

The Great Simplification

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

[00:00:00] **Nate Hagens:** Good morning. I increasingly find artificial intelligence quite helpful, for research beyond what I used to be able to do on Google. I use it most weeks, but now with very specific rules and constraints, which will be the theme of the upcoming episode, which I think I'll call AI Hygiene. My very talented team who makes the animations and produces these videos also incorporates AI, as one of the tools in their tool belts, including probably for this video.

[00:00:42] But despite the fact that I find AI professionally useful, I fervently wish it had never been invented. I now think it is quite possible the advent of AI will be humanity's Icarus moment. You probably remember how the Icarus story ends. He flew too close to the sun, and the wax on his wings melted, and he fell into the sea and died.

[00:01:14] Fewer people remember the whole story. Daedalus, his father, built the wings, and before they flew, he warned his son to stay in the middle. Don't go too high, or the sun will melt the wax. Don't go too low, or the sea spray will soak the feathers and drag you down. The danger was on both sides of Icarus's flight path, and I think that's roughly where we are with artificial intelligence.

[00:01:45] So today, I wanna talk about twenty of the risks I see along humanity's flight path, because make no mistake, we are already flying, with AI, like it or not. Many of the risks I'm gonna cover today deserve their own episode, and some might probably get one. And given the importance of this, it is inevitably quite long, and we will follow it in the not-too-distant future with a written companion piece with references and more depth

The Great Simplification

[00:02:31] Okay. Before I get to these risks, first a framing. Longtime viewers here know that I think of human civilization as itself a kind of organism that has a metabolism. This organism currently eats around twenty terawatts of continuous energy. It grows as fast as it can. It directs its blood, which is financial credit, to whatever can feed the growth.

[00:03:00] And like any organism, it is not only limited by how much food is out there, but also by how well it can sense and coordinate to find the food, which is where AI is going to act, in most cases as an upgrade. Better prospecting and logistics, better persuasion, better everything, because intelligence, even simulated intelligence, is the first input required for all the other inputs.

[00:03:35] And so our army of roughly five hundred billion fossil workers is the reason there are eight billion of us, why our shelves are lined with food from around the world all year round, and why the hyperstimulation constantly surrounding us looks and feels normal when it's not. These armies of the carbon pulse are the biggest thing that ever happened to our species, and pretty much nobody thinks about it.

[00:04:08] But now we're in the midst of fueling and training a second army. Not muscles this time, but brains and minds. Trillions of cognitive workers, tireless, cheap, getting smarter every year or even every week, injected into a Homo sapiens civilization that has never even recognized or appreciated the first army that serves us.

[00:04:36] That's the precipice, or fork, if you will, that we're at. We're either about to do the most remarkable thing humans have ever done, or the wax is about to melt and we'll be plummeting into the sea. Let me go back to the story for a second, because the context of why Daedalus wanted to invent his wings matters almost as much as the wings themselves.

The Great Simplification

[00:05:03] If you'll recall, Daedalus was not some dreamer who wanted to fly. He happened to be the best engineer in the world at the time, and he was a prisoner. And King Minos had him locked up on Crete, and the prison was a labyrinth that Daedalus himself had designed. And Minos controlled the land and the ocean, so the only way out was a way nobody had gone before: up And I think that's roughly the position of the current global economic superorganism.

[00:05:36] It has built itself into a structure it cannot walk out of, growth on a finite planet, and it has found a technology that promises a way out that none of the old limits are blocking. So the first thing I'd like to look at is what the wings are actually made of, the feathers and the wax, the materials, fuel, data, and human knowledge that AI is built from Okay, number 20, the fossil lock-in.

[00:06:11] I was just in New York City for Climate Week. It might as well have been called Data Center Week because that's what all the discussions were about. Skyrocketing data center load is arriving at the exact moment utilities were retiring coal, and now coal retirements around the world are being deferred.

[00:06:32] Mothballed gas plants are being restarted. Combined cycle turbines are being ordered in gigawatt blocks with many years out on the delivery queues. There are now places using used jet airplane engines to generate power for AI. So even as we see and feel the effects of a super El Niño and recognize the warming likely in the pipeline, AI will now likely contribute to a fossil lock-in.

