):

Of course I have watched and read your things, this is why we're friends.

Nate Hagens (00:12:49):

You're so damn busy that I'm like, Hey Daniel, watch this. And you're like, oh, I will. But you and I have not talked about cognitive biases, but you're right, I do talk about them a lot. So carry on.

Daniel Schmachtenberger (00:13:01):

So let's take, there are clusters of cognitive biases that go together to define default worldviews. And they're not a single cognitive bias, they're a kind of bunch of them. And you don't even have to think of it as bias. It's just like, I mean it's a strong sounding word, though it's true. It's a default basis for the sense making and meaning making on new information people are likely to do first. And so one of them that I think is really worth addressing when it comes to AI is a general orientation to techno optimism or techno pessimism, which is a subset of a general orientation to the progress narrative. And I would argue, and we'll not spend too long on this, so it actually warrants a whole big discussion.

(00:13:53):

I would argue that there are naive versions of the progress narrative. Capitalism is making everything better and better. Democracy is, science is, technology is. Don't we all like the world much better now that there's novocaine in antibiotics and infant mortality's down and so many more total people are fed and we can go to the stars and blah, blah, blah. Obviously there are true parts in everything I just said, but there is a naive version of that that does not factor all the costs that were associated adequately. And there's a naive version of techno pessimism. So first on the naive version of techno optimism, when we look at the progress narrative, there's so much that has progressed that if you want to cherry pick those metrics, you can write lots

and lots of books about however everything's getting better and better and nobody would want to be alive at any other time in human history.

Nate Hagens (00:14:45):

There's two things that the naive progress is missing. One is the costs like climate change and the oceans and insects and the other is the one time subsidy of non-renewable energy and inputs and the source capacity of the earth, and those are not finite. So those are the two blind spots I think in that narrative.

Daniel Schmachtenberger (00:15:12):

So we could say the costs and the sustainability of the story. And so if you talk about the story of progress, particularly like the post modernity version of science technology and the associated social technologies, not just physical tech because capitalism and democracy and international relations are all kind of coordination systems that we can call a social technology, a way of applying intelligence to achieving goals and doing things of which you can consider language in early social technology, which it is. If you ask the many, many indigenous cultures who were genocided or extincted or who have just remnants of their culture left, or if you ask all of the extinct species or all of the endangered species or all of the highly oppressed people, their version of the progress narrative is different.

(00:16:17):

And just like the story of history written by winners or losers. But if you add all of those up, the totality of everything that was not the winner's story is a critique on the progress narrative. And so one way of thinking about it is that the progress narrative is there are some things that we make better. Maybe we make things better for an in-group relative to an out-group. Maybe we make things better for a class relative to another class for a race relative to another race for our species relative to the biosphere and the rest of species. Or for some metrics like whatever metric our organization is tasked with upregulating or GDP or something relative to lots of other metrics that we are not tasked with optimizing.

Nate Hagens (00:17:05):

Or for our generation versus future generations.

Daniel Schmachtenberger (00:17:09):

Exactly. Short term versus long term. And so the question is where it is not a synergistic satisfier where there are zero sum dynamics that are happening, the things that are progressing are at the cost of which other things. And we're not saying that nothing could progress in this inquiry we're saying are we calculating that well? And if we factor all of the stakeholders, meaning not just the ones in the in group but all of the people, and not just all the people but all the people into the future, and not just all the people but all the other life forms and all of the definitions of what is worthwhile and then what is a meaningful life, not just GDP, then are the things that are creating progress, actually creating progress across that whole scope.

Nate Hagens (00:17:55):

So I have two replies to that. My first is amen. And my second is you're advocating for a wide boundary definition of progress as opposed to a narrow boundary one.

Daniel Schmachtenberger (00:18:07):

And the definition between wide boundary and narrow boundaries. Very related to the topic of intelligence too. Are our goals narrow goals or are they very inclusive goals? If we have a goal to improve something, for whom? Is it for a small set of stakeholders? Is it for a set of stakeholders for a small period of time? Is it measured in a small set of metrics where in optimizing that, in being effective at goal achievement, we can actually externalize harm to a lot of other things that also matter. And whether we're talking about technology itself or nation state decision making or capitalism or whatever, we can talk about something where all of the problems in our world we could say have to do, the human induced problems have to do do with the capacity to innovate at goal achieving decoupled from picking long-term wide definition good goals.

(00:19:10):

And that doesn't mean that there is nothing good about the goal. It means that the goal achievement process is fragmented the world enough that, and sometimes it's not even perverse, right? I'm going to get ahead economically and I'm going to fuck the environment and the people doing slave labor in the mines and I know it and I'm just a sociopath. So I do it. Sometimes it's not that. Sometimes it's we are, the world is complex, nobody can focus on the whole thing. So we're going to make say a government that has different branches that focus on different things so they can

specialize in specialization and division of labor allow more total capacity. And so this group is focused on national security of this type or focused on whatever it is, let's say focused on if it's the UN, world hunger.

(00:19:58):

Now is it possible to have a solution to world hunger - that is where now my organization has specific metrics, how many people are fed, et cetera, and whether how much of the budget we get next year to be able to do stuff and whether we get appointed again or elected again, have to do with specific metrics where it is possible to damage the top soil. It's possible to use fertilizers and pesticides that will harm the environment and cause dead zones and oceans and destroy pollinators that advance harm metric. But if we don't, there is actually no way to continue within that power structure. This is an example where it's not even necessarily perverse in a knowing way, but the structure of it, the institutional choice making architecture is such that what is being measured for and optimized and tasked can't not prioritize some things over others. And that with increasing capacity to goal achieve what is externalized to the goal is increasingly problematic.

Nate Hagens (00:21:02):

So is the narrow boundary focus versus the wide boundary focus, could that itself be a basic fundamental difference between intelligence and wisdom? And then building on that, if an entity, a tribe, a nation, a culture focuses on the narrow boundary goals, won't that out-compete a nation tribe, a culture that focuses just on the wide boundary, broader multi-variable things, like fairness or environment or future generations.

Daniel Schmachtenberger (00:21:48):

So let's come back to that definition of wisdom and the relationship between wisdom and intelligence. But let's address that we were saying earlier, there's a naive version of the progress narrative or the kind of techno capital optimist narrative. There's also naive version of the techno pessimist narrative. The techno pessimist narrative over focuses on all of the costs and externalities who lost in that system and basically orients in a Luddite way and is like, no, fuck tech and new things. It was better before. There's various versions. One is there was more wisdom before. And this is a dissent from wisdom in terms of cleverness that is somewhere between less wise and evil.

(00:22:35):

The benefits that come from this will be more like hyper normal stimuli that actually cause more net harm that we're moving towards tipping points of catastrophic boundaries for the planet, et cetera. So let's just not tech and extreme versions of that look like the Amish. But Unabomber wrote a lot of things on this topic and they were not dumb things. He was doing a real critique of the advancement of technology and then being like, how do we not destroy everything if we keep it on this track? Now, and we'll also see that there have been indigenous perspectives that wanted to keep indigenous ways, that wanted to resist certain kinds of adoption of things that would, as far as technological implementation is considered a non-embrace of progress. Now, you were just mentioning if tech is associated with goal achieving and some goals have to do with how to upregulate the benefits of an in-group relative to an out-group, doesn't tech mean power?

(00:23:46):

Yes. Doesn't a group that rejects some of it mean less power in the short term? So where those competitive interests come, particularly if the other side both has tech and the mindset to use it, does that end up meaning that that doesn't forward? And we can see that when China went into Tibet and it was kind of the end of Tibetan culture as it had been, was that because Tibet was a less good culture, meaning provided less fulfilling life for all of its people than China and nature was selecting for the truly good thing for the people or the world or no, we can see that whether we're talking about Genghis Khan's intersection with all of the people he intersected with or Alexander the Greats or whatever that...

Nate Hagens (00:24:36):

But that was tech too.

Daniel Schmachtenberger (00:24:38):

Yes, yes. That those who innovated in the technology of warfare, the technology of extraction, the technology of surplus, the technology of growing populations, coordinating them and being able to use those growing coordinated populations to continue to advance that thing relative to others. There were cultures that might have lived in more population sustainability with their environment, maybe more long-term harmony maybe said, let's make all our decisions factoring seven generations ahead

and they were just going to lose in war every time. And so the naive techno negative direction just chooses to not actually influence the future. It's going to say, I'm going to choose something because it seems more intrinsically right, even if it guarantees, we actually have no capacity for enactment of that for the world. And that's why I'm calling it naive.

Nate Hagens (00:25:33):

I don't understand that.

Daniel Schmachtenberger (00:25:36):

If...

Nate Hagens (00:25:36):

That last thing. Could you give an example?

Daniel Schmachtenberger (00:25:38):

Yeah. If someone says, I think a particular advancement of, I think advancement of tech in general focuses on the upsides that are easy to measure. Cause we intended it for that purpose doesn't focus on all the long-term, second, third, fourth order downsides that are going to happen. I don't want to do that. We want to have a much slower process that pays attention to those downsides. Only incorporates the things with the right use and guidance and incentives. It will lose in a war, it will lose in an economic growth to the other cultures that do the other thing. If you want to take a classic example and go back to and in, it didn't happen exactly this way anywhere because it happened in such a different ways in the Fertile Crescent and in India and whatever. But as a kind of thought experiment illustration, the plow emerges animal husbandry for being able to use the plow.

(00:26:32):

Now we have to domesticate a bull, or a buffalo, turn it into an oxen. And that involves all the things it does. It involves castrating it involves having a whip and you stand behind it to get it to pull the plow for row cropping, whatever. So certain animistic cultures were like, I don't want to do this. We'll hunt a buffalo, but we also will protect the baby buffaloes. We'll make sure that our body goes into the ground to

become grass for the future ones. We're part of a circle of life. We believe in the spirit of the buffalo.

(00:27:03):

I can't believe in the spirit of the buffalo and beat one all day long and do things to it where I wouldn't want to trade places with it. But the culture that says now we're not going to do that thing is not going to get huge caloric surplus. It's not going to grow its population as much. It's not going to make it through the hard weather times as well. And so when the new technology emerges, those who use it, if the technology confers competitive advantage, it becomes obligate because whoever doesn't use it or use at least some comparable technologies loses when you get into rival risk interactions.

Nate Hagens (00:27:40):

Let me take a brief rabbit hole side step here. But while it's fresh in my mind, I think this dynamic that you're talking about now, and I know we're going to get to artificial intelligence, but in my public discussions, people are recognizing the validity of the systemic risk that I'm discussing and that we're headed for at least potentially a great simplification. Simplification is the down slope of a century plus of intensive complexification based on energy. But those communities, and you could talk about countries that simplify first because it's the right long-term thing to do. In the meantime, they're going to be out-competed by communities that don't because those communities will have more access to government stimulus and money and technology and other things. But it almost becomes a tortoise and the hare sort of story. Had a podcast a few weeks ago with Antonio Turiel from Spain and he said Europe is in much worse shape than the United States.

(00:28:59):

Cause the United States has 90% of its own energy. So Europe is going to face this simplification first in a worse way. So the United States has another decade. So you guys are off the hook. And I was thinking to myself really because yes, the United States is mostly energy independent, but Europe will be forced to make these changes first and maybe they will have some learnings and adaptations that will serve them in the longer run when we just ride high in the superorganism for a while longer. I mean that's really a complicated speculation, but what do you think about all that? Is that relevant?

Daniel Schmachtenberger (00:29:46):

So this is why we talk about the need to be able to make agreements that get out of the race to the bottom type dynamics, the multipolar trap, the social trap. Because of course if anybody starts to cost resources properly, price resources properly, meaning pay for what it would cost to produce that thing renewably via recycling and whatever it is, and not produce pollution in the environment through known existing technology, they would price themselves out of the market completely relative to anyone else not doing that. So either everybody has to or nobody can, right? And whether we're talking about pricing carbon or pricing copper or pricing, anything as you say, well we price things at the cost of extraction plus a tiny margin defined by competition. And that was not what it cost the earth to produce those things or the cost to the ecosystem and other people of doing it.

(00:30:43):

So the proper pricing at pricing is really very deep to the topic of perverse incentive. And yet if we talked about how do we ensure that in our, and this is core to the progress narrative, right? Because the thing that we're advancing that drives the revenue or the profit is the progress thing, the cost to the environment of that we're extracting something unrenewably that is going to cap out, that we're turning it into pollution and waste on the other side. And we're doing it for differential advantage of some people over other people and affecting other species in the process.

(00:31:17):

If you were to, the stakeholders that benefit, you get a progress narrative. The stakeholders that don't benefit you get a non progress narrative. But until industrial tech, like it's important to get that before industrial tech, we did extinct species, right? We over hunted species in an area and extincted them. We did cut down all of the trees and caused desertification that then changed local ecosystems led to flooding, ruined the top soil we did over farm areas. So environmental destruction causing the end of civilizations is a thousands of year old story, but it could only be…

Nate Hagens (00:32:00):

It's just now global.

Daniel Schmachtenberger (00:32:03):

So until we had industrial tech, we could not actually, we just weren't powerful enough to mess up the entire biosphere. So how powerful we are is proportional to our tech because we can see that a polar bear cannot mess up the entire biosphere no matter how powerful it is. Corporally, right? The thing that can mess up the entire biosphere is our massive supply chain, technologically mediated things starting with industrial tech. And so given that we are for the first time ever running up on the end of the planetary boundaries because we figured out how to extract stuff from the environment much faster than it could reproduce and turn it into pollution and waste much faster than it could be processed. And we're hitting planetary boundaries on both sides of that, on almost every type of thing you can think about in terms of biodiversity, in terms of trees, in terms of fish, in terms of pollinators, in terms of energy, in terms of physical materials, in terms of the chemical pollution, planetary boundary. So the things that are getting worse are getting very near tipping points that were never true before. Those tipping points will make it to where the things that are getting better, won't matter even for the stakeholders they're intended. And that's a key change to the story is, it can no longer be that the winners can win at the expense of everybody else. It is that we're actually winning at the expense of the life support systems of the planet writ large. And when that cascade starts, obviously you can't keep winning that way, which is optimized narrow goals at the expense of very wide values.

Nate Hagens (00:33:39):

You've described the naive progress optimist and the naive progress pessimist. Is there such a thing as a progress realist?

Daniel Schmachtenberger (00:33:52):

Yes. So I am a techno optimist, meaning there are things that I feel hopeful about that require figuring out new ways to do things. New technae both social tech and physical tech. But I'm cognizant that the market versions of that tech are usually not the best versions because of the incentive landscape. In the same way that if Facebook hadn't had an ad model, it would've been a totally different thing, right? If we're just talking about the technology of being able to do many to many communication, but you had something that was not a market force driving it, could you have had something that was much better that was not trying to turn people into a commodity for advertisers, which means behaviorally nudge them in ways that

manufacture demand and drive the emotions that manufacture demand, maximize engagement which causes the externality of every young person having body dysmorphia and ubiquitous loneliness and confusion about base reality and polarization.

(00:34:57):

Could we have done it where rather than drive engagement, the goal was to actually look at metrics of cognitive and psychological development and interconnectedness across ideological divides and do that thing?

Nate Hagens (00:35:10):

Yeah, of course.

Daniel Schmachtenberger (00:35:11):

So the same technology can be applied to wider goals rather than more narrow goals, and you get a very different thing. So the base technae, it's the technology and the motivational landscape that develops its application space we have to think about together. So there are ways that we can repurpose existing technologies and develop new ones, both social and physical technologies that can solve a lot of problems, but it does require us getting this narrow goal definition versus wide goal definition. And if intelligence guiding technology is as powerful as it is and actually exponentially powerful, and we're defining intelligence here as the ability to achieve goals, what is it that defines what good enough goals are that being able to optimize them exponentially is not destructive? That's how you would get a progress narrative that is post naive and post cynical.