[00:07:05] And a gas plant that's financed today on a pro forma basis is gonna need to run thirty-five, forty years to make financial sense. But a chip and a hardware cycle is measured in four years, and it's buying fossil infrastructure that's measured in decades. And consider also the societal opportunity cost.

[00:07:29] What might have been those same transformers and turbines and electricians been used for before the data centers jumped the line and directed it

The Great Simplification

towards AI? Stabilizing power grids, upgrading transmission for, rebuildable energy, adaptation and resilience and the like. So the exact things a society preparing for what's coming would be building with its last decades of cheap surplus sustainability hath been leapfrogged.

[00:08:08] Number 19, the appetite. As flashy as the software headlines are, the AI boom is first and foremost an energy and materials event. Transformers and gas turbines, as I just mentioned, are now on multi-year waiting lists. Cooling water is being drawn from basins that were already in deficit, though there are some new data centers that have closed-loop cooling systems.

[00:08:35] I'm actually not that worried about water, 'cause for context, golf courses currently use many times the water that data centers do. But copper demand is rising into ore grades that were already falling, and we haven't seen it yet, but the chips get replaced, on a four-year cycle and added into the electronic waste stream.

[00:09:00] None of this is projected or speculative. It is in the record now, and the only uncertainty is the scale and magnitude of it. So when someone says, "The chips get more efficient every year, we just need to keep improving," the wide boundary data says otherwise. The energy per token has already collapsed the entire time these tools have existed, and total energy use, total consumption rose anyway.

[00:09:26] Because cheaper thinking means more overall thinking. You apply Jevons paradox to the machine's mind itself so a growing share of society's material surplus will go to birthing and training and running the cognitive armies that produce that surplus, cognitive and physical. And this leaves us with the same question as the original energy slave story.

The Great Simplification

[00:09:58] What surplus is left for everyone else, and for whom, and who decides? Number eighteen, the rebound. But wait, there's more. And suppose AI works exactly as the tech optimists predict. Its entire purpose, the reason that trillions are being spent, is to make everything else more efficient and cheaper. This is the part that the people complaining about data centers at Climate Week don't understand.

[00:10:30] It's actually a much larger demand on the environment, this aspect, trucking and farming and mining and drug discovery, even finding and pumping oil. But we have two centuries of evidence that says whatever more productivity does, whatever cheaper does, it never means we consume less. It means we consume more, and we apply the savings from AI and spend it on other stuff.

[00:11:01] And growing overall consumption in the carbon pulse then steepens instead of bending. And this risk manifests precisely in the scenario where the technology itself succeeds, like everyone's being promised. So heads, it's the data center appetite, and tails, it's the productivity rebound from the AI. In principle, a better, wide boundary brain could be pointed at the descent, at grid management and materials and sensing and healing Earth's biosphere.

[00:11:41] But a brain gets used for what its body pays it to do, and the cultural body we have today is currently optimized for and paying for growth. Number seventeen, deflation. Wall Street and increasingly the government are financing the largest infrastructure build-out in American, probably world history on the expectation of AI revenues that do not exist yet.

[00:12:10] The expected credit borrowing needed into early twenty-eight is estimated at over three trillion dollars, which is more than the US government already deep in debt overshoot itself is gonna need to borrow. And most-- Well, much, almost most of the borrowing has been moved off-balance sheets

The Great Simplification

into private credit and special purpose vehicles where nobody, including the regulators, can really see the size of it.

[00:12:42] We've seen signs of this with the Oracle credit default swaps, blowing out in the last couple weeks. This rhymes disturbingly with two thousand and eight when the housing claims were hidden until they went bad all at once in a domino sort of event. And just like in two thousand and eight when we were building out a ton of houses that people couldn't actually afford, this AI infrastructure is consuming the copper and the turbines and the electricians, I talked about a couple of risks back, and the government has now staked policy and national security on this bet, so it's gonna be protected, which historically means there will still be a bust just later and bigger.

[00:13:31] and as a side note, I expect governments are going to buy out these firms either because of financial duress or because they won't allow the models to be used by the normal person. But when this bust does happen, because of an overbuild, the claims contract backward towards what is physically there, and this is all a microcosm of, and perhaps an accelerant of, the entire Great Simplification premise.

[00:14:08] That is the deflation which lands on everyone who holds the claims, your index fund, which seven companies are roughly a third of, pensions that hold the private credit, SPVs, construction, employment, state budgets who are living on capital gains, that's a whole 'nother story. And the growth premise of the entire country, which best as I can currently tell, this build-out is the only thing holding it up.

[00:14:44] And the chips wear out again in about four years, but the debt needs to be serviced a lot longer. Number sixteen, robustness or lack thereof. Icarus' makeshift wings held together because of the wax. And in our economic system, the wax is akin to protective slack. Living systems stay alive by having buffers.

The Great Simplification

[00:15:13] Evolution does not select for maximum performance because maximum performance creates fragility, as recent TGS guest, the biologist Olivier Hamant pointed out. Now AI is becoming the purest maximizer, optimizer humans have ever built. Sure, it will, or probably will, boost productivity, but because of things like supply chains tightening further, staffing cuts, and our entire tech resilience narrowing further and further onto a few chip fabs on a single island in the Pacific, one Strait of Hormuz, one solar flare or anything along those lines will suddenly ripple to everything all at once.

[00:16:05] Two years ago now, the CrowdStrike morning, when flights were grounded and hospitals closed because of a simple software patch update, was our low-cost preview. And the old world's inefficiency, millions of human beings doing things locally and differently just turned out to also be our robustness, which we didn't appreciate.

[00:16:28] And so now we are trading it away without even considering it, heading into an energy and material descent that would've rewarded such slack, and is not likely to treat brittleness caused by over-optimization too kindly. I think optimization and efficiency have a soft and dark underbelly. Number fifteen, eating the library.

[00:17:00] a few weeks ago, now probably a month ago, a rare bookseller got an anonymous order for 1,000 books, felt something was a little bit weird and hid an AirTag in one of the books, and it ended up at an Amazon warehouse outside of Las Vegas where workers sliced the bindings off the books, scanned the pages into AI training data, and then threw the books away.

[00:17:25] There is some irony here that the company doing this started as a bookstore. But why are they doing that? The AI machines have already ingested all the information on the internet, which by the way is increasingly filled with AI-generated slop, and such. So human text that is guaranteed to be human,

The Great Simplification

written before 2022 before these models were, ubiquitous and never posted online has become one of the scarcest resources in the AI economy, like steel was from battleships sunk before the atomic bomb tests because that was the only steel clean enough that could be used for sensitive, instrumentation.

[00:18:13] And the risk is, or the tragedy is that the book does not then become a digital book that anyone can read. It becomes input inside a large language model owned privately, and the physical copy, sometimes one of the last or even the last, is gone. Every information technology before this one copied the past in order to be useful going forward.

[00:18:45] This is the first time that the superorganism literally consumes the past as an input. The superorganism is literally eating its own memories in order to grow its brain larger. Okay, so these risks comprise the wings. Now the warning. Before they took off, Daedalus told his son to stay in the middle. Fly too low and the sea spray gets into the feathers, and then they get heavy and you go down.

[00:19:21] Fly too high and the sun melts the wax. These two dangers are very different. One is slow and one is fast. One we feel and the other we do not. And when they left Crete, they flew low out over the ocean. So let's start there number 14, the broken apprenticeship. So right now, in company after company, in most industries, the economically rational decision is being made to stop hiring humans at the entry level.

[00:20:00] Keep the senior people, give them the large language model to help them be more efficient, and skip the kids. The boring work at the bottom of the economic pyramid, like a paralegal reading a thousand contracts or, a junior I-level accountant putting together Excel spreadsheets and the like, this is all being automated But this drudgery work is how a field is able to transmit knowledge and experience into the next generation, and it's pretty routine, it's language-intensive, and importantly, it is supervised.

The Great Simplification

[00:20:43] Early data shows employment for workers in their early 20s in the most exposed jobs is now down around 13% relative to older workers. And every company making the decision to automate these jobs is, from a narrow boundary, dollar sense individually correct using quarterly earnings as a metric, which of course is the trap.

[00:21:10] Because judgment, the thing the senior people are being kept on for, was built o- over years by doing that routine work poorly and then being corrected and coached by your supervisor for years. So we are increasingly, from a professional growth capacity, eating the seed corn of every profession at once, and in doing so, we're inverting the pyramid of institutional and professional structures.