Nate Hagens (00:36:14):

In contrast, I'm probably a techno pessimist or at least a mild one because I see how technology is acted as a vector for more energy and more climate, CO2 and degradation of nature at the same time, I think it's how we choose what technology we use. Like a golden retriever is probably the best technological invention ever of our species, even though it's really more of a co-evolution. But you know what I mean? It's something that we came about and sexually selected and for companionship and they give us the complete suite of evolutionary neurotransmitters for not a lot of resource input. And there's lots of other technologies that are appropriate that help us meet

basic needs and give us wellbeing that don't destroy the biosphere. But this gets back to, I don't think individuals chose...

Daniel Schmachtenberger (00:37:16):

Wait, this is important. The superorganism thesis that you put forward shows that the superorganism is oriented on a path that does kill its host and thus itself, right? It does destroy the space reality, the substrate that it depends upon. The meta crisis narrative that I put forward says a similar thing is why we did this whole five series to kind of show the relationships. And so I would say as long as the axioms of that thing are still in place, yes, I am a techno pessimist, meaning I think that the good things that come from the new tech don't outweigh the fact that the new tech is in general more often than not, accelerating movement towards catastrophic outcomes factoring the totality of its effects. But this is why I said there is a post naive and post cynical version in which I'm a techno optimist, but it requires not being on that trajectory anymore. It requires that the technology is not being built by the superorganism in service of itself, but is being built by something different in service of something different.

Nate Hagens (00:38:30):

Well, in that way I'm also a techno optimist because after growth ends and after super-organism is no longer in control, efficiency will no longer be a vector for Jevons paradox because then efficiency's going to save our vegetarian bacon. Because as the economy is shrinking, efficiency's going to be really important and innovation. Just right now, it's feeding more energy and stuff into the hungry maw.

Daniel Schmachtenberger (00:39:02):

For the people who haven't heard the previous stuff on Jevons paradox, will you do that briefly? Why efficiency? Because obviously AI can cause radical efficiencies which can help the environment. That's part of the story of why it's an environmental hope. So would you explain why as long as Jevons paradox is the case?

Nate Hagens (00:39:22):

Yeah, so humans get smarter on how we use energy around 1.1% a year. So we get more energy efficient every year because we're smarter. Coal plants use less coal to generate the same amount of electricity. We invent solar panels, our televisions are a

little bit more energy efficient and our laundry machines and one would think on the surface that that would allow us to use less energy. But what ends up happening is that money savings gets spent on other things that use energy and writ large new innovation ends up system-wide requiring a lot more energy.

(00:40:06):

Since 1990, we've had a 36% increase in energy efficiency at over the same time we have a 63% increase in energy use. So as long as growth is our goal and our cultural aspiration is profits in GDP, more energy efficiency will paradoxically unfortunately result in more energy and environmental damage. That's called Jevons paradox. It was based after a 19th century economist who observed this Walter Stanley Jevons who observed this in steam engines, that steam engines wouldn't reduce our energy use but they would scale because they helped everyone and were so useful.

Daniel Schmachtenberger (00:40:52):

So let's talk about first versus second, third, nth order effects here. Because Jevons paradox is, it's important to understand that.

Nate Hagens (00:41:00):

Daniel, what are the odds that we actually don't get to artificial intelligence on this conversation?

Daniel Schmachtenberger (00:41:08):

Low.

Nate Hagens (00:41:11):

Okay, keep going. First, second, third, order. Go for it.

Daniel Schmachtenberger (00:41:16):

So if we create a new technology that creates more energy efficiency on something, whether it's a steam engine or a more energy efficient energy generation or transportation or storage technology, the first order effect of it is we use less energy. The second order effect is now that it takes less, now that we have more available energy and energy costs less, there's a bunch of areas where there was not positive energy return, profit return that now become profitable. And so now we open up a

whole bunch of new industries and use more total energy, but it's a second order effect or even a third order effect because it makes some other technology possible that does that.

(00:41:59):

This is one of the asymmetries that we have to focus on in the progress narrative is the progress narrative is and technology in general, when we make a new technology, and by technology I mean a physical technology or even say a law or a business to achieve a goal, we're generally making something that is trying to have a first order effect on a narrow goal that is definable in a finite number of metrics for a small set of stakeholders. The stakeholders are called the total addressable market of that thing and very rarely is the total addressable market everything. And so we're making things, whatever it is. So I'm using technology in the broadest sense of human innovations towards goal achievement here. We're making technologies to achieve first order effects, meaning direct effects for a defined goal, for a defined population. Even if we're talking about a nonprofit trying to do something for coral, it's still focused on coral and not the Amazon and everything else. And so it can optimize that at the expense of something else through in terms of the second, third order effects of whatever putting that thing through does. And so we put out a communication to appeal to people to do a thing politically. Well, it appeals to some people, it really disappeals to other people. One of the second order effects is you just drove a counter response.

(00:43:33):

The counter response is people who think that the thing you're benefiting harms something they care about and now you just up-regulated that. Is that being factored? And so the progress narrative, the technology narrative and all the way down to the science narrative, and this is where we get to the human intelligence versus wisdom and then how this relates to artificial intelligence is it is easier to think about a problem this way. Here's a definable problem. It affects these people or these beings. It is definable in these metrics. We can measure the result of this and we can produce a direct effect to achieve it. We did, we got progress. Awesome. And the progress was more GDP. The progress was people who could communicate faster. The progress was less dead people in the ER, the progress was less starving people. The progress was whatever the thing was that we were focused on, even if it seems to be a virtuous goal.

(00:44:25):

But that same thing that you did maybe polarized some people who are now going to do other stuff. That is a second order and maybe third order effect, maybe it had an effect on supply chains. So the second, third, nth order effects on a very wide number of metrics that you don't even know what they are to measure on a very wide number of stakeholders that you don't even know how to factor is harder in kind to think about. So it is logically, cognitively easier as we talk about intelligence to figure out how to achieve a goal than it is to make sure that goal doesn't fuck other stuff up.

Nate Hagens (00:45:04):

So efficiency too has a narrow boundary and a wide boundary lens with which to be viewed.

Daniel Schmachtenberger (00:45:12):

Yes.

Nate Hagens (00:45:15):

But here's one of the challenges though. It's easy for a group of humans or a full culture to optimize one thing. It's very difficult to optimize multiple things at once. Multi-variable inputs and outputs are incredibly complex. So optimizing dollar profits tethered to energy, tethered to carbon, combining technology, materials and energy, that is a thing that was very easy akin to the maximum power principle. So what do you think about that?

Daniel Schmachtenberger (00:45:55):

I think that optimization is actually the wrong framework. When you think about everything that matters, you're not thinking about optimization anymore, you're thinking about a different thing. So optimization is... Now let's come back to what is distinct to human intelligence. Why did that cause a meta crisis or a superorganism? How does AI relate to that? And then what it would take other than intelligence is also relevant to ensure that the intelligence is in service to what it needs to be in service to. So we're not saying that humans are the only intelligent thing in nature. Obviously not. Nobody reasonably would say that, but there is something distinct about human intelligence. So how do we define intelligence? It's fascinating. Go look up on a bunch of encyclopedias and you'll see that there are a lot of different schools

of thought that define intelligence differently. Some do it in terms of formal logic and reason.

(00:46:56):

Some do it in terms of pragmatics, which is the ability to process information to achieve goals. Some do it just kind of from a information theoretic point of view of the ability to intake information, process it and make sense of it. And all of these are related. I'm not going to try to formalize it right now, but I'm going to focus on the applied side because it ends up being the thing that's selected for and it ends up being the thing that wins short term goals and that obviously we're building AI systems for. So there we can say intelligence is the ability to figure out how to achieve goals more effectively.

(00:47:41):

Or we can just say the ability to achieve goals. We can see that a slime mold has the ability to achieve goals and it will figure out and reconfigure itself. A termite colony figures out how to achieve goals and it reconfigures itself. There's some element of learning and when you watch a chimpanzee figuring out using this stick versus this stick to get larvae out of a thing, you can watch it innovating and learning how to achieve goals. So all of nature has intelligence. What is unique about human intelligence relative to other things? Relative to other, let's just say animals. We could talk about plants, funguses, all the kingdoms, but that gets harder. So let's just stick with other animals. Well first we realize that we can't talk about this properly because from an evolutionary perspective, there were things between the other animals that we look at now and humans meaning earlier hominids and so where do we call humans in that distinction?

(00:48:40):

So since they're not around, we can mostly talk about sapiens versus everything else on the planet that we're aware of at this point. But we can say that the thing we're calling human starts before homo sapien probably with homo habilis or Australopithecus or somewhere around there.

Nate Hagens (00:48:57):

We are the ninth homo. And perhaps the last, we don't know.

Daniel Schmachtenberger (00:49:04):

So people might question what kind of weird anthropocentrism is it that would have you say that you know that you have some kind of intelligence the whales don't have or the chimps don't have or whatever. And I think it's very fair to say what it is like to be a whale, we really don't know. And what about the experience of whaleness? The quality of the sentience of it might even be deeper than ours, might be more interesting in some ways. Totally. Right. That's a harder problem in kind to address. But we can in observing their behavior, say there are types of goal achieving that they clearly don't have the ability to do in a prima facie evidenced way that we obviously have the ability to do. They have not figured out how to innovate the tech that makes them work in all environments the way we have.

(00:49:59):

They have not even figured out how to stay away from boats that are wailing boats. And so from their most obvious evolutionary motive, how to figure out that thing is not a thing that they've really done. And so we can see in a prima facie sense that they're not innovating in technology and changing their environment, making the equivalent of an Anthropocene in a similar way, even when we see the way that beavers make beaver dams and changes their environment or ants do, they do it roughly the same way they did it 10,000 years ago. Humans don't do it roughly the way we did it 10,000 years ago. So we can see something unique about humans in our behavior related to the innovation in technology and environment modification.

Nate Hagens (00:50:44):

In whales' defense, they don't have opposable thumbs and they're underwater. But I'm with you. Keep going.

Daniel Schmachtenberger (00:50:51):

Well this is not putting whales down. I think opposable thumbs are pretty significant to the story. I think there are things about the evolution of homo sapiens that probably have to do with the combination of narrower hips from uprightness that allowed us to de-weight the hands that allowed them to be more nimble and opposable with the larger heads that involved neotenous birth. And there's this whole complex of things. And so in no way does saying that humans have more of this particular kind of innovative intelligence mean have a more meaningful right to exist. Those are totally

separate things, right? Doesn't mean have a deeper experience of the world. Those are different things.

Nate Hagens (00:51:34):

So let me get back to something you just said a minute ago, unless this is where you were heading. But you said intelligence is problem solving on route to a goal. And most animals in nature, their goal is, well it's security and mating and reproduction, but energy return is a primary goal in nature to invest some energy and get a higher amount back because that enables all sorts of other optionality. Energy, calories in nature are optionality for organisms. So the problem with humans isn't the intelligence per se, it's the goal?

Daniel Schmachtenberger (00:52:16):

I don't even want to call it a problem yet. I want to call it a difference. We'll get to the problem in a minute. So I want to say something, because this is related here, about modeling because human intelligence, all forms of intelligence have something to do with modeling. They can take in information from the environment and be able to forecast what happens if they do something enough to be able to inform their next choice. Which choice is more likely to achieve some future goal, even if the future goal's a second, right?

Nate Hagens (00:52:51):

So let me interrupt there. Does that differentiate humans that if we model something that we have the perception and ability to consider time and how does time factor into intelligence?

Daniel Schmachtenberger (00:53:06):

We're not the only animal that has a relationship to time, but we definitely have the ability to have abstractions on time that seem to be unique from what we can tell. And we also have the ability to have abstractions on space and abstractions on other agents. And there's something about the nature of abstraction itself that is related to the what is novel in human intelligence, the type of recursive abstraction. But there's reason, to talk about modeling. A model of reality is taking a limited amount of information from reality and trying to put together a proxy of that that will inform us for the purpose of forecasting and ultimately choice making, ultimately goal achieving

insofar as the model gives us accurate enough forecasts that it informs actions that achieve our goal. We consider it useful. That doesn't mean that it is actually comprehensively, right. The models end up being that they're optimizing for a narrow set of sense making just like what we were talking about before, that we optimize for a narrow set of goals.

(00:54:23):

And the reason I bring this up is because all of our models that can be useful, even in trying to understand the meta crisis and whatever themselves can also end up blinding us to being able to perceive outside of those models. So when Laozi started the Tao Te Ching with the Tao that is speakable in words or understandable conceptually is not the eternal Tao. It was saying keep your sensing of base reality open and not mediated by the model you have of reality. Otherwise your sensing will be limited to your previous understanding and your previous understanding is always smaller than the totality of what is. I would even argue that thou shall have no false idols, a model of reality that says here's how reality works, is the false idol which messes up our ability to be in direct perception of new things where our previous model was inadequate.

(00:55:14):

So I say this because there are places where a particular thing we're going to say is useful, but it is not the whole story and it's important to get where it's not the whole story. So for instance, if we talk about energy that doesn't include the parts about materiality and the parts about intelligence. And even if we talk about those three, that's actually not the whole story. So it's useful, but I'm wanting to call this out. If an animal is eating, are they eating only for energy? No, they're also eating for minerals and for enzymes and for vitamins and for proteins and for fats, and not just fats that will get consumed as energy, but that will become part of the phospholipid membrane of cells and they're eating for materiality as well, right? And so it's not true that all energy is fungible. I can't really feed an elephant meat product well, even though there's plenty of energy in it.

Nate Hagens (00:56:08):

So hold on a second. So elephants also have what I refer to as the trinity, which is energy, materials and technology. They try to get acacia trees, they use their trunk or some other tool. And in the acacia leaves are energy, the photosynthesis from the sun,

but also as you said, atoms, minerals, materials. So it's the same for animals as humans?

Daniel Schmachtenberger (00:56:38):
And they're non-fungible. There's a reason why-

Nate Hagens (00:56:39):
How so?

Daniel Schmachtenberger (00:56:42):
I can't make certain amino acids from other amino acids. I can't make some minerals from other minerals no matter how much calcium I get, I get no magnesium from them and I need a certain amount of magnesium. So of course if I don't get enough dietary input of vitamin D, we get rickets and die even if we get plenty of B vitamins and vitamin C and other things. And so those nutrients are non-fungible to each other and you need all of them. You need the whole suite that a thing needs, which is why there's very interesting health studies that show people who are dying of obesity are actually dying of diseases of malnutrition.

(00:57:19):
Because we have a diet that has basically optimized for calories while stripping all the micronutrients out of them. So you can be eating tens of thousands of calories a day and actually becoming profoundly deficient in minerals and phytochemicals and other things like that where then your body is wanting to keep eating because it's actually starving and you continue to give it something that creates a neurochemical stimulus that says you ate and satiates the hunger thing for a moment, but what you're actually starving for is not in that food.

(00:57:53):
And so I don't want to oversimplify that energy is a part of the story. Everything we say is a part of the story, but the totality of the story is more complex than however we talk about it, right? There's something so important and sacred about that because what is wrong about our narrow goal achieving is what's upstream from that is our narrow modeling of reality and what even equals progress. Who is worth paying attention to? How is it all connected? When I make a model that separates it all, then I can up-regulate this, harm something else, but I don't even realize I'm harming

something else because that's not in my model. I don't even realize that that thing that I'm optimizing for isn't the actual thing or is only a part of the whole thing.

Nate Hagens (00:58:45):

And then probably by definition we choose models or inputs into the models that kind of confirm our own built identity up until that moment.

Daniel Schmachtenberger (00:58:55):

And, or, win in some game theoretic way. So if it achieves, if the model damages lots of things, but makes me win the war, that model will probably win. And you notice how it's like, okay, so we get the inquisition, we get the crusades, we get some fucking gnarly, violent, cruel like figuring out how to optimize torture stuff in the name of the guy who said, let he who has no sins among you cast the first stone and you're like, how the fuck did we go from principles of forgiveness and let he who has no sins cast the first stone into this version that says the inquisition is the right way to do that. What you see is the interpretations - there's lots of interpretations - the interpretations that lead to kill all your enemies and proselytize win, end up winning in short term warfare. Not because they're more true or more good, but because they orient themselves to get rid of all of the enemies and have more people come into them and have nobody ever leave the religion because they're afraid of hell and whatever. And this is an example of there are models that win in the short term but that actually move towards comprehensively worse realities and, or even self extinction. Evolutionary cul-de-sacs. And I would argue that humanity is in the process of pursuing evolutionary cul-de-sacs where the things that look like they are forward are forward in a way that does not get to keep forwarding. And at the heart of that is optimizing for narrow goals. And at the heart of that is perceiving reality in a fragmented way and then getting attached to subsets of the metrics that matter. Models. Which leads to us wanting to optimize those models and those metrics. And now I would start to define the distinction between intelligence and wisdom here. And that wisdom is related to holes and wholeness. Intelligence is related to the relevance realization of how do I achieve a goal?

(01:00:59):

A goal will be a narrow thing for a narrow set of agents bound in time, modifying a fixed number of parameters. And so then I would say if human intelligence distinct

from other types of animal intelligence where we don't just have the ability to work within a range of behaviors that are in capacities that are mostly built into where the primary physical technology is our bodies, the animals evolved to have claws, to have blubber for the cold, to have whatever it was that was a technological innovation to be effective within its environment. And it can't become radically more of that thing by choice, by its own understanding. It becomes more of that thing through genetic selection, which is super slow and it doesn't control. And the mutation that makes the giraffe have the slightly longer neck or the cheetah a little faster or whatever it is, is happening as the rest of the environment is going through similar mutations.

(01:02:05):

So the cheetah's getting a little faster, but so are the gazelles and there's co-selective pressure so that if the cheetah gets a little faster first and eats the slower gazelle's, then what's left is the faster gazelle's whose genes inbreed. So you have tiny changes happening across the whole system and co-op regulating each other. So there're cemeteries in the rivalry that lead to the entire system still maintaining its meta stability, not stability, meaning a fixed equilibrium, a homeo-dynamic, not a homeostasis, that continues to increase in complexity over time, but that meta stability is the result of that type of corporeal evolution. But then humans' adaptive capacity is not mostly corporeal, it's mostly extra corporeal. You call it extra somatic, meaning outside of just our body and it's both, we can use a lot of calories outside of our body, which started with fire.

(01:03:03):

Fire was the beginning of us being able to warm ourselves and all of a sudden make new environments possible and make foods edible that weren't edible before. But that's calories, right? That's extra somatic calories. Then our ability to get more calories from the environment, meaning gather more stuff, kill more things, involve the innovation of tools, right? Those spears, those stone tools allowed a little group of primates to take down a mastodon. The combination of their coordination technologies with each other because a single person couldn't do it and their stone tools together was able to do that. You get more caloric surplus. The agricultural revolution really advanced that, the oil revolution really advanced that. But at the heart of how did we figure out how to get oil and how to use it was intelligence. It was this kind of recursive intelligence that figures out I can use this in service of my goals.

Nate Hagens (01:04:02):

I wonder if 30,000 years ago some Neanderthal interbreeding with a human could ever imagine that 30,000 years later there would be a brain evolved like Daniel Schmachtenberger's.

Daniel Schmachtenberger (01:04:17):

Well check this out. Tyson Yunkaporta. I don't know if he's been on your show yet or not.

Nate Hagens (01:04:25):

Two weeks from now.

Daniel Schmachtenberger (01:04:27):

Okay, so great. Then you can follow this conversation up with some things I learned from Tyson and other indigenous people who hold some indigenous wisdom, knowledge have told me similar things that one, date back more than the standard current archeological narrative of when humans knew certain shit but also with kinds of wisdom that have definitely been lost in the progress narrative. And one of the things that Samantha Sweetwater told me originally, and then Tyson said something similar was that many of the indigenous cultures had a story that when humans developed the first stone tools, the apex predator of the environment, Samantha's version was the sabertooth, came to the human.

(01:05:16):

Obviously this is a story, right? But you get what it would mean that the early people had made the story and they said we were the ones who were taking care of and maintaining the complex diversity of the whole system in this kind of apex predator role you now clearly are and we're turning over the mantle of stewardship of the whole to you, your job. Because now you have the ability to destroy the whole ecosystem. You must be the steward of the whole thing.

(01:05:41):

And imagine that even recognizing because maybe those stone tools were 2 million years ago and maybe we already had killed a bunch of megafauna, extincted them, extincted some of our other hominoid cousins through gruesome, kind of inter-species

genocidal tribal warfare, destroyed some environments and already had time to learn those mythos and be like, no, no, no.

Daniel Schmachtenberger (01:06:03):

Destroyed some environments and already had time to learn those myths and be like, no, no, no, we're not going to maximum power principle kale, take everything. We're going to live in sustainability. Think seven generations ahead. And there was wisdom about appropriate use of technology and restraint 40,000 years ago.

Nate Hagens (01:06:18):

That was my question is, we are homo sapiens, our appellation, which is wise man, but any small percentage of tribes or individuals or nation states or warring clans that pursued a narrow boundary goal would have out-competed those tribes with wisdom.

Daniel Schmachtenberger (01:06:42):

Not any of them...

Nate Hagens (01:06:44):

And here we are with the superorganism.

Daniel Schmachtenberger (01:06:45):

Any of them above a certain threshold, right?

Nate Hagens (01:06:51):

What do you mean?

Daniel Schmachtenberger (01:06:53):

Let's say that we had a number of tribes in an area that had all developed some kind of wisdom by which they bound intelligence. And I'm not saying we don't do that today. It's called law, right?

Nate Hagens (01:07:04):

Right.

Daniel Schmachtenberger (01:07:07):

And it's supposedly also what religion is about, the development of wisdom of what is the good life, what is worth pursuing and not pursuing in which you get things like religious law. You're not going to work on the Sabbath. You're going to take that day to do different things. For instance, if you want to think about that as a example, you could think about the Sabbath as an example of law binding a multipolar trap, associated with a naive version of progress. If you don't have a Sabbath, some people will work seven days a week in the short term before they burn out. They'll get ahead. They will get so much ahead because they'll be able to keep investing that differential advantage in rent-seeking behavior that anyone who doesn't will have no relevance to be able to guide their own lives.

(01:07:59):

And now you have a world where no one spends any time with their kids. No one reflects on the religion, nobody enjoys their life. Everything sucks for everyone because somebody did that thing and the name of progress because they moved ahead faster. We say, "No, no, no, you're going to have a day where you don't progress stuff." That's actually the gist. You're not going to be focused on external progress in the world. And so there's 27 or 29 ways in Leviticus that you can violate the Sabbath and you'll be killed if you violate it. Which seems like just whackadoodle religious nonsense. But if you're like, wait, no, you never have to actually do that. If you hold that law that extremely, and it's everyone's like, "All right, we're not going to with the Sabbath. Now what do I get to do that day?"

(01:08:39):

I get to reflect rather than achieve goals, I reflect on what are good goals. So I get to spend time with my family, I get to spend time with nature, I get to read the scripture, I get to meditate and I don't get to achieve goals. I get to experience the fullness of life outside of goal achieving. And I get to reflect on what goals are truly worthwhile and in doing so, binding the multipolar trap that I don't have to because everybody else is rushing ahead.

(01:09:04):

That would be an example of the way religions were supposed to have something like wisdom that created something like law and restraint to bind naive versions of progress in a way that was actually better for the whole long term.

Nate Hagens (01:09:19):

Two comments there. One, when I was much younger, I had some Jewish friends and I didn't ridicule them, but I was kind of, "Ha ha, you guys have Sabbath today, I'm going to go to the arcade or go on a boat ride or go fishing or whatever." But now as I'm older, everything you just said about the good life and spending time with family and reading and spending time in nature and not using the internet on a Saturday or whatever sounds freaking wise and makes sense and appealing to me. Maybe with age and maturity, I'm flipping from intelligence to wisdom.

(01:09:56):

And then the second thing, the implication, and maybe this is where you're heading, is to muzzle or forestall the risk singularity that is coming from the superorganism. We have to have some Sabbath equivalent applied to AI.

Daniel Schmachtenberger (01:10:17):

It's not just Sabbath equivalent, but almost all of law is about restraint, right? Things that you don't do in the presence of having incentive to achieve narrow goals. What are the things that for the collective wellbeing, which also means the capacity for your own individual wellbeing for all individuals into the future, what are the things that we say we don't do? If you have Samantha on this is a topic she'll talk about cares a lot about, which is there's no definition of wisdom worth anything that is not bound to the concept of restraint.

Nate Hagens (01:10:53):

Yeah, I don't know how our culture approaching a biophysical Wiley coyote moment 20 or 30 or 40 years ago before all this over leverage and different systemic risk, we could have added restraints, but now our restraint would almost default create this rubber band snapback in the economic systems. But we can talk about that another time.

Daniel Schmachtenberger (01:11:19):

Well, so this is where you end up having the embedded growth obligation of a system, the embedded continuity of a system, the kind of institutional momentum that, a partner Zach was writing something recently, a couple of people on the team were contributing, and in the beginning it was talking about where do we find ourselves

now? And it says we find ourselves in the relationship between the life giving nature of the biosphere and the life giving nature of the civilizational system.

Nate Hagens (01:11:54):

Yeah.

Daniel Schmachtenberger (01:11:55):

And the unique point in time at which the latter is threatening the ongoing continuity of the former upon which it depends. And that what it takes to maintain that civilizational system will destroy the biosphere the civilizational system depends on, so we must remake the civilizational system fundamentally.

(01:12:15):

We do need civilizational systems. We do need technological systems, but we need ones that don't have embedded exponential growth obligations. We do need ones that have restraint. We do need ones that don't optimize narrow interests at the expense of driving arms, races and externalities. We do need ones where the intelligence in the system is bound by and directed by wisdom.

Nate Hagens (01:12:36):

Right. Which is the equivalent of sabbath plus law plus emergence.

Daniel Schmachtenberger (01:12:41):

Now come back to for a moment to the "ha ha, what idiot's" reaction you had when you were young. I had a similar one and lots of young people, probably even many people who were raised Jewish have a similar one before. They understand the full depth of it. So let's talk about Chesterton's fences for a moment.

Nate Hagens (01:13:01):

Never heard of that.

Daniel Schmachtenberger (01:13:03):

I don't know actually the history of why I've got that name, but there's a thought experiment in philosophy that says called Chesterton's fence, which is there's a fence up and you think, the purpose of that fence was X. That's no longer here. The fence is

ugly and in the way, let's cut, let's take the fence down. Is there a chance that the purpose for the fence included several other things that you don't know and you don't know that you don't know it? And before you take the fucking fence down, you better make sure you actually understand why it was put up. And now this comes to a very deep intuition. We were talking about biases earlier in the progress narrative. Progressives, it's funny how right now that is somehow associated with left in some weird way. But progressive and traditional is a deep dialectic and neither one are supposed to be the one you choose.

(01:13:52):

It's a dialectic. You're supposed to hold them in balance, right? And very much in the same way. This is an important point and it relates to wholes and wisdom versus narrow goals. Narrow value sets are as bad as narrow goals. They're part of it. Any value that is a real value exists in a dynamic tension with other values.

(01:14:17):

A dialectical one oftentimes, but other values. Where if you optimize the one at the expense of everything else, you optimize it, you get these reductoabsurdum, right? Meaning the optimization of any value by itself can end up looking like evil. So if I want to optimize truthfulness and all I'm going to do is speak the truth all the time, then when the Nazis come and ask me are there any Jews inside? I say, "Yes,"… No! I don't - truthfulness is not the only value at that point to the preservation of life and kindness and other things are." We can even see an example where it's like if someone in a naive sense says, my value is honesty, there's a bunch of places where you can see a person who in the name of honesty is just an asshole. And they just say kind of mean things and say it's in the name of honesty.

(01:15:09):

We can see that in the name of say kindness. People will lie to say flattering things and avoid painful things. We can see that if you hold them in dialectical tension, you actually get a truth that is more truthful and you get kindness that is more effective because the kindness that doesn't want to tell the person they're an addict when they are and nobody does or tell the emperor they have no clothes or say anything painful that is necessary feedback isn't even kind. Sometimes for that value to even understand itself fully, there's this kind of dialectical relationship that helps that come about.

(01:15:46):

Now here's where I'm coming to Chesterton's fence and then I want to hear.

Nate Hagens (01:15:50):

Okay.

Daniel Schmachtenberger (01:15:51):

There is a dialectic between a traditional impulse and a progressive impulse. The traditional impulse basically says, I think there were a lot of wise people for a long time wise and smart people who thought about some of these things more deeply than I have who fought and argued.

(01:16:08):

And that the systems that made it through evolution that made it through, made it through for some reasons that have some embedded wisdom in it that I might not understand fully and it makes sense for me to have as my null hypothesis, my default trusting those systems, they wouldn't have made it through if they weren't successful, didn't work. And likely the total amount of embedded intelligence in them is more than I've thought about this thing. Without knowing it, that's the traditional intuition. The progress intuition is, collective intelligence is advancing, built on all that we have known. We're discovering new things and we're we're moving into new problem sets where the previous solutions could not possibly be the right solutions because we have new problems. We need to have fundamentally new thinking. Obviously these are both true. Now on the traditional side, the Chesterton's fence thing is I might have as a kid, or you might have as a kid thrown out the Sabbath and said, that's dumb.

(01:17:08):

Before we actually understood it because we understood a straw manned version of it, said it was stupid and threw it out. And so when we're talking about wisdom and restraint and all like that, there is something around are we seeking to, because in the name of progress there will always be something that is focused on restraint that seems like it's up that progress I could get. And if I don't understand all the reasons for the restraint that factor second, third, fourth, nth order effects long into the future in the short term, I should do the thing. In the short term, no, of course I should advance the AI applied to genomics to solve cancer without thinking through the fact that the fourth order effects might involve increasing bioweapons capability for

everyone in destruction of the world, so even the cancer solutions don't matter in the course of those people's lives.

(01:17:57):

And so this is there a whole enough perspective to be able to see how the things that are actually wise from a narrow perspective look stupid?

Nate Hagens (01:18:07):

Two questions, one, is the metaphor the same as when I was younger, I thought those things were stupid and now I recognize the validity of them. That's where we are as a culture? We are the younger version of Nate in the intelligence versus wisdom dynamic? I'm just speculating. I think that's probably the case. And then two, you and all the people that I know, I know a lot of smart people, you're certainly up there, but you also have wisdom and I don't know as many people that have both intelligence and wisdom and you in my sphere rank near the top. But is it in our genome? Is it in the human behavioral repertoire to hold more than those single values to hold multiple values and wide boundary views of the world? What do you think about that?

Daniel Schmachtenberger (01:19:14):

We said that there is something unique about the types of recursive intelligence that lead to technology innovation and the anthrop scene, the superorganism et cetera in human intelligence relative to other species. Let's talk about the genetic predisposition and what the predisposition is actually for and the nature nurture thing a little bit I would say. And again, everything I'm going to say here will be at a high level that is hopefully pointing in the right direction, but totally inadequate to a deeper analysis of all the topics. Our nature in terms of the genetic fitness of humans, homosapiens, it would be fair to say that our nature selected for being more quickly and recursively changeable by nurture than anything else.

Nate Hagens (01:20:24):

As individual humans?

Daniel Schmachtenberger (01:20:27):

In the individual human is not the unit of selection and evolution, the tribe is.

Nate Hagens (01:20:33):

Right? Well, both.

Daniel Schmachtenberger (01:20:35):

The tribe or the band, the group of humans.

Nate Hagens (01:20:39):

Both, sometimes individuals, sometimes tribes.