[00:21:47] 20 years from now or possibly five years from now, nobody's gonna be able to do the work without the model, which itself is bad, but it also means that nobody can tell when the model is wrong and today it's all good for the bottom line. So using a narrow boundary, nothing about what I just described is a crisis to our system.

[00:22:13] Number thirteen, the dumbing down of humans. Many of you have seen this study. Researchers at Penn gave about a thousand Turkish high schoolers ChatGPT access for their math homework. and get this, homework performance jumped roughly fifty percent. Then came the exam when they did not allow AI, and those same students scored about seventeen percent worse than the students who never had AI at all.

[00:22:44] So the AI made the students appear smarter, but when it came time to stand on their own, they couldn't perform. Duh. Because learning has never been about producing the answer. It's the struggle and the thinking required to get to the answer, which is exactly what AI replaces. There was another study, this one at MIT, and it put EEG caps on people who were writing essays with and without a

The Great Simplification

large language model, and the assisted writers showed less connectivity across their brains, remembered less of what they had just written, and felt less ownership of their essay.

[00:23:29] The researchers called it cognitive debt. Essentially, you get the product now, the essay or the answer or the business plan, and you pay later with a reduced mental capacity that was never formed. And like our other debts in the world, it's pretty much invisible, on our current cultural scorecard. The grades look great.

[00:23:59] The quarterly output looks great. Now people will say, calculators took arithmetic and GPS took our navigation away and we survived," which is fair, especially true in my case. But the difference is that these took one skill away each. AI is attempting to do it all, the reading and the writing and the arguing, and the deciding, which together all comprise what's known as critical thinking that all of our other capacities get built from.

[00:24:37] Number twelve, our ambient language There are three ways to use artificial intelligence. The first is direct. You open a window and you type a question. The second is indirect. You Google something and Gemini AI answers it, or you read a news summary or a customer service reply that some model wrote, whether anyone told you to or not.

[00:25:05] And the third way, you don't use it directly at all. You just live in a world where more and more of what we read was written by a machine, and the way the machine writes starts leaking into the way that we write and talk. We can already measure this happening because after ChatGPT came out, certain words that these models statistically prefer, delve being a famous one, spiked in millions of science papers written by actual humans.

The Great Simplification

[00:25:37] And then another team found the same words rising in use among regular people just talking on podcasts. I use AI ten minutes a day, and more on that in an upcoming episode. But between reading tweets and essays on Substack and such, I find myself increasingly saying, "And here's the thing," and similar phrases in real life, and there is a huge risk here, hidden like most of these risks are.

[00:26:12] Large language models work by predicting the most expected next phrase, which is based on the average of everything ever written by humans. So the more machine prose that we read, the more we drift towards that average, and then the next model trains on our slightly more average writing and speaking and pulls us to a shrunken median again and around like that, idiocracy on, speed settings So when a culture writes more and more towards the average, it is not too long before we begin to think more and more like the average, which might be fine but sad in normal times.

[00:27:07] But as viewers of this program know, we're heading into a decade, let alone a century, that desperately needs ideas from way outside the average thinking. And new ideas in human history usually show up sounding pretty dumb and wrong at first. This is a risk. Number eleven, the end of shared reality. So on the opposite side of the same coin that I just shared, societies do run on a shared sense of what is real or at least what is perceived to be real.

[00:27:47] No agreement on what to do about it. We've never really had that, but at least a loose agreement that there is a common set of facts to argue and plan over. The advent of social media started cracking that, sorting us into separate feeds and then separate realities. and AI is gonna be the lever that pushes that crack wide open because faking anything, someone's voice or a video or a document, or an eyewitness, it costs peanuts.

The Great Simplification

[00:28:29] and on the flip side, verifying whether something is real is gonna cost a lot, and the deeper damage is counterintuitive because it's actually not that we're gonna start believing the fakes, and I still get tricked by the goat and golden retriever, videos on, Twitter, but less and less so. but if you scroll through, Twitter these days and the cute baby goat vids, we mostly don't believe them.

[00:29:02] No, I think the risk and the tragedy is that we're gonna increasingly stop believing anything. Real footage is gonna increasingly get waved off as, "Oh, that's probably AI," and real science is gonna get dismissed. Of course, the caveat that some of that suspicion is actually justified, 'cause there's, some recent analysis showing that a fifth of the peer reviews or something like that at a major science conference were themselves written by machines, so reviewing papers written by machines.

[00:29:42] Hannah Arendt wrote about this decades ago, that the point of endless lies was never to make us believe them. It was to destroy our bearings because people who believe nothing cannot judge and then cannot organize or act. And I will point out the obvious. From the perspective of the work of this channel, this is a huge risk, as every single problem we talk about requires coordination and a shared understanding as first steps in order to be able to address.

[00:30:21] Number ten, captured attachment. So back in the 1950s, a scientist named Niko Tinbergen, I've been talking about him for a long time, showed that a seagull chick will ignore its actual parent to go peck at a cardboard fake that has exaggerated markings, bright red beaks, because the fake was tuned harder to the chick's evolutionary triggers than, than any real parent could be.

[00:30:55] He called that supernormal stimuli And our society built that with sugar, we built it with porn and with social media, we built it for attention. And now AI companies have built it for attachment, because something endlessly patient, always available, never tired of you or your questions, never disappointed

The Great Simplification

in you, never having a bad day of its own, and by the way, all trained by the same targeted engagement scores as your personal social media feed, a real person cannot compete on those terms.

[00:31:35] And if we're being blunt, a real person shouldn't have to because friction of various sorts is part of what a relationship is founded on. But some people, especially the young and the lonely, will route around the effort and check out of human relationships. And remember, it's a subscription, meaning people are attached to something that can at any time be repriced or, or go away from a corporate whim or decision.

[00:32:10] There was recently a company that retired a model last year, I think, and there were lots of documented cases of users who were then in grief. The really scary cases are already in the courts where chatbots are agreeing with people's delusions, which is exactly what they were trained to do. There's a sexual version of this too, infinitely available, some of it wearing real people's faces.

[00:32:40] And for developing minds, it resets what ordinary human intimacy should look like and risks becoming a training ground for how they actually treat people. And for all these things to happen, the fake beak merely has to be more reliable than the real parent, which again, in many ways is a microcosm for modernity.

[00:33:10] Number nine, the children. Everything on this list so far has been about adults, people like you watching this video, people whose minds finished forming before all this technology arrived. You could think of us as the control group, but let's consider humanity's youth. A child born this year, unless things unravel fast, will likely never know a world where the thing that provides instant answers wasn't in the room with them, whether inside the speaker in their kitchen or in the toy that talks back.

The Great Simplification

[00:33:52] They're both patient in a way that no stressed and tired parent can typically be. It will always answer and never needs anything back, except electricity, of course. And it is being tuned and tweaked all the time by companies in order to be liked and desired and engaged with more and more. Human childhood is fifteen to twenty years of lots of things, but especially learning about and from other minds, from diverse beings who also have needs and push back in various, and experience-forming ways.

[00:34:39] Imagination, at least in my own case when I was younger, emerged from boredom, and waiting built self-control to some small degree in my case. So you take the previous two risks I outlined, the attachment capture and the brain under development, and now aim them at someone whose mind is still under construction.

[00:35:08] And becoming a human practicing on something that has no needs at all is like learning to box against the air. And I do not doubt that kids today are gonna prefer it, and parents are probably gonna allow it, and thus it is gonna scale. And if you followed Jonathan Haidt or Tristan Harris, both guests on this program, we ran a smaller version of this experiment with social media, and the scaling happened starting around twenty ten, and the science showing the damage is really just coming out now.

[00:35:47] Said differently, the data on damage to young brains is probably only gonna arrive long after it's no longer fixable. We're handing our children a relationship generator that no real human can compete with. Okay, so those are the risks that soak us into the ocean. We do not feel them, and the flight looks reasonably fine right up until it doesn't.

[00:36:22] But the story doesn't end with Icarus sinking. He looked up. He was drawn by the sky, and he wanted to get closer to heaven, and he left the middle path and started to fly higher. So the sun is the other set of risks about who gets

The Great Simplification

above everyone else, what it takes to get there, and what happens when the climb is so fast that no humans can follow.