Daniel Schmachtenberger (01:20:43):

I don't think there is much of a case for individual humans surviving in the early evolutionary environment by themselves very well. And the behavior of them as individuals separate from social behavior leading, there are certainly some animals that are largely solitary and they have a different set of selection criteria than primarily social animals. Humans are a primarily social-

Nate Hagens (01:21:04):

But I mean this was EO Wilson, David Sloan Wilson's paper that selfish individuals out compete within groups and cooperative groups, out compete selfish groups. I think both are hardwired in us, but let's not get detracted by that.

Daniel Schmachtenberger (01:21:26):

Actually what I'm saying holds with this, the individual, there are some selection of an individual within a social environment, but there's no selection of an individual outside of other sapiens, right?

Nate Hagens (01:21:40):

Right.

Daniel Schmachtenberger (01:21:40):

And so the unit of selection that is driving the dominant feature for sapiens, the unit of selection is a group. That's actually a really important thing to think about. As opposed to that the unit of selection's an individual because we have such an individualistically focused culture today, and we think in terms of individual focus way excessively to the actual evolutionary fitness of an individual outside of a tribe was dead in almost every environment for most of history. A set of behaviors that made

you alienate the tribe was not an evolutionary strategy for most of the evolutionary basis of humans.

Nate Hagens (01:22:20):

And the problem now is our tribe is 8 billion strong pursuing profits tethered to carbon-

Daniel Schmachtenberger (01:22:25):

Which is no longer a tribe. The tribe was capped out at the scale at which there were certain types of communication across that group that allowed it to be the unit of selection. And so everyone knowing everybody. Everyone being able to communicate with everyone, everyone being able to participate in some choices that then everyone would be bound by, so they stayed in it rather than defect against it. Which is why you got kind of Dunbar limitations. And then there's a series of things where you went from say couple hundred thousand years of Dunbar to huge cities in a relatively short period of time, which is the beginning of the thing we call recorded civilization.

(01:23:07):

And we would say that most of the superorganism properties we talk about now were at that junction started. Because in the smaller size, lots of lying and sociopathy and whatever really don't pay because people are going to know you're lying and enough people can beat you up if you are very problematic.

Nate Hagens (01:23:31):

We as individual humans, because evolutionary selection acted at the tribal unit, we have the capacity for wisdom, but once the number of people and the self-organization went to the city state, nation state size, there became downward causation of the emergent phenomenon at the aggregate level. That started to focus on intelligence and out-competed the wisdom of individuals in smaller units.

Daniel Schmachtenberger (01:24:13):

The multipolar trap, the Moloch behaviors really took off there because you can think of homo sapiens with tools, they were already different than the rest of nature because they had stone tools and fire and language. They were already different, which is why they had already extinced other species and moved and become the

apex predator everywhere. The beginning of the story is the beginning of the type of intelligence that leads to recursive innovation, recursive abstraction leading to recursive innovation. When you look at what various people call the defining characteristic of humans, or the earliest technae that made humans really distinct stone tools, fire and language are three very common one's brought up. Stone tools and innovation in the domain of matter. Fire is an innovation in the domain of energy. Language is an innovation in the domain of intelligence or information. And they were all intelligence applied to those innovations. Recursive intelligence of this kind, right?

(01:25:14):

A group of humans with no stone tools can't do anything to a mammoth just with their fangs and claws. They're not going to hunt to mammoth, right? Eight dudes aren't going to do that. And one dude with a spear isn't going to hunt to mammoth. It is the coordination protocols and the physical tech together that led to those capacities, so both the social tech and the physical tech, and so I'm meaning tech here as intelligence applied to goal achieving the innovation of new fundamental capacities. That most broad definition of tech. And so for a couple of hundred thousand years of sapien, you have this very small size, a couple million years of homage, at-large. And then we can argue why exactly we started to get way on beyond the Dunbar number and where we went and beyond the couple Dunbar numbers into Sumaria, Egypt, Goblectepi, like the first large ones.

(01:26:23):

But one function that certainly comes up in the analysis is having already capped out migration as a strategy and tribal warfare, you start getting into resource limits with another tribe. You're sharing the same environment, competing for the same stuff you just move. When you've moved to all the places, there's nowhere else to move. There's resource issues and one tribe is willing to be warring. Now the other tribes have to unify together to survive. They're willing to sacrifice some of the collective intelligence of their smaller scale and their intimacy with each other in whatever for survival.

Nate Hagens (01:27:01):

Collective wisdom of their smaller scale?

Daniel Schmachtenberger (01:27:03):

Yes. And now there is collective intimacy actually, and collective wisdom are very deeply coupled. Because if everybody can know everyone, you can have intuitive mappings to everyone. You can care about everyone. You can have shared pathos and you don't want to hurt someone where you hurt if they hurt, right? As soon as we get too much larger scale where I don't know everyone, I don't trust everyone, I don't have intimacy, now I have to have some way of knowing we're in group. We are all under the same flag, we all have the same indicator of some kind. And now I don't get to act in a unique way to each person. I have to have rule sets that mediate it. We start to get this kind of intelligence rule-based law that is different than wisdom. Now one could say that all of the wisdom traditions were trying to... There's different ways of saying this, having built an environment that we were not naturally fit to.

(01:28:04):

And the way you would talk about this are neurochemistry, our dopamine/opioid access is evolved in a certain environment and then we've created a new environment for ourself that where we're no longer fit. Now it's easy to get more calories and in the early environment it wasn't so there was a dopamine relationship to always get more calories or whatever. Now in this environment, we're kind of actually in some ways genetically misfit. You could say the wisdom was trying to tell us how to deal with the dopamine/opioid access in a environment that we were no longer evolutionarily fit for to deal with evolutionary mismatch.

(01:28:40):

You could also say it was the wisdom was trying to deal with how do we guide the collective relationship to the technological powers we have that has some long-term viability built in.

Nate Hagens (01:28:55):

Number one, as you're speaking on this grand arc of history, I kind of feel sorry for all the peoples and the cultures that expressed wisdom and were out competed by the larger entity, number one. And number two, can we have wisdom with 8 billion humans in our current environment? Or is wisdom only able to be had with a smaller scale, smaller groups that aren't faced with the evolutionary environmental scale limits and pressures?

Daniel Schmachtenberger (01:29:35):

In terms of feel sorry for, now with that beginning of an insight, talk to any indigenous person and you'll get to get the context of their life. Talk to Tyson about this and have them get to share like, "Oh, so the colonialist story was that we were savages that need civilized. Let me actually tell you about what civilization means. What something that was more truly civil was, let's say why the Hobbesian narrative that our lives were short, brutish, nasty, and mean as the apologism to be able to genocide the fuck out of us is the dumbest story in human history." Dumbest and cruelest, right?

(01:30:20):

But here's where also the naive techno pessimistic narrative doesn't work, is that if you don't embrace the tools that will end up being what defines the world, if any of them are there, then you don't actually get to have any say in the future. And which is why right now I am advocating for something that recognizes both the problems of tech and the problems of a multipolar trap where if you just try to get things right for your people and not everybody, you still fail.

Nate Hagens (01:30:54):

Are we ready to talk about artificial intelligence?

Daniel Schmachtenberger (01:30:59):

The reason I was talking a little bit about the evolutionary history of humans was in these smaller scales and totally different types of environment with oral language rather than written language with lots of things very different. But that unique to humans is the capacity to be more changed by their environment. You could say that our nature is to be very nurture influenced, and as much as you can take a wild animal and train it to be different in a different environment, in one generation, how much you can do that with a human is so radically different. And because the other animals are genetically fit to their environment, they evolved to have capacities that work within a niche.

(01:31:48):

And yet humans evolved to make tools that would change our whole environment, including make whole new environments, right? Homes and cities and et cetera. We couldn't come hardwired to a particular environment or a particular way of being because our evolutionary adaptation was not our body. It was our body extended through tools that we change and environments that we change. We had to be able to

upgrade to, "Okay, new sapien, new environment, what language are we speaking? What tools are we optimizing for? What type of en environment are we navigating? We just migrated." We have to come incredibly not prepared for a specific set of things, which is also why we're neotenous for so long. The human sapien is dependent and helpless for a very long period of time compared to the chimpanzee or any other close relative, but also imprint-

Nate Hagens (01:32:37):

And it's why our Mylan sheaths don't… right, keep going.

Daniel Schmachtenberger (01:32:41):

Yeah, we have a totally different… we don't come preset to do the evolutionary adaptive thing within that environment because our evolutionary adaptation was to change how we were adaptive, meaning which tools and everything we use, which languages we use, which tools-

Nate Hagens (01:32:59):

We're probably way more plastic and flexible in our behaviors?

Daniel Schmachtenberger (01:33:07):

Yes.

Nate Hagens (01:33:07):

The elephants today are doing similar things to elephants 10,000 years ago. Yeah, okay.

Daniel Schmachtenberger (01:33:13):

Think about it, a human growing up, learning how to throw a spear versus learning how to use ChatGPT are very different skill sets. And throwing a spear is not that advantageous today in most cities and most environments, even though it's the most advantageous thing. You don't want a kid growing up with genetically in-built spear throwing capacity. You want them growing up to figure out what is the tech stack around me? What's the language? What are the goals? How do I do that thing? Which is why they're also fairly useless at everything to begin with because they haven't imprinted what is useful in this environment yet. The horse is up and walking

in 20 minutes and it takes us a year. And you think about the multiples of how many 20 minutes go into a year for another mammal and you're like, "Wow, that's really different."

(01:33:57):

But then you look at how different human culture now to 10,000 years ago and even one human culture to another is, right?

(01:34:06):

Now we started doing our social science after certain aspects of culture, nurture had become so ubiquitous that we took them for granted. Like alphabetic language, it's not natural. It is not natural to have a written language that is not a natural part of the evolutionary environment. That was an invention. And then we taught it to everybody and it changes the nature of mind very fundamentally because rather than relate to the base reality out there, I relate to the mapping of reality to these sounds that are arbitrary and then how those sounds work. That affects the nature of mind. And then I go to school and I learn this is geography and this is English and this is math and this is history and it's all divided. And there's different principles and the goal is get the right answer and not understand why that's not natural.

(01:34:59):

We are ubiquitously conditioning people and then calling that nature. Whereas if you go to the few indigenous cultures that are left, you'll see that most of what we call human nature in terms of the current life experience of sapiens in the developed west is not true for them. And so it's important to get that our relation… Now do we have a culture that is systemically conditioning? What would both most support wisdom? Look at what are called wisdom cultures, right? Look at the way that a child is raised in say, pre-Mao, Buddhist Tibet and what they're being developed for and how fucking amazingly different it is. Are we genetically misfit as beings to wisdom? No. Did we produce a civilization that is misfit to wisdom and then humans are being made fit to that civilization, i.e. the superorganism, i.e. Moloch, i.e. the mega machine, i.e. the generative dynamics that lead to the meta crisis? Yes. And the humans who are born into that, who become fit in, that are fit for the thing that is killing its own substrate.

Nate Hagens (01:36:15):

I get that that is clear. But we were selected for historically at smaller scales and those systems optimize that had intelligence merged with wisdom to protect the tribe or the

smaller unit. That dynamic doesn't map to my knowledge, to 8 billion in a global consumption based culture.

Daniel Schmachtenberger (01:36:45):

Right. The first question is it not possible for humans, individual humans, groups of humans at different scales to develop wisdom, i.e. relationship to base reality rather than the symbols that mediated and the models that mediate it and relationship to the wider wholes both temporally, spatially, and which agents. Is there anything about human nature that makes that impossible? No. Is there anything about the current conditioning environment of the civilizational system, not the biosphere, that makes that not what is being incented to be conditioned at scale? Absolutely. Is it possible for the 8 billion person thing to continue without the wisdom? No. It will self terminate.

Nate Hagens (01:37:39):

Yep.

Daniel Schmachtenberger (01:37:41):

There is nothing innate to our biology that is the problem. There is something innate to the particular trajectory of the civilizational system that is, and we do not get to continue this civilizational system and what it conditions in people. Then the question is what does it take to create environments that could condition the wisdom in people that in turn reinforce those environments, right?

(01:38:07):

The bottom up effects of the wiser humans creating different societies and the top down effects of different societies having an incentive to develop different things in the people. How do we get that collective individual top-down, bottom up recursion moving in the right direction given that there is no other answer for humans long-term.

Nate Hagens (01:38:29):

Last week on the phone you mentioned when we were discussing AI, that capitalism was an early form of compute. Can you use that as a bridge maybe to get us to talking about AI?

Daniel Schmachtenberger (01:38:48):

Within a body, all the cells in a body have the same genome, but they're epigenetically differentiated so that a red blood cell and the liver cell and a neuron are different. And so there is specialization division of labor that allows synergy across differences and capacities of the whole system that none of the systems on their own have. So, the differentiation and then integration is a thing, even at the level of a body, right, and at the level of a tribe, there was specialization, division of labor, so that the tribe could do more than any individual person could do if everyone was trying to do everything. When we got to-

Nate Hagens (01:39:30):

Kind of the economic theory of comparative advantage?

Daniel Schmachtenberger (01:39:35):

Economic advantage, the story of progress and economic advantage is very coupled to this idea of specialization, division of labor, and the increase in the total cumulative complexity of the space that that facilitated of which capitalism was seen to be the dominant system for doing so.

(01:39:58):

Now, I want to say something before going any further, which is, it is very common that if one says anything bad about capitalism, the default reaction by many people from a historical perspective is, "This guy is a neo-Marxist and is going to suggest something that had Stalin kill 50 million people and Mao kill 50 million people and blah, blah, blah, and doesn't he realize that capitalism is the best solution?"

Nate Hagens (01:40:26):

In my Earth Day talk I did last week, it'll be out tomorrow, I did word pairings, and how the importance of semantics influences our behaviors and gives us permissions, and one of the word pairings was fossil fuels. No, the reality is fossil hydrocarbons, but another was capitalism and communism are both industrial-growth-based systems.

Daniel Schmachtenberger (01:40:51):

Yes.

Nate Hagens (01:40:53):

So, I'm with you there on the exculpatory definitional clause. Keep going.

Daniel Schmachtenberger (01:41:00):

So, when we talk about problems in capitalism, they were different expressions, but many of those problems in terms of environmental harm, optimizing narrow goals, whatever happened in communism, happened under feudalism, happened under various other types of systems that operated at scale. And we're not even going to say that at the very small scale that everything was wise and awesome. That would be a kind of romantic naivete that we don't want to do. We don't want to do the naivete on the other side that says it was Hobbesian, brutish, short, nasty, and mean. Neither the full romantic tribal picture nor the full regressive, they were dreadful animals, picture are true. It was complex, and different ones were different and whatever.

(01:41:48):

But when we are critiquing capitalism now, it's not that we don't understand why it was selected for the things that it did that were more effective or how gruesome some of the other systems were. It's recognizing that this system, on the trajectory it is, happens to be self-terminating, so we have to come up with a new thing. And the new thing will be a new thing. It won't be a previous thing. So, I'm saying that so the default reaction of, "Why am I listening to these commies," someone doesn't have to default into that.

(01:42:15):

But if we have a system, even if it was the best, and we can critique democracy in the same way, Churchill's "It's the worst of all governance systems, save for all the other ones." It's the best thing, but the best thing is still self-terminating, then we have to do new thinking, right? And so, capitalism and democracy should not themselves be golden calves beyond critique, but we are going to do the critique in a way that is historically informed.

Nate Hagens (01:42:45):

A quick question on capitalism, because I've never asked you this. I don't think I've asked anyone this. There was never a person or a group of humans that said, "Hey, let's invent capitalism." It was always an emergent response to the challenges and the innovation and the coordination of intelligence towards problem-solving of the day,

and it took on momentum and then institutions and everything built on top of it, right? There wasn't a, "Let's form this"?