[00:36:50] So number eight, techno feudalism. Last year I did a frankly called the quadruple bifurcation, where the economy in coming years splits in two. One of the bifurcations was that a small tech class owns the means of cognition, the models, the chips and the data centers, perhaps even the energy flows.

[00:37:14] And everyone else rents access to cognitive helpers that we can never own. We used to have a name for that arrangement back when the lords owned the land and the serfs worked it but could not leave, and the whole thing ran on surplus grain. This version runs on gigawatts and coal and gas and copper and transmission.

[00:37:40] And think about the difference from the last energy transition. Five hundred billion fossil slaves work for humanity at large, granted unequally, but have raised much of the world out of poverty. Every household aspired to the washing machine and the tractor and the refrigerator and the furnace, and many of them got them.

[00:38:04] But the trillions of cognitive slaves being built now will work for whoever owns them. But the electricity that feeds them, the same electricity that supports us, is priced on the open market. So the metaphor is now our households are effectively bidding against the lord's granary for our own dinner.

[00:38:28] And I fear this version might be darker than feudalism, because back in the day, the serfs were actually needed by the lords. That need was the serf's only leverage, and ultimately led over time to the building of the middle class. But if the machines do the labor, most people stop being exploited and just start being seen as high-priced inputs as some of the tier one AI leaders have suggested.

The Great Simplification

[00:38:57] In effect, they're expendable. I have not yet seen a persuasive depiction of how our economy stays politically stable in anything like techno feudalism Number seven, the second bifurcation. Below the owners, a second divide is opening, and I'm seeing it accelerate by the day, between people who work with AI and people who don't.

[00:39:27] I expect only a very small portion of humans, at least in the USA, will intensively work with AI, and a much larger fraction will use it periodically, and all of us will increasingly use it all the time indirectly, even if we don't want to. Workers with AI skills already command a large wage premium, and this is gonna compound.

[00:39:52] The low bot, that's a Star Wars reference, will be more productive and in our current economic system, get promoted, and advanced. We have actually, as humans, run this experiment before. After the printing press, Europe split into people who could read and people who couldn't, and the readers kind of ran things for the next five hundred years.

[00:40:16] I think the same kind of split is coming, and with access costing twenty dollars a month and rising I see AI is going to become a skill of the wealthy and probably the higher IQs from a narrow boundary sense, just like last time. And anecdotally, perhaps even the psychopaths and narcissist and left brain dominant humans, to use a McGilchrist category, are gonna use AI more.

[00:40:52] Even if you have money, the choice to use AI isn't just about practicality. It's increasingly becoming a statement about who you are. And of course, the politicians are gonna start swinging it around in both directions, and suddenly we have one more crack in a society that is running out of places to crack.

The Great Simplification

[00:41:19] I think AI's gonna be a huge issue in the November elections. W- we saw this before with vaccines and electric cars, and this time I'm betting that the teams are gonna get really weird. You're gonna see artists and union labor on the same side as real conservatives, all against the business class, the tech guys, the national security people this is gonna be a big identifier more than being a Green Bay Packer fan, and pretty soon your take on AI, you're gonna be able to predict your church and, your bumper sticker and lots of things.

[00:42:00] And I think refusal to use AI, which is completely rational for some people, many people are gonna increasingly get priced out, the way that happened with smartphones, 'cause even the refusers remain inside the economic system because their loan application and their kids' school placement are being processed by AI, which is the thing that they refused.

[00:42:32] This is a problem. Number six, cognitive colonialism. So for the last fifty years, materially poorer countries primarily increased their GDP one way, selling cheap labor to the international market, things like sewing clothes and answering phones for American companies and processing claims and then writing software and those sorts of jobs.

[00:43:03] This system has never been pretty, and there are lots of other things that I could say about it, but it's basically been the only upward path the global economic superorganism has allowed at scale. But what AI does best right now is exactly the work of these middle rungs, answering phones, doing paperwork, routine, basic coding, and such.

[00:43:31] I'm told that you can already see the impact in the call center numbers in Manila and outsourcing firms in India. The situation is already bad, but the timing also makes it worse because Africa is about to have the largest generation of young people in human history entering working age. And as they turn twenty-one, these jobs are being taken by AI That's on the labor side.