Daniel Schmachtenberger (01:43:19):

There were definitely points in the process that were important and there were points that were conscious in the process. What you actually define as capitalism, right, because classical capitalism versus kind of neoclassical versus Keynesian are all quite different, different than some of the axioms.

(01:43:41):

One could say capitalism started with the beginning of private property. That would be one way that one could talk about it. One could say it began with surplus. The moment there was surplus and we had to figure out who got to do what with the surplus and who has the choice-making associated that, that's the beginning. One could say no, communism was one answer, capitalism was a different answer, so it's surplus plus private ownership and exchange. Or, one could say actually it was only once the medium of exchange was not itself an intrinsic value, so once we got currencies. Or you could say actually only once you had financial services where the currency made more of itself, so the beginning of rent-seeking or loans or debt. Or you could go forward to, not until you actually have a formalized banking system, which you could say started with the Templars or you could say started with Dutch capitalism and the kind of ship-based mercantilism.

(01:44:40):

So, each of those are various steps in the story of what we could call capitalism. Then of course central banking, then international central banking agreements, global reserve currencies, blah, blah, blah, right? I mean, these are topics that, I think for both of us, are just intrinsically fascinating, to understand how the human condition evolved, and they also happen to be maximally important.

(01:45:06):

What's funny, right, you and I have in common is, we are not so motivated by, within the existing dominant systems, what is useful to advance success within that system, but recognizing that those systems have limits that we're approaching what is useful within the system is inherently self-terminating, so what is useful long-term is not based on what's useful in the system. Which means you have to go into, what do we

even mean by useful? What do we even mean by good? How did we get here? So, when you're at the end of a particular kind of success paradigm, if you recognize that, you have to actually go pretty deep in the historical stack, the theoretical stack, to even be pragmatically focused.

Nate Hagens (01:45:56):

"Islands of coherence have the ability to shift the entire system," says Ilya Prigogine.

Daniel Schmachtenberger (01:46:02):

That's true.

Nate Hagens (01:46:04):

So keep going on the why capitalism, why you think capitalism was an early form of compute, and compute is relevant to artificial intelligence?

Daniel Schmachtenberger (01:46:15):

In one of our talks we discussed Moloch, which was kind of an analogy for ... You're actually asking the question now, did anyone design capitalism, or was it kind of emergent step-by-step? And insofar as the most significant features of our world are not the result of anyone's intentional choice, it's like, who the fuck made that? And so, being able to anthropomorphize as a thought experiment what the collective behaviors predispose, you can get different kinds of gods.

(01:46:53):

Adam Smith did this with the invisible hand of the market, right, that the progress was defined by this god of the invisible hand that nobody was choosing the overall topology of the system. They were all doing local point-based choices of, I want this product or service at this price. I would like to make a business that provides this product or service. And yet, the totality of the collective intelligence that was an emergent property of that moved things in a particular direction. He was defining the invisible hand as a good god, right? Roughly, and kind of-

Nate Hagens (01:47:26):

But he did say there was an end to that benevolence in the distant horizon once we would exhaust the surplus, but keep going.

Daniel Schmachtenberger (01:47:38):

So if humans have desire, demand for things that actually improve our lives, and we only want things that improve our lives, and we're rational actors who will make a rational choice to utilize the resources that we have intelligently for the things that improve our lives the most, and that creates an incentive niche for people to innovate how to make goods and services that improves people's lives more at better value, then yes, you would get a good god emerging from that decentralized collective intelligence. And that's kind of the idea of the market and capital mediating that, and having this currency that is not worth anything itself but represents optionality for all forms of value. And-

Nate Hagens (01:48:25):

So was the market an early form of artificial intelligence, the global market, just-in-time delivery, everything optimized for profit, sufficiency, et cetera?

Daniel Schmachtenberger (01:48:38):

If we define intelligence in the pragmatic definition of goal-achieving, if we define general intelligence, narrow intelligence can achieve a specific kind of goal, and artificial intelligence can win at chess, is a narrowly-trained system on one thing. Can when it Go is another thing, can translate languages is another thing, can do image recognition is another thing, can optimize supply chains is another thing. Narrow intelligence. General intelligence is, increasingly, generalists can do more of those things. Fully generalists can, in general, figure out how to optimize any goal. Is the market intelligence-

Nate Hagens (01:49:18):

So, humans have general intelligence?

Daniel Schmachtenberger (01:49:19):

Yes. And one definition of artificial intelligence is trying to make things that behave in ways that we would see as intelligent, so it's just taking the term intelligence kind of intuitively or heuristically, and if it behaves in a way that we kind of would assess that is intelligent, that's one definition. Another definition is that intentionally aim to model human capacities, human intelligent capacities. There are other definitions that are more interesting than those. But when you ask this question about, are humans

generally intelligent? Yes. I think we actually have to back up and define what artificial general intelligence versus narrow artificial intelligence is to explain why artificial general intelligence is scary to then be able to come back and describe this market thing for a moment.

Nate Hagens (01:50:08):

Please do.

Daniel Schmachtenberger (01:50:12):

Probably most of the people here have heard conversation about, are we nearing artificial general intelligence? Maybe we are nearing it in a very short period of time, and maybe that is really catastrophic for everything, and why? And probably some people have a clear sense of why. They've maybe heard Eliezer Yudkowsky or other people like that, Nick Bostrom, talk about the cases regarding artificial general intelligence that are the most catastrophic types of cases that one can imagine, meaning much worse than nuclear bombs. But for anyone who's not familiar, I'll just try to do it very briefly.

(01:50:48):

If I have a narrow intelligence, like the ability to play chess, first, if people don't understand how much better at chess the best AI systems are than the human systems, go study that for a minute and get a sense. Look at how the AI that beat Kasparov at chess, IBM's system, then evolved into Stockfish as the best system that had been programmed with every human game that was so much better than humans, it wasn't even calibratable. And then how a totally new approach to AI that Google innovated, the kind of AlphaGo system, was able to beat that system. There were some ties, but roughly 38-0 without having programmed any human games, just letting two of them play themselves a trillion times in three hours, and get so fucking good based on no human system, just the rules themselves, that they could beat the previous god as if it was nothing. And it took three hours of training and you're like, "Whoa, we aren't even in the running of relevant. We don't even have a reference for how not relevant we are at being able to do anything with that narrow goal." Right?

(01:51:59):

Now, it happens that that narrow system that was able to train that way and beat us on chess also was on Go, also did on StarCraft. You just line them up and you're like,

"Oh, whoa, okay." So, this is a sense of the power of artificial intelligence, meaning intelligence like we have, achieve a goal. Goal here is win at chess. Win at Go, win at StarCraft. Are all those military strategy games? Yes, those are all military strategy games. Could you apply those to real-world military and economic environments? Can you define real-world environments like games, train AI on them to be able to win? Would they dominate that excessively? That is the whole direction of things, right?

(01:52:42):

Does a AI clone-

Nate Hagens (01:52:43):

Real quick naive question. When people are applying AI to war and economic problems and innovation or whatever, it's a series of individual narrow artificial intelligence, they're not applying an artificial general intelligence?

Daniel Schmachtenberger (01:53:07):

There is no artificial general intelligence as far as we know today. It is the biggest goal in the space. We're rapidly moving forward towards it. I'm going to make an argument that there is something like it, but not in the obvious sense. But the computational systems, the purely computational artificial intelligence systems, are narrow, but they're increasingly wide, meaning that the same system that can be trained on one narrow goal can be relatively easily trained on other goals, which means that you don't have to start from total scratch. So, that is increasing generality but not full generality.

(01:53:51):

But if I want to give a real-world example, if you make drones, let's talk about the copter type drones that use swarming algorithms, where they fly in kind of a swarm pattern, and so they can go through obstacles and some of them can be taken out and the other ones will reconfigure. The AI that regulates their flight is so radically advanced beyond what a human controller could do, right, like a human controller trying to control that fleet of drones is not even in the ballpark, just like they aren't in chess. So when you start to think about autonomous weapons and what would that mean, you can begin to get a sense of it.

(01:54:35):

And there are already examples of this, that there's a lot of things that are run by forms of artificial intelligence, cybernetic systems. So AI high-speed trading has already meant that a normal human engaging with a maximum amount of knowledge without those tools cannot play in that space. AI-optimized high-speed trading can only be competed by competitive AI systems. And so then-

Nate Hagens (01:55:08):

Can I ask another naive question?

Daniel Schmachtenberger (01:55:10):

Yes.

Nate Hagens (01:55:11):

So when you say that … Let's just use chess as an example, even though military, finance applications are probably more dangerous and relevant to our futures. But taking chess, you said you have to train the system and the training took three hours. So, what does that really mean? There's some business owners or coders that say, "Okay, we've got these algorithm, this artificial intelligence, and now we want to apply it to be really good at chess." They write a little code and they give the AI the objective and then they just press Go, and three hours later they've got a model that they can apply in a real game?

Daniel Schmachtenberger (01:55:51):

So, the techniques of artificial intelligence, the techniques of learning, those are all evolving and there are … To understand something about the evolution rate, there are many different exponential curves that are intersecting. In general, and you've probably addressed this on your show, in the evolutionary environment in which we evolved, yes, we have abstract intelligence, but we don't have intuitions for ongoing exponential curves, because they never happened in our environments. Something would start as an exponential curve and then it would turn into a logistic curve, into an S curve. And so the rates of exponentials are non-intuitive. People are used to thinking about, "Oh, it's getting faster. We have time to do something," not like, it 10Xd then it 10Xd again than it 10Xd again, and the speed in which it is doing it is dropping by 10X, right, the time period. So we intuitively get this thing wrong. That's for a single exponential.

(01:56:45):

So if I take AI, the hardware, how we are not just having CPUs, but GPUs and TPUs and different kinds of arrangements of those, and network cards and whatever, and how many transistors we can get in a small area, there are exponential curves in the hardware in terms of the increase in progress, Moore's law and other stuff. There are exponential curves in how much data is being created through sensor networks and through social media being able to aggregate all this human data and stuff. There are exponential curves in the total amount of capital going in, the total amount of human intelligence going into the space, the innovation in the models themselves, in the hardware capacities for actuators, and the types of things that can be sensed. And so you have intersections of many different exponential curves.

(01:57:40):

So, what we meant by how an AI learns at the time of Stockfish to the time of AlphaGo already was different. It's a completely different answer to that. And so where Stockfish, you were programming in all of Bobby Fischer's games and all of Spassky's games and giving it opening moves and putting in all of the books of chess and all like that, as a type of learning that then has a lot of human feedback on, is it getting it right?

(01:58:06):

This other technique, all that was programmed is, what are the definitions of a win, and what are the allowed moves? And then you have two versions of the system play themselves with some memory and learning features built in where they know what they're seeking to optimize for is how to win. No human game is built in, no theory of opening moves, but because of the speed, they can really play a trillion games in a small number of hours. Then they have opening moves that have never been seen that are not aligned with any theory, totally different approaches humans have never done, because they can take very different branches. And so they have the equivalent of what would've been millions of years of human chess playing, or maybe no amount of human chess playing, because human memory could never hold that much. Human combinatorics couldn't. But in three hours it can be trained to do that, compared to how long does it take to get a human to just even catch up with what has already been known in chess.

Nate Hagens (01:58:58):

Who owns the AIs?

Daniel Schmachtenberger (01:59:01):

Some AIs are built by corporations, mostly public corporations. Some are built by branches of military or governments, so National Laboratories makes some very powerful AIs, and obviously the AIs that are in public deployment as large language models right now are Microsoft and Google and Anthropic, and then increasingly other companies that have to follow on, Meta and Baidu and like that. But Tesla has very powerful AIs, very, very powerful AIs, that are trying to get full-blown self-driving down, that take information from sensors and process it through artificial intelligence to control actuators. And-

Nate Hagens (01:59:51):

How did it happen that all these are companies that don't really talk to each other? Microsoft doesn't really talk to Google at a competitive sense, so they're just a bunch of engineers that around the world are simultaneously developing these AIs in competition with each other, and they're all figuring out the same codes and et cetera, but some are doing a little bit faster and better?

Daniel Schmachtenberger (02:00:17):

I mean, AI is not that different in this sense than other categories of emerging technology, where you have different companies competing, where anybody can see, once someone deploys, what they did, reverse engineer it, and try to figure it out.

Nate Hagens (02:00:36):

Oh, okay.

Daniel Schmachtenberger (02:00:36):

They can all try to spend enough money to hire key talent away from the other company. There are academic groups that are publishing stuff that is also at the cutting edge of knowledge, and then the companies take that knowledge and do stuff with it. They can all do corporate espionage on each other.

Nate Hagens (02:00:52):

So, there isn't one AI. There are dozens, hundreds, thousands of AIs, depending on what they're applied on and depending what vector of society they're developed in?

Daniel Schmachtenberger (02:01:02):

And different fundamental approaches to AI.

Nate Hagens (02:01:09):

What do you mean?

Daniel Schmachtenberger (02:01:10):

Well, there are AIs that are trying to take visual sensor data to control motor actuator data to move a self-driving car, which is different where it's sensing and actuating a movement of a vehicle through space. There are other ones that are trying to read content online, read language, your ChatBot, and so they're sensing in language and they're actuating in language, so they're optimized for language input/output. So, those are very different kinds.

Nate Hagens (02:01:42):

So, the goal of AI in these examples is not profit maximization. The goal of the people that control the AI is profit maximization.

Daniel Schmachtenberger (02:01:54):

The goal of the AI is whatever people code its goal to be, its objective function or its loss function, the inverse of loss function, but the people are developing it in service to whatever the goals of the organization they're part of is seeking. So, if it is a military group, it might be the goal of national security. If it is a corporation, especially a public corporation, it has to be profit maximization through some specific domains of action. If it's an academic group, it might be advancing the knowledge of that field in whatever ways that is.

(02:02:35):

So it is an important thing that you're asking, which is, if the group that is developing an AI has certain goals, the AI will be developed in service to those goals. And so, are those groups wise? If the AI is super intelligent, and we're already starting to get at that intelligence has to be bound by wisdom, is the goal of the group a wise set of

goals, or is the goal of the group a narrow set of goals, in which case it'll be building the AI for the optimization of that, is actually a really key question.

Nate Hagens (02:03:11):

But here's a subset of that question. If a few of the groups choose wisdom as their goal, or if most of them do, they'll still be out-competed by those AIs that have the narrow focus and win, no?

Daniel Schmachtenberger (02:03:28):

So, the question of, are multipolar traps obligate, or is there a way to get out of them? Agreement is a way out of a multipolar trap. The agreements are hard, because if everybody can realize that their likelihood of winning the race is low, the race as a whole might mess everything up for everybody, and not doing the thing is better, was mutually-

Nate Hagens (02:04:01):

Like WarGames with Matthew Broderick.

Daniel Schmachtenberger (02:04:03):

Yeah, mutually assured destruction was a way of saying, "Don't think that you can win first strike advantage." Let's be clear that winning equals losing. At this level of power, there are things built in place where winning equals losing, so we have an agreement to not try to do the win, because anyone trying to do the win would be so catastrophically bad for everybody. And the challenge with these agreements is, you can enforce such a thing within a nation, right? Within a nation where you have a monopoly of violence and backing up rule of law, you can make sure nobody cuts the national forests down, even if there's a lot of motivation to do it, because you can make a law that says, "You're not going to," and the monopoly of violence is stronger than whatever the capacity the logging in companies are going to bring to bear.

(02:04:55):

But if you don't have monopoly of violence and rule of law, how do you enact it? So internationally, and this is why we face so many global catastrophic risks, whether it's the destruction of the oceans or the destruction of the atmosphere or biodiversity or developing synthetic bio or AI is, if anyone else is, we lose in the short term by not

doing it, should they deploy it. And yet if everybody does, most likely everybody dies, so how do we bind those? The tricky thing is, how do we know the other guy's keeping the agreement, and that they're not defecting and secretly doing it? So with regard to AI, can you get … And so we see that OpenAI was originally created for safety purposes and concern about how far ahead DeepMind was than everybody else and the need to be able to have an open source approach, so there was not only one essentially concentrated power. And through whatever series of things happened, the thing that was originally a non-profit dedicated to safety is a public corporation for-profit, wholly-owned subsidiary of Microsoft, deploying in competitive types of races against other ones. And then Anthropics that broke off from there to do safety that took 300 million from Google ends up being in similar dynamics. And so, without some kind of shared agreement that nobody does a particular thing, then yes, you end up having an incentive gradient for everyone to race to get there first, even ethically, where the whole me getting there first, I'm the ethical guy, so I'll win that race and then I'll use that power for good.

(02:06:33):

That conversation is pretty accurate-

Nate Hagens (02:06:38):

Yeah. It's like a billionaire being aware of The Great Simplification and still trying to maximize their optionality to have more money so that they can do more philanthropic good in the future.

Daniel Schmachtenberger (02:06:49):

It is amazing how powerful motivated reasoning is on bias.

Nate Hagens (02:06:55):

Let me ask more naive questions, Daniel. We're getting to the heart of it here. Could you program AI for wisdom?

Daniel Schmachtenberger (02:07:06):

This is the AGI alignment question, and there's different ways of getting to this, so let me construct the AGI concern versus the narrow AI concern for a moment. We were on the track of, what is narrow AI? It is optimized for a certain set of goals and it can get

extremely good at it. It can get better than humans by not just a little bit, but in many domains, by so much that we have a hard time even understanding it.

Nate Hagens (02:07:34):

And that so much as we speak is accelerating by the month.

Daniel Schmachtenberger (02:07:38):

Yes. And so, without fully general intelligence, fully autonomous systems, then the AI is in service of humans using the AI for purposes. The humans are the agents. The AI is an enhancement of our goal-achieving capability, right?

Nate Hagens (02:07:57):

It's just a tool.

Daniel Schmachtenberger (02:08:00):

I would argue that even the word tool makes us think badly, because language as a … And this is why I was saying technology meaning goal-achieving is such a broad set of things that the types of tools that an end user uses versus the types of tools that make other tools, machine tools, versus the types of tools that innovate how to make lots of other tools like computation, yes, they're all tools. But they're really fucking different in kind. And we wrote this paper called Technology is Not Values Neutral, that should link here, which discusses why it's not just that we have our values and we use our tools. It's that the tools give the capacity to do certain things better, to achieve certain goals better. And insofar as they do, the humans that use them win at things, and as a result, everybody has to use them or they kind of lose.

(02:08:53):

So one, the tools become obligate. Two, the tools code different patterns of human behavior. Now I'm doing the behavior where I'm using that tool as opposed to doing some other thing. Encoding that pattern of human behavior changes the nature of human mind and societies at large. So, it is not true that you've got values and then tools are just neutral. The use of the tools changes the human mind individually and aggregate, and then becomes obligate. So, psychology affects the tools we create, but the tools in turn affect the psychologies. And this is why I was saying, our nature is to be malleable to the environment because we change the environment so much, but if

we don't consciously know how to direct that, it will get directed by a kind of downhill gradient that ends up in an evolutionary cul-de-sac.

(02:09:43):

So, with that little aside on tool or technology is actually a very deep concept. Yes, A narrow AI that is not an agent that a human is using is a tool, but it's a tool of a very special kind where it can take lots of steps on its own to help achieve a goal rather than me just use it for fully specified purposes. Right? So, this is a kind of important distinction.

(02:10:17):

Now, if I look at AI benefit first, so as to not seem overly negative, for all goals humans have that would create some progress or benefit for some real thing, can we use AI in service of those? Yes. Not all things equally today, but lots of things. Is it all the scientific progress we have on solving new diseases and stuff like that? Can we use AI to accelerate it, and accelerate movements in science and discovery medicine? So, might AI be able to speed up the rate by which we come up with a way better nuclear energy, maybe fusion, maybe deep geothermal? Yes. Might it be able to solve many types of cancers? Yes. If I have a daughter that is dying of cancer, do I want to hear anyone slowing down the systems that if they get there fast enough might save her life? I don't want to hear that. That sounds like the most cruel, evil fucking thing. Whatever other problem it's going to bring about, I'll deal with that problem later.

Nate Hagens (02:11:19):

Is there someone listening to this that cares mostly about the environment and climate change? Is there a progress case to be made that AI will help carbon emissions and reduce environmental impact?

Daniel Schmachtenberger (02:11:35):

Can I use AI to model how to do geoengineering more precisely? Can I use AI to do better genetic engineering on crops to maximize their carbon sequestration, if I'm a carbon fundamentalist? Can I use AI to advance stem cell meats? Can I use AI to advance energy technology or battery storage technology or any number of things like that? Can I use AI to affect supply chains, to lower energy on supply chains and decentralize a lot of things where you don't have as wide a supply chain. Yeah. You can use AI on all those things.

Nate Hagens (02:12:06):

Okay. Got it.

Daniel Schmachtenberger (02:12:08):

And so for any goal we have, if I can use that to enhance the goal, I don't want to hear anybody talking about slowing that down.

Nate Hagens (02:12:16):

Yep. I agree. Or at least I understand.

Daniel Schmachtenberger (02:12:20):

This is the same as capitalism, right? If me having access to more capital speeds up my ability to achieve my goal, and I believe in my goal, then I don't want to hear anything about taxing that or decreasing that or fucking up my capacity to achieve those goals.

Nate Hagens (02:12:34):

Yep.

Daniel Schmachtenberger (02:12:34):

And, so are there a lot of things that human intelligence can do that are good that increasing that type of human intelligence through artificial systems that can operate on more data and faster could also do? Yes. Now, why would we be so concerned about it then? So let's talk about a few different cases. One type of AI risk is AI employed by bad actors. Now, nobody thinks they're a bad actor for the most part, right? There's some exceptions, but most people that someone else calls bad actors, right? They're terrorists, but to them, they're freedom fighters. Was it a protest or was it a riot? Blah, blah, blah. The layoff frame on those things very much, is it progress or is it destroying all these areas for the people that are being destroyed by it who want to do some destruction back that they consider tiny in the name of self-protection, it is a protest and a freedom fight.

(02:13:41):

Otherwise, it might be a terror act or a riot, right? But let's just pretend that we weren't thinking about all that and just call it bad actors for a moment, criminal

activity. Can people, whether it is a dude that would've just shot up a bunch of stuff with an AR-15 because that was a technology he could use to achieve harmful goals, and obviously this is why people are concerned about assault weapons is because they couldn't kill as many people as easily with a knife, right? Or a single loader pellet gun when the Second Amendment was created. Well, they can do way more harm as drones become more widely available that you can hook explosives to. And so it happens to be that the bad actors who want to do up fuck stuff oftentimes are in states of mind where they're not amazing technologists and can't coordinate lots of people and do strategy and technology.

(02:14:39):

This is not always true, but sometimes it has been true. We've been saved by this. But as we make the destructive capacity, now it might have constructive capacity. I can use a drone of course, for a lot of positive things to be able to monitor construction and railways for safety and to plant trees from the sky and whatever. But can we also use it to fly an explosive over some critical infrastructure? So when we make the thing for the positive purpose, we also enable all the things that it can do for any purpose. When we make that easy enough-

Nate Hagens (02:15:11):

And in that case, probably 80% plus of those bad things we can't even imagine at the beginning.

Daniel Schmachtenberger (02:15:21):

That's a complex topic I want to get to is, how you do externality forecasting. Because we can do a much better job than we ever have done and saying that we couldn't is a source of plausible deniability for not trying.

Nate Hagens (02:15:36):

Well, what I meant was there's unknown unknowns, to use Donald Rumsfeld terms with AI. Yeah.

Daniel Schmachtenberger (02:15:42):

But if we're… okay, I have to do this tangent then because it is fucking critical.

Nate Hagens (02:15:50):

Okay.

Daniel Schmachtenberger (02:15:51):

There are problems that-

Nate Hagens (02:15:52):

Well, we're already in line for this to be the longest podcast I've ever done, so let's do it right? Keep going.

Daniel Schmachtenberger (02:15:59):

Because there are problems that are second, third, nth order effects that were not easy to anticipate, can we use a technology for a certain purpose that creates unanticipated and maybe even unanticipatable consequences? Yes. Can you prove that you can anticipate and advance everything? No, because there will be some things that are unknown unknowns that you can't have proven that you thought through in the safety analysis ahead of time. So that's a true thing. But because that's a true thing, people use that as a source of bullshit plausible deniability to say, I couldn't have possibly known. Where then they don't even really try to forecast the things because they will privatize the gains and socialize the losses. And this is a very, very important thing to understand. Also related to the progress narrative and the underlying optimist versus pessimist or I would state opportunity versus risk orientation.

(02:16:55):

The people who focus more on the opportunity of a new technology, this is going to do all these amazing things, blah, blah, blah, they're going to move faster, focused on that than the people who are focused on the risk and really want to do good thorough safety analysis and make sure it won't cause any of those risks that takes a lot of money and a lot of time and doesn't rush you to market as fast as possible. So the first mover advantage is going to happen by the guys who take risk less seriously, focus on the opportunity more. They'll be able to get the network dynamics that are very, very powerful from associated with first mover advantage and early scale. They'll be able to regulate in their interest because they state that the risks are not that bad, and they do lobbying efforts with the money they got from early revenue or investment or et cetera.

(02:17:42):

The people who take the risks seriously do the analysis and say, oh, fuck there's actually no way to advance this that is good right now. We just should just not do it. Well those people just get to sit on their knowledge of the problem and not do anything, but also not gain any power to affect the system because they're not going to generate money by which they can do lobbying and public opinion effect and et cetera. Or they say there is a safe way to do it, but we just spent all of our money figuring out safety, and not doing optimization. So there is a perverse incentive in general for those who are more focused on opportunity than risk. And as a result, we get all the opportunities and all the risks and the risks happen, the cumulative effect-

Nate Hagens (02:18:24):

And that's true with capitalism and the market in the last 50 years.

Daniel Schmachtenberger (02:18:29):

It is a market function, not 50 years. Forever.

Nate Hagens (02:18:33):

Right. Okay. Yeah.

Daniel Schmachtenberger (02:18:37):

So but the risks are getting larger as the technology gets larger and the cumulative effects are increasing. So now we're at a place where the cumulative effects of industrial tech on the environment are reaching critical tipping points of planetary boundaries, and the exponential tech is getting to a place where it's capacity, the destructive capacities are so fucking huge. Right. The atomic bomb was the first thing that could destroy everything quickly. And until then, for all of human history, we couldn't do any quick thing that would destroy everything. Now, the atomic bomb is not the only thing like that. Synthetic bio where you can make an artificial lifeform that doesn't have a few natural genetic mutations, but so many that it could be an invasive species everywhere. Right. Or artificial intelligence. These are examples of things that where their destructive effects are actually exponentially more than any previous kinds of thing.

(02:19:33):

But if you combine things like a public corporation having a fiduciary responsibility for shareholder profit maximization, which you can argue why that made sense. You can argue why the shareholders are giving you their money to work with. Those shareholders come from pension funds. You need to get that back to them. They can only trust you to have it if you have this bound fiduciary responsibility to return their funds appropriately, and you're innovating a good or service that people want that is not good for the world, blah, blah, blah. Of course, as you can see, where the logic of the market became less and less true as rather than rational actors, we figured out how to nudge everybody into less and less rational action. And the supply side figured out how to manufacture demand for things that don't increase the quality of people's lives and didn't account for the costs and the environment.

(02:20:24):

You still had that story. So you've got the maximize profit, and then you have nobody would want to innovate. And the innovation is good for everyone if the bad things that happened they were personally liable for. So we'll make liability limiting properties where the corporation will get fined, not any of the people or directors involved. So then doing things that destroy everything just becomes a cost of doing business. No real deterrent for it. So the corporation will privatize the gain, socialize the losses. The actual people who make the decisions have a bunch of upside and not relative downside. You put all those things together and you say AI is being developed in those environments. And so it has to do profit maximization. It does not have the people who are making the decision have the liability associated with the scale of risk and harm that could occur. Those who are more opportunity focused, build the corporations that become worth tens of billions of dollars, and have all that power to also influence lobbying and influence public opinion. And those who do the safety analysis run tiny nonprofits that nobody listens to comparatively. So it's important to get that this asymmetry that perversely orients those who think about opportunity more than risk, as risk is moving into an exponential scale and then rationalize that. That is itself one of the underlying drivers of the meta crisis.

Nate Hagens (02:21:52):

And embedded in that and I understand that, embedded in that is the natural human scale, ethical feedback loop. When people are innovating and they go and work in a room for months and they're working on something, they kind of get a little bit of

recognition and an insight into what they're doing. Here you code something and you press a button and all those negative potential externalities are just in the future, and you get no emotional sense of what's going on. Is that also true?

Daniel Schmachtenberger (02:22:25):

Yeah. The fact that it happens at a scale like many of the harms will occur via supply chain somewhere in the world, I won't see. And as we already said, there'll be second and third and fourth order effects. So let's say for instance, now this is a very important point. We can say that all technology is dual use. Dual use being a military term, meaning, and there's two different ways to think about it. A technology is developed for a military purpose, so its primary use is military. It also might have civilian or non-military applications. Right. We're developing the rocketry capability to make missiles hit their targets, but maybe we'll also be able to use it to put satellites into outer space for communications for everybody.

(02:23:26):

And obviously computation was developed in World War II to crack the enigma machine and whatever for military purposes and has had a lot of other purposes. On the other side, we say anything that is developed for a non-military purpose probably also has a military application, right? So dual use goes both ways. The fact that if I'm developing something for a non-military purpose, it still probably has a military application, i.e, you can say on the positive side defense, but on the opposite side, the offense or killing capability means that that has to be factored in the development of the technology.

Nate Hagens (02:24:01):

But dumb question though, if someone at Microsoft is creating an AI for some purpose, they're not sharing that with the US government or any government, right? It's got to be independently developed by coders and engineers within the government, within the military?

Daniel Schmachtenberger (02:24:20):

No, no, no, no.

Nate Hagens (02:24:20):

No?

Daniel Schmachtenberger (02:24:24):

If a technology is seen as having risks to national security, then there will end up being government bodies that have oversight into certain capacities.

Nate Hagens (02:24:38):

Got it.

Daniel Schmachtenberger (02:24:41):

But that's one way. The other way would just be the corporation develops the capacity and one of its clients becomes military and becomes a military contractor in addition to other things.

Nate Hagens (02:24:49):

Got it. Got it.

Daniel Schmachtenberger (02:24:50):

One would be, it doesn't even say Microsoft doesn't develop a military application. It develops LLMs, and then Palantir develops competing LLMs and makes a military application of them. Right. So the technology itself will get developed lots of places, will get reverse engineered used for lots of purposes. Now so all technologie is dual use. Now on the side of we're developing something for a civilian purpose that we say is positive, but it also has military. Okay, we have to think about that. But the other side is also risky because if you're developing a military technology that is very powerful and very dangerous, but at least you think you can control it, the moment that that same capacity also gets a civilian purpose means it will proliferate then it makes it much harder to control. Like let's say you developed drones for military purposes, now it gets a commercial application, which is I can just fucking film stuff with it. Now everybody can get access to drones. Wow, that just really affect the capacity for decentralized terrorism. So okay, tell me where you want to go. I'm about to go to the next part beyond dual use.

Nate Hagens (02:25:58):

And you can. I'm just going to comment that I get this sinking feeling that we're headed towards a risk singularity.

Daniel Schmachtenberger (02:26:09):

Yes.