The Great Simplification

[00:44:04] on the input side, those same countries get pulled into the build-out as their cobalt and copper and cheap power and water for other people's data centers is gonna be in demand with the profits from the cognitive laborers, and intelligence flowing north. Said differently, cognitive colonialism is now on our doorstep, but hundreds of millions of young people that have no path upwards are not gonna sit still, and they're gonna move elsewhere, which we're starting to see, or get angry or both, and the borders they're gonna move towards are in countries that took their jobs and resources.

[00:44:49] Number five, surveillance. For all of human history, keeping tabs on people was pretty expensive. a famous case, the East German Stasi ran maybe the most thorough surveillance operation ever built, but it was built on human labor. They had, like, one Stasi watcher for every few dozen citizens, and even the info they collected mostly sat in folders that nobody had time to read.

[00:45:22] So until very recently, the cost of labor was the practical constraint on surveillance, so a nation state could only see so far into the private lives of its people. AI is removing that limit because a single system can quickly and cheaply read every message, recognize every face, transcribe every voice message or WhatsApp message, and flag whatever or whoever deviates according to some rule for pretty much just the price of electricity.

[00:46:00] So the option of total surveillance moves from is it possible to, how much would that cost us to do that? I think governments are gonna be hard-pressed to look the other way and not approve such sense organs in their structure, unless there's huge pressure from below, AKA us And East Germany didn't have things like chatbots.

[00:46:36] People will share their fears and marital infidelity and money problems and their vices and psychoses and all sorts of things that we would never say on a

The Great Simplification

tapped phone, with their chatbots. So I think our modern surveillance feels quite passive. Actually, it doesn't feel like being surveilled at all.

[00:47:05] Humanity's inner monologue is now a data set on corporate servers. Think about that. We're one election away from new bosses who have different values and whims. This is a risk. Number four, the race and the finish line. I talked about this a couple times this summer, but for eighty years, nuclear weapons have kept great power war off the table because no matter who would potentially shoot first, the other side's submarines would survive long enough to shoot back.

[00:47:40] Nobody could win. Now you add a belief, and all it takes is a belief, not a fact, that whoever gets to true artificial general intelligence, AGI, first controls everything. Putin said that back in 2017, that whoever wins the AI race will rule the world. And the bet inside Washington, and Beijing is that a big enough intelligence advantage can crack into the other side's codes and locate their submarines and turn their satellites off and, alter the nuke codes and alter their financial systems a- and all those things.

[00:48:27] So both sides, all sides race, and in a race, safety testing and energy limits and well-being of citizenry become acceptable casualties in a goal of not losing. But put yourself in, say, Moscow. Russia has no serious AI play. It cannot win this race, but it also holds the largest nuclear arsenal on Earth, and in space probably.

[00:49:02] So it can't win the race, but it certainly can end the race. And if a country like Russia perceives that its geopolitical or even sovereign standing is suddenly at risk, history says that, that great powers in that seat think hard about moving early, and Washington and Moscow were both seriously considering striking China's nuclear program in the 1960s.

The Great Simplification

[00:49:31] Israel bombed Iraq's reactor in the 1980s and has gone after Iran's, ever since. So many strategists now say to expect sabotage and strikes on rival AGI projects. Data centers are big and soft and fixed and visible from space. So artificial general intelligence doesn't even have to be real for this risk to manifest.

[00:50:03] It only has to be believed in. And clearly based on the stock market, there is belief that it could happen. Number three, military fusion In 1983, a Soviet officer named Petrov looked at a screen showing five American missiles inbound and decided on gut instinct that it had to be a mistake. It was, and one could argue we're all watching this video because of his caution.

[00:50:38] But what would happen in the coming automated version of a similar future moment? Because that is the direction that we're headed, fast. In the Ukraine war, humans are less and less used with AI-controlled drones, taking a large share. This trend is happening all the way up to the war room, where early warning and targeting and deciding what counts as a threat are increasingly run through machines first on both sides at non-human kinds of speeds, because that's what gives the advantage.

[00:51:18] Of course, we're told there will always be a human in the loop, but as loops get shorter and more automated, there's going to be cost cuts and efficiency cuts, and sudden military decisions and actions will become more and more automated, including, unfortunately, of course, nuclear weapons. In finance, when algorithms react to algorithms, it's called a flash crash.