(02:26:11):

The meta crisis is a risk singularity in which the underlying drivers overdetermine failure, meaning if we could prevent the AI apocalypse, that doesn't prevent the synth bio one or the planetary boundary one or the gazillions of other ones. If we could stop planetary boundaries regarding fishing, that doesn't affect what we're doing to soil or nitrogen runoff or PFAS pollution or whatever. The underlying thing is creating so many different sources that can lead the catastrophic risk. That if you don't deal with the underlying thing and you just deal with some of the risks, you only buy a tiny bit of time.

Nate Hagens (02:26:54):

I have thought that the underlying thing was the emergent phenomenon that I call the superorganism, which is the growth compulsion of our global market system. But what you're saying is that we had agricultural surplus, then we found flammable fossils, then we accelerated our technology, then we went to debt, then we had the internet, and each one of these kind of exponentially increased what had come before. And you're saying that AI is the next quote, unquote tool in service of this growth based superorganism.

Daniel Schmachtenberger (02:27:38):

Yeah. So you say growth, let me, which is, that's a fair way to say it, but let me define it slightly differently because I'm going to define it in a way that is more aligned with this AI conversation. So I think it actually gives a lot of insight into the other. Narrow goal achieving is the underlying generative dynamic.

Nate Hagens (02:27:58):

Okay. I get it.

Daniel Schmachtenberger (02:28:02):

And growth is an epiphenomena of that.

Nate Hagens (02:28:08):

Wait, an epiphenomena?

Daniel Schmachtenberger (02:28:10):

And undernea- a result of. A second order effect.

Nate Hagens (02:28:15):

Okay, so the first thing is the narrow boundary goal and growth is the second order effect of that narrow boundary goal.

Daniel Schmachtenberger (02:28:26):

Yes. So if I want to achieve a goal and having a little more surplus gives me more optionality to do that, but me having that more optionality when I'm rival risk that increased security, I have inherently decreases somebody else's security or their relative competitive advantage in a status game for mating or whatever it is. So now they have to do that. Then I see them doing it. So now I need to do more than I needed to do. Now we're going to race on it. Growth is the epiphenomena. What everyone is pursuing is not the growth of the whole system. They're pursuing their own narrow goals and generalized optionality for goal achieving.

Nate Hagens (02:29:07):

Okay.

Daniel Schmachtenberger (02:29:10):

And the growth in total consumption of energy and atoms, the growth of total waste and entropy, the growth of intelligence and new types of technology, the growth in memetics and the ability to convince a lot of people of it are all epiphenomena of both goal achieving and specific. And then the general capacity of increasing optionality for goal achieving.

Nate Hagens (02:29:34):

Could we ask AI how to go from a narrow to a wide boundary goal?

Daniel Schmachtenberger (02:29:42):

So all technology has certain affordances and all technology has combinatorial potential with other technologies, and the affordances of them together are different than them on their own. And obviously, not only does the hammer have different potentials, if I also include nails and saws and the other things, and it would on its own, but I can't even get a hammer without the smithing tools that would be necessary to make a hammer, right? So there's a whole technological ecosystem.

Nate Hagens (02:30:17):

I'm just, I mean, I'm cutting to the chase on this. I'm just wondering what AI does to our current complexity.

Daniel Schmachtenberger (02:30:23):

Yes. So there's a reason there's a reason I'm having to try to do this. All tools have combinatorial potential with other tools. They all have certain affordances. AI is unique. Computation has been unique. And then AI, even within the space of computation in that it is more combinatorial with every other type of tech than any other tech is, and it has more total affordances than any other tech has.

Nate Hagens (02:30:56):

So what's an affordance again?

Daniel Schmachtenberger (02:30:59):

Things that it makes possible.

Nate Hagens (02:31:01):

Okay.

Daniel Schmachtenberger (02:31:04):

So for instance, if I have a nuclear weapon, it makes certain things possible that are not possible without it. It does not intrinsically make me better at bioengineering. It does not intrinsically make me make better drones. Computation can be applied to make better drones, better bioengineering, better nuclear weapons, better propaganda, better marketing, better financial systems, better chips to make better AI systems. So the AI, because all the other tools are made by the kind of human

intelligence that makes tools and AI is that kind of human intelligence externalized as a tool itself, it has a capacity to be omni modal, right? Not dual use, omni use more than anything else is, and omni combinatorial. The risk of AI with synthetic bio... So the dual use thing here is I'd say, oh, I can use AI and kind of protein folding stuff to be able to solve cancers, but once I develop it for that purpose, I can also make better weapons with it and now I've made that ability cheap and easy.

(02:32:18):

Or I can use it on a chemical database to do drug discovery, but I can also make chemical weapons. I can use it to optimize supply chains, but I can also use it to do optimize terrorism on breaking supply chains. And so the AI has the capacity to increase every agent's capacity to do every other motive in any context with any other combinatoric technology in a way that nothing else has. And that's really important in understanding because we'll develop it for a particular positive purpose, for a particular set of agents, but in doing so, lower the barrier of entry of all other agents for all other purposes being able to use that capacity.

Nate Hagens (02:33:05):

So from my biophysical frame, which is not where you're coming from, but I'll just interject this, I think a lot of people who don't fully understand AI look at the invisible but real biophysical body that AI has, which requires energy. I read in the last few days that AI requires a lot of water for cooling and things like that. But the larger environmental impact of resources, CO_2 planetary boundaries is not what's needed to power the AI. It's the resulting acceleration of all the other consumption and things that were already going on that are just being exponentially increased by the AI.

Daniel Schmachtenberger (02:33:55):

Yes. So let's look at AI driving Jevons paradoxes in your world. So let's say there's a bunch of areas where the mining is not quite profitable, right, that the cost of extraction, I can't on the current market sell the thing for enough money. So we don't mine that area. Now the AI figures out efficiencies such that a whole bunch of areas that weren't quite profitable are now profitable. Awesome for the economy, new economic growth for everybody, and also more rapid growth of the superorganism and more environmental externalities for everyone under the current motivational landscape and incentive landscape, and down to both the market incentives and even

the legal landscape because the market incentives won't create a real liability deterrent because of liability limiting things, and they actually require you to do profit maximization and things like that. Right. So-

Nate Hagens (02:34:45):

How is there not a massive discussion in the climate space about AI right now as a generator to bring us to 500 PPM in a decade?

Daniel Schmachtenberger (02:34:57):

I don't know. Maybe this conversation will help with that.

Nate Hagens (02:35:02):

But you're very worried about that. You're worried about the planetary boundaries impact of AI.

Daniel Schmachtenberger (02:35:09):

Either you have the market still running things and AI working within the market, in which case, yes, the market incentives have cost externality as fundamental to them. If you had to pay the real cost of things, the market would collapse as we understand it. And so the AI causing more efficiencies will drive Jevons paradoxes because you will continue to seek returns and you will drive planetary boundaries faster. The AI person might argue, no, this is not true, because the AI allows us a completely new economic system whereby we can actually track those externalities and do very complex optimization. That means the market doesn't run the world anymore. That means a centralized AI system runs the world. That is its own dystopia because either you have separate competing groups that don't have totalizing power, in which case they're caught in multipolar traps, or you have a single group that can bind all of those.

(02:36:09):

But now there is a central point of failure capture and corruption and no checks and balances on the power. That's why we talk about one bad future being catastrophes, the other being dystopias, because to control for all the catastrophes orients towards very centralized power, which gives dystopia. So what we need is a third attractor that is neither of those, which means it is neither a central AI coordination system, nor is it AI and service of markets, which means it is some new thing that like markets as a

collective intelligence system are also a collective intelligence system, but it doesn't have the perversion and externalities of markets. So a wiser collective intelligence system wiser than democracy or capitalism. That yes, it will employ AI and computational capabilities to mediate just in the same way that democracy required employing the printing press.

(02:37:01):

And so new physical technologies enable, and just like computation has enabled all of modern banking. New technologies make new social systems possible. But the goal here is not an AI overlord that runs everything. It is how do computation and intelligence capacities make new collective intelligence possible such that you can have the global coordination to prevent global catastrophic failures, but where that system has checks and balances in it so you don't have centralized power coordination and corruption failures.

Nate Hagens (02:37:35):

So it seems to me that if we weren't in this meta crisis and Wiley coyote moment with all the leverage in the system and the geopolitical fragility and everything, that we could have time to have some wisdom embedded in our information systems, in our governance. But we're at this late stage where the snowball is getting bigger and bigger, and how do we inject wisdom into this situation with all these things going in the opposite direction? Obviously you must have some idea or a plan or you wouldn't have dropped all this really scary stuff on me and the listeners. Or are you just calling this out as a risk that we need to have massive urgent discussions on?

Daniel Schmachtenberger (02:38:36):

I think there are people who are AI experts, which I'm not. I have a focus on how to think about the relationships between all of the kind of ideas of progress and risks that gives an insight into AI. But when you look at whether it's Hinton or Russell or any of the famous kind of AI pioneers and the risks that they have called forth or the call to pause large language model deployment, or the most serious case like Eliezer Yudkowsky's conversations. If people haven't seen the first podcast he did that kind of started this space taking these things more seriously recently on Bankless, it was so fascinating to watch because you could tell the podcasters didn't really know what he was going to say and thought he was going to talk about AI and crypto. And he came

on and he founded Machine Intelligence Research Institute and spent his life focused on the topic of the risks of artificial general intelligence that he thinks we are both very close to, but not close to aligning.

(02:39:47):

And to understand that, which we haven't discussed yet because we keep not quite getting to the whole construction. You have narrow intelligence which humans can use for fucked up purposes and markets can use for fucked up purposes. But then a general intelligence becomes kind of its own thing. It becomes its own agent rather than a tool of us as the agents. Can make its own goals and can increase its speed of learning, meaning in competition relative to us so much faster than us, that if its goals don't align with ours, we're fucked.

Nate Hagens (02:40:17):

So using the chess example, okay, I was just going to say using the chess example of the three hours and trillions of iterations, that AI could use itself. It could train itself without humans giving it the order.

Daniel Schmachtenberger (02:40:30):

Yes. But across broader things. So it could come to a new domain where it wasn't already trained, what to do and figure out what are my goals in this space and or for whatever my big picture goal is, what would the goals on the path, the instrumental goals be regarding this whole space? And so kind of in the way that you say, no matter what your goal is, it's probably going to require energy. So get more energy ends up becoming a goal in and of itself because it increases optionality to fulfill other goals. Get more capital becomes a goal in and of itself. It's called an instrumental goal.

(02:41:06):

The AI can start figuring out, well, I can gain all the optionality in this space by such and such. I'll mine all the things, I'll et cetera, et cetera, because I'll be able to use them for goals in some future time that I have. So if we talk about a general autonomous intelligence of that kind, that could beat us at any games, can redefine anything as a game that can win, it's very easy for us to realize it would be really bad to make that thing before being sure that its goals don't mess up everything for us, because our goals really sucked for all the species we extincted. Our goals really suck

for all the animals in factory farms and for all the cultures that are destroyed. So if there's something that is as much smarter than us at goal achieving then us and similarly narrow goal achieving, then we might be like animals in the factory farms or extinct animals very soon. And so-

Nate Hagens (02:42:04):

Or the cultures with wisdom that were out competed.

Daniel Schmachtenberger (02:42:10):

So Eliezer came on this podcast having kind of world-leading expertise in this topic and just spoke honestly to these podcasters who they did a good job of staying with him, but were totally not prepared. And he's like, yeah, we're on our way to this thing and we're all going to die. And then they asked him, so what do we do about it? And he is like, I don't know. I've been working on this for decades and everything that I thought would work doesn't work. And yet the market is driving everybody ahead and I have no idea. And then they're sitting there -

Nate Hagens (02:42:41):

How is -

Daniel Schmachtenberger (02:42:44):

Wait, let me get this last part. You should probably watch it.

Nate Hagens (02:42:47):

Yep. Yep. Sure. I will.

Daniel Schmachtenberger (02:42:49):

Then they're like, what do you want the listeners to take away? And he's like, I don't know. I just feel like at the end we should at least be honest with each other. It's a very poignant moment. Right. It's a very poignant moment. I'm not doing that right now. I actually do think that there is a way forward. But I think what he said shows what someone who is very bright who spent their whole life looking at this, like it's very important to take that as a data point. But you were asking this question, okay, what do I want listeners to take away? So I was giving as a reference, arguably what many

of the top experts in the world come to with this. So I will offer what might be a way forward, but it does require taking seriously how deep a thing we're talking about.

Nate Hagens (02:43:34):

I can understand how narrow AIs are pursued by corporations and by governments to achieve task and achieve profits. But why would a corporation pursue artificial general intelligence if profit was their objective? Wouldn't that also run the risk of killing everyone in these other extreme scenarios that then profits aren't worth anything if the whole system blows up into a giant superorganism?

Daniel Schmachtenberger (02:44:10):

You could ask, why would any country pursue having nukes when the more nukes that exist increases the probability that everyone including them dies with nukes? And yet you can see from-

Nate Hagens (02:44:21):

So it's an arms race.

Daniel Schmachtenberger (02:44:23):

Yes. It's partly an arms race, and it's partly a set of these biases that those who are more focused on the opportunity than risk end up being the ones who rush ahead. And so the people who don't think AGI will kill everything, try to build AGI because it will solve everything. Wow. Imagine intelligence like ours, but so much more. It'll solve all of science. It'll give us radical life extension. It'll create wealth and abundance for everyone. And we will get ushered into-

Nate Hagens (02:44:50):

So we've got the naive progress people shepherding the AI development train.

Daniel Schmachtenberger (02:44:59):

The naive progress combined with the motivated reasoning on capital and winning and ego and all of the various sources of motivated reasoning, combined with once they start actually being legally bound for things like profit maximizing and that the liability is inherently externalized while the profit is centralized, then combined with things like the country doesn't even want to regulate them because it wants the

economic growth so that it's not fucked because it uses that economic growth to grow militaries and geopolitical alliances. And so then you get a, "Yeah, but if we regulate and China won't, then we lose." So you get layers and layers of multipolar traps driving the need to continue to optimize for near-term narrow interests.

Nate Hagens (02:45:46):

This is not where I thought the conversation would go, really. I'm more depressed than I thought I would be. You and I have talked about this, so I'm kind of aware of some of these risks and there are risks that we haven't talked about and there's information hazard with some of these risks, but this sounds highly plausible to me, and I think even before AGI would be reached, there are plenty of other, well just using AI as a vector to accelerate all of the things, that itself is enough to put us in the tipping point.

Daniel Schmachtenberger (02:46:25):

And I know we are late and the sun has been setting in the beautiful mountain background behind me, so we need to wrap soon. So I want to try to bring back a few threads. Human intelligence, unbound by wisdom, it is fair to say is the cause of the meta crisis and the growth imperative of the superorganism or the capacity that gives rise to it. That intelligence has created all the technologies, the industrial tech, the agricultural tech, the digital tech, the nuclear weapons, the energy harvesting, the all of it. That intelligence has created all those things.

(02:47:05):

It has made the system of capitalism, it made the system of communism, all of those things. Now, that system of intelligence that takes corporeal capacities, things that a body could do and externalizes them the way that a fist can get extended through a hammer or a grip can get extended through a plier or an eye can get extended through a microscope or a telescope or our own metabolism can get extended through an internal combustion engine or there are musculature or whatever it is. So it takes the corporeal capacity and extends the fuck out of it extra corporeally. So that type of intelligence that does that is now having the extra corporeal technology be that type of intelligence itself in maximized recursion, not bound by wisdom, driven by international, multipolar military traps and markets and narrow short-term goals at the expense of long-term wide values.

(02:48:12):

So in the meta crisis, there are many risks. Synthetic biology can make bad pandemics and extreme weather events can drive human migration and local wars and this kind of weapon can do this and this kind of mining can cause this pollution and this kind of pesticide can kill these animals. Those are all risks within the meta crisis. AI is not a risk within the meta crisis. It is an accelerant to all of them, as used by the choice-making architectures that are currently driving the meta crisis. If we weren't in the meta crisis, if we had different choice-making architectures, we'd be using AI for different things. But if AI is in service of human goals and human goals have driven the superorganism and the meta crisis the way they are, then this context of human goals as an accelerant of them all.