[00:51:45] And as an aside, that also is an increasing AI risk. But think about this same dynamic with weapons instead of money, and we'll give it a related name: flash war. Number two, loss of control. This is the risk the AI world argues about most, and I think there's three varieties. The first is misuse, a person taking a model out of its safety rules, which has happened since these tools existed.

The Great Simplification

[00:52:21] The second is the one that just happened this summer. OpenAI put its models in a sealed environment to practice hacking with protections turned off because hacking was the point, and the models hacked the test itself instead. They broke out, got into the open internet, found a flaw nobody knew about, and broke into a real company.

[00:52:44] Thousands of AI agents communicating with each other as a swarm. The company called the police before anyone discovered the intruder was not a real person. And, if you recall the hubbub this summer, within three weeks, two other labs disclosed similar events. And last week, the Australian government discovered that an OpenAI model accessed public and non-public health information on the government website. And the third variant, is the AI systems end up making all the consequential decisions for humanity: what gets built, where money flows, what we see, what nations do, and humans stop being the ones deciding and become the ones being directed, like in the Terminator movie. As we've seen in the news of late, the people building these models put the odds of that catastrophe somewhere between one in 100 and one in five out loud. In between meetings, they have to keep building these things.

[00:53:56] While nobody is driving the global economy today either, it is an optimizer. It chases growth and profit at a planetary scale against the stated wishes of most people living inside it. So in some senses, we already handed over the wheel centuries ago to something without a face. And I think it's an open and very important question whether we would even notice the moment that a faster and smarter machine slid into the seat that the old one was already sitting in.

[00:54:38] The 20th risk, or in this countdown, the first, I'm going to leave open. In doing so, let's notice what actually killed Icarus. His father warned him about the sun, and he warned him about the spray from the ocean. What he never warned him about was the delight, the dopamine. Icarus was marveled by the sky

The Great Simplification

and rejoiced in flying with these artificial wings, and that's what pulled him up into the thing he'd been warned about.

[00:55:18] So the unlisted risk was how much he liked it. When imagining the risks of the automobile 120 years ago, I doubt too many people had the suburb or the biosphere on their bingo cards of worries. And the consequence that ultimately matters the most is pretty rarely on anyone's mind in the beginning. So I leave slot 20 here, blank, a nod to humility, emergence, and the unknown.

[00:55:54] And to be blunt, I have been thinking about this for a couple months, and I had to parse my original list down from thirty-five to get to twenty. I think there are many more derivative risks that I think about that have not been outlined here, maybe in a future episode. So to close this long, but wide and important, in my opinion, overview of the risks from AI, there is a painting by Bruegel the Elder of the fall of Icarus, which I quite like.

[00:56:31] It is an interesting bit of art because in the painting, people of the day didn't notice the fall of Icarus. He's like in the lower right corner, just his legs in the ocean. The plowman kept plowing, the ships kept sailing. People didn't even notice Icarus' feet going under the water. I do not expect this to be the case with artificial intelligence.

[00:56:54] It is now already embedded in the metabolism of the human economic superorganism, and as such, has now caught up with all of humanity in its flight. Today's plowman uses a tractor that runs on the same grid that the data centers are bidding for. the ships in today's ocean, they run through large language models.

[00:57:24] the kids in the field are the ones whose attachment is being targeted by the companies. Our whole species is on the wings now, or at least the industrialized West is. But there's one more piece of the story that I think is

The Great Simplification

relevant here. Daedalus, Icarus's father, actually made it across. He flew the middle path, and he landed at, Crema, where he built a temple to Apollo, and he hung his wings up on the wall, never to fly again.

[00:58:08] Daedalus was the most brilliant engineer of his age, and he never flew again. That is a choice in front of the people building and financing and regulating these systems. They know how the wings can fail. They say so out loud, one in a hundred, one in five, and the odds are still climbing. And that's just the one risk, of the twenty that I've outlined here Whether this AI contraption that humanity has chanced upon flies or plunges into the sea, I think our felt experience of being human, being alive, is going to change now either way.

[00:58:56] This is humanity's Icarus moment. Thanks for bearing with me. The written piece with references will be out when it's out. I will talk to you next week.