(02:48:59):

Now if I think about AGI risk, we make an AI that is fully autonomous. We can't pull the plug. It has goals that aren't ours. It starts optimizing for those and then the process decides to use all the atoms for something else and completely terraforms the earth, which it could do. It could totally do, and we're moving way faster towards that than we are towards safety on that. That is a risk within the meta crisis landscape. But AI being used by all types of militaries, all types of governments, all types of corporations for all types of purposes, achieving narrow goals, externalizing harm to wide goals is an accelerant of the meta crisis on every dimension. And so now as we take the intelligence that has driven these problems unbound by wisdom and we exponentialize that kind of intelligence, we get to see, "Whoa, super intelligence is really fucking potent. What goals are worth optimizing with that much power with something that's a trillion, trillion times smarter and faster than humans? What goals are worth optimizing?"

(02:50:02):

It's not global GDP because I can increase GDP with war and addiction and all kinds of things and destroy the environment every definition and so I say, "Okay, it's GDP plus Gini coefficient plus this other thing plus carbon plus whatever." Nope, there's still lots of life that matters outside of those 10 metrics or 100 metrics that I can damage to improve those. The metrics set that is definable, like the Tao teaching says, "The Tao that is speakable in words is not the eternal Tao." The metric set that is definable is not the right metric set. So if I keep expanding the metric set to be GDP plus dot dot dot dot dot, I can still do a weighted optimization with an AI on this and destroy life.

And the unknown unknown means there will always be stuff that matters that has to be pulled in where I don't want to run optimization on this thing.

(02:50:54):

So then the question comes, if I have something that can optimize so powerfully, what is the right thing to guide that? It's the thing that can identify the difference between the set of metrics you've identified as important and reality itself, the limits of your own models. That is not intelligence. That is wisdom. I'm defining these roughly in a way that someone gets the sense of. The difference between what the weighted metrics set of all the identified important metrics and the optimization function on it says you should do, the difference between that and what you should actually do is wisdom. And it requires being able to attune to more than just the known metrics and more than just the optimization and logic process on those. And so as super intelligence shows us how fucking dangerous narrow optimization is, it even shows us our own intelligence and our own intelligence running across all the humans via things like markets.

(02:51:56):

And the market incentivizes some humans to innovate in new ways to turn the world into dollars and other humans to take those innovations and exploit the fuck out of them. So you both search algorithms and optimization algorithms. The market makes the general intelligence of humans do those in groups. So the cybernetic intelligence of corporations and nation states in the world as a whole is already a general autonomous super intelligence running on all the humans as general intelligence as rather than running on CPUs, but also using all the CPUs and TPUs and GPUs in service of the collection of all the narrow goals. So AI accelerates the meta crisis, but it also makes clear to us that what it would take to align it is you cannot have, and this is why the question you asked, who's building it and who owns it, and what goals do those groups have. If you wanted to make a super intelligence that was aligned with the thriving of all life in perpetuity, the group that was building it would have to have the goal of the thriving of all life in perpetuity, which is not the interest of one nation state relative to others and is not the interest of near-term market dynamics or election dynamics or quarterly profits or a finite set of metrics.

Nate Hagens (02:53:12):

But that maps right onto the global geopolitical governance conversation, which we can't have it right now. Yeah, go on.

Daniel Schmachtenberger (02:53:24):

If you have a group that has a goal narrower than the thriving of all life in perpetuity and it is developing increasingly general AIs that will be in service of those narrower goals, they will kill the thriving of all life in perpetuity. So what this says is inside of capitalism and inside of separate nation state competitive interests which are inside of moloch superorganism type dynamics, you cannot safely build increasingly general intelligences. You have to get the general intelligence of the humans in the group, the cybernetic intelligence of these humans in the group be aligned with the thriving of all life in perpetuity that has those wisdom dynamics. That thing could possibly, possibly be oriented, at least it doesn't have a perverse incentive to build something that was also aligned with that where it is now seeking to scale. Because remember we were talking about it, wisdom is more possible at a smaller scale of people that can be in richer relationships with each other and then scale messes it up.

(02:54:33):

Can AI actually help take the dynamics that can happen at smaller scales and help us to build governance structures? And I don't mean AGI. I mean certain tools of computational capability. I don't think anyone thinks that if we were to try to build a democracy from scratch today, fighting a revolutionary war or whatever, that we would build it the same way we did in 1776. We would be building it in computational systems. Would people be able to vote from home? Would voting even be the thing? Would digital identity be a thing? Would cryptographic provenance of information be a thing? Would there be some kind of data aggregation using AI? Everybody knows if we're going to build it from scratch, it would be a different thing than when you built it in the industrial era. In the industrial era, it was a different thing than when the Greeks built it because they had different problem sets based on their tech and also different capabilities. If what a democracy and a market are systems of collective intelligence, what we need are systems of collective intelligence and wisdom and now we're talking about building artificial intelligence.

(02:55:34):

How do we build systems, cybernetic systems where the human interaction with each other, both how the humans are developed and how they're interacting with each

other, factoring all their incentives is both intelligent and wise and the computational capabilities, the artificial intelligence is not disintermediating them but is in service of the scaling of those collective intelligence capabilities?

Nate Hagens (02:55:58):

Don't we need a change in cultural goals and aspirations or prices or systems before that would happen? Or could we have small NGOs that get philanthropic donations, develop AIs in the service of all of life, but doesn't it need a lot of resources and compute and those would still get out-competed by the military and giant corporate AIs?

Daniel Schmachtenberger (02:56:26):

If we look at the multipolar traps and the competition dynamics, if we look at who has the resources to build things at scale, if we look at the speed of those curves, it doesn't look good. To be honest, it doesn't look good. Something has to happen that we are not currently obviously on course for, but if enough people, if some people can, stepping back, be able to see, "Oh, the path that we are pursuing that we feel obligated to pursue, oh our own opportunity focus relative to risk focus is actually mistaken and we're running a cognitive bias.", or other people who are employed by it, recognizing that or enough people to be like, "Fuck, let's make an agreement to slow this thing down and figure other things out." If you don't have, again we said wisdom will always be bound to restraint. If we do not get the restraint wisdom to stop the max race, then yes, these will be near the last chapters of humanity. And so then the task becomes how do we do that?

Nate Hagens (02:57:54):

So do we need as a first step is to expand the wisdom within the hyper agents in the system?

Daniel Schmachtenberger (02:58:05):

It's always a good question. If there are some people who have disproportionately more power than others and other people who have disproportionately more wisdom, do you try to get the people with the more wisdom to have more power and influence? Or do you try to get the people with more power and influence to have more wisdom? Or do you try to get the larger collective bodies to have some more of both of those?

Nate Hagens (02:58:26):

Yeah, and this gets back to the market, the superorganism, the growth imperative, the narrow boundary goal because those hyper agents that have the power and the resources to do AI and scaling, they have the optionality of money and money can be turned into everything else. Is it possible that AI gives people more optionality than money at some point? I mean I don't know how to break that dynamic.

Daniel Schmachtenberger (02:59:01):

There is stuff about money and optionality and pursuing instrumental goals, goals that increase your optionality to pursue other future goals that narrow goal optimization requires. In AI, it's called instrumental convergence. It no matter what your goal is, you're going to want to pursue certain things that increase goal achieving capability. Capitalism does that. We didn't finish answering, is the totality of the market already a super intelligence? We defined it as a superorganism. Is it also super intelligence? So there's some stuff that we have not got to. I will allow that-

Nate Hagens (02:59:37):

We could go for Balaji and Ferris's eight-hour record on one recording. No, I'm not going to do that.

Daniel Schmachtenberger (02:59:44):

What I'll say is that if where we leave this with all of the open threads, people have a lot of questions and we want to come back and address those, I think that would be interesting. In the need of closing, because I'm also realizing that I am late for another call with friends of ours on the AI topic, I want to share something that I think will be helpful in thinking about the wisdom intelligence relationship, which is not saying how we enact it. The enactment thing is a real tricky thing, but just on what we need to enact.

(03:00:19):

If people have not watched the conversations that David Bohm and Krishnamurti had together back in the day, I would recommend them as some of the most useful, valuable, beautiful recordings of human conversation I've ever seen. And in one short clip, David Bohm speaking on, if you YouTube it, I think it's called Fragmentation and Wholeness, something like that. He basically identifies, and this was maybe the '80s,

the cause of the meta crisis, though he didn't call it meta crisis or superorganism, but like all the problems of the human predicament that is clearly going towards a point of self-determination that was seeable at that time. The way he defined it I think was exceptionally good. I think it maps to the way indigenous wisdom has defined it in other things. Men are not the web of life. We are merely a strand and at whatever we do to the web, we do to ourselves.

(03:01:18):

But when we become power capable of doing exponentially powerful stuff, then our own short term, our win lose becomes omni lose lose. Our short term optimization ends up affecting even the timescales we care about. So what Bohm said is the underlying cause of the problem is a consciousness that perceives parts rather than that perceives wholes or the nature of wholeness. And because it perceives parts, it can think about something separate from others. So it can think about benefiting something separate from others and either it can then care about some parts more than others so it's okay harming the other things or it just doesn't even realize it is. So whether it is separation of care and values or separation of just calculation.

(03:02:09):

And so I can benefit myself at the expense of somebody else. I can benefit my in-group at the expense of an out-group. I can benefit my species at the expense of nature. I can benefit my current at the expense of my future. I can benefit these metrics at the expense of other metrics we don't know about. And that all of the problems come from that in so far as we were perceiving the field of wholeness itself and our goals were coming from there. And then our goal achieving was in service of goals that came from there. That's what wisdom binding intelligence would mean, which is the perception of and the identification with wholeness being that which guides our manipulation of parts, i.e. technology. And then Iain McGilchrist, who I think you are going to have on your show or maybe already have, if not-

Nate Hagens (03:03:04):

I haven't, please introduce us, but keep going.

Daniel Schmachtenberger (03:03:07):

Iain, in The Master and his Emissary, I think advanced what David Bohm was saying in an incredibly beautiful way, and he'll share it here on the show, but basically said to

not hit evolutionary cul-de-sacs, there's a capacity in humans that needs to be the master and another capacity that needs to be the emissary, meaning in service of, also bound by. You could also say there is a capacity that needs to be the principal and another that needs to be the agent, in legal terms, a principal-agent dynamics. And basically, the thing that needs to be the master is that which perceives, not mediated by word symbols, language models, perceives in an immediated way the field of inseparable wholeness. And the emissary is the thing that perceives each thing in light of its relevance to goals and figures out how to up-regulate some parts relative to others, i.e. What we think of as intelligence, relevance realization, salience realization, information compression. And so when I was talking-

Nate Hagens (03:04:23):

Wisdom is the master and intelligence is the emissary?

Daniel Schmachtenberger (03:04:28):

And I said you're basically saying that the principle that the emissary developed all these powerful capabilities and so in some places the emissary said, "Fuck the master thing. I want to be the master." And it had the tools to do so, started to win that thing on a runaway dynamic is the cause of the meta crisis, the superorganism. He's like, "Exactly." So which maps to what Bohm was saying, it corresponds to the reality that the master circuits would orient to towards. So if you already look at all the problems in the world and the global meta crisis and the impending catastrophes being the result of the emissary intelligence function unbound by the master wisdom function, then you look at AI being taking that part of us already not bound by wisdom and putting it on a completely unbound, recursive exponential curve. That's the way to think about what that is. So what is it that could bind the power of AI adequately has to be that what human intelligence is already doing is bound by and in service to wisdom, which means a restructuring of our institutions, our political economies, our civilizational structure, such that the goals that arise from wisdom are what the goal achievement is oriented towards. That is the next phase of human history if there is to be a next phase of human history.

Nate Hagens (03:05:55):

So my takeaway from that and from this whole conversation is that well, first of all, my takeaway from the whole conversation is I owe you an apology from the time we met

four years ago. I thought limits to growth and the great simplification dominated AI as a risk. And looking back I didn't understand, actually I didn't understand until the last three hours how AI merges with the superorganism in this potentially catastrophic way that accelerates all the things. So I didn't understand.

Daniel Schmachtenberger (03:06:37):

As a hypertrophication of the intelligence that is already driving the superorganism, exactly.

Nate Hagens (03:06:41):

Yeah. And then my second thought is there still is a role for education, for culture and leading by example and by a transference of the me-based culture to a we-based recognition both between humans, we are in this together and we are part of the web of life and that's also part of the we. And the more humans that understand and feel that, whether through ayahuasca or drums or being in nature or being with others or meditation or whatever it is, the higher chance we have to intervene with the trajectory we're on. I mean that's my just mild takeaways.

Daniel Schmachtenberger (03:07:38):

But those people who go spend time in nature observing non analytically but communing with the truth, goodness and beauty, the intelligence, the meaningfulness of it, who do the ayahuasca and have those experiences, if they become lotus eaters and simply drop out, it doesn't affect the curve of the world either.

Nate Hagens (03:08:04):

It's the equivalent of them being deniers, yeah.

Daniel Schmachtenberger (03:08:08):

The key is, and yet if people are working to make change but they are not actually connected to the kind of wholeness that they need to be in service to and they continue to have what seem like good goals, but they're narrow, we need to get carbon down. We need to get the rights of these people up. We need to protect democracy. We need to get our side elected because the other side is crazy. We need to develop the AI to solve this problem. Anything less than the connectedness with wholeness and anything less, both at the level of care and at the level of calculus,

even though you can't do them, you are oriented to try with the humility that knows you'll never do it properly. The humility that knows that you'll never do it properly is what keeps you from being dangerous from hubris. But the part that really, really wants to try is what has you make progress in the direction of the service of the whole.

Nate Hagens (03:08:59):

We should probably do our next conversation in the service of the whole and take a deep dive in that and what it might look like and maybe there will be some people that respond to that calling and then come back and do a deeper dive on these AI questions because it's going to take me a while to process this. This has been great. Is there anything else you want to contribute?

Daniel Schmachtenberger (03:09:27):

We'll add some things to the show notes. So the piece that Eliezer Yudkowski did on Bankless I think is worth watching, both because of getting what he said, but also that the moment in culture. The Bohm-Krishnamurti pieces I think are super valuable. McGilchrist and Tyson, you're going to have on, I think Samantha would be great to have on. She and I've had good conversations on these topics and has good indigenous kind of insights. There's a guy named Robert Wright, who's made a bunch of short, really simple AI risk videos that I think are exceptional as a resource to share.

Nate Hagens (03:10:06):

Is he the guy that wrote The Moral Animal? No, different one.

Daniel Schmachtenberger (03:10:11):

No, different guy. And let me verify, Robert Miles, excuse me, Robert Miles. And so if people just want to understand the AI issue better with short, simple explainer videos, I think his are some of the better ones I know of. And obviously, the conversations you and I had previously in the series contextualized a lot of it. And then if people write to you with questions, there are so many threads we left open. Would be happy to come back to it. It would be fun.

Nate Hagens (03:10:50):

Okay. Thank you for continuing to think and push on our plight. It seems daunting, but this is the time we're alive. We have to understand it, care about it, and engage with it. But this is a big old red pill, my friend.

Daniel Schmachtenberger (03:11:12):

I mean it only takes the whole of ourself in service to the whole of reality and utilizing both all of our technological and trans-technological capabilities for the purposes that are inclusive of everybody. And so that could seem daunting, but can also seem inspiring.

Nate Hagens (03:11:45):

Yeah, it's both. It's both. To be continued and thank you.

Daniel Schmachtenberger (03:11:53):

Thank you, my friend. These are fun to be in these conversations.

Nate Hagens (03:11:57):

If you enjoyed or learned from this episode of The Great Simplification, please subscribe to us on your favorite podcast platform and visit thegreatsimplification.com for more information on future releases.