What remains scarce after AGI? – Alex Imas and Phil Trammell

76 MIN · YOUTUBE · HTTPS://YOUTU.BE/JJ-KBHZUOHS?SI=U876YLEENUGL2YY6
https://youtu.be/Jj-kBHzUohs?si=u876YlEENUgL2yy6

SUMMARY

The discussion features Alex Imas and Phil Trammell exploring the implications of advanced AI and automation on the economy, particularly focusing on labor share, wealth distribution, and the nature of scarcity in a future dominated by AI. They emphasize the uncertainty surrounding predictions in economics, the potential for new job creation alongside automation, and the importance of understanding human preferences in a changing economic landscape.

- *The relational sector, where human involvement adds value, may remain scarce even as automation increases in other areas.*
- *Historical economists like David Ricardo predicted automation would lead to unemployment, but new job creation often offsets this.*
- *Economists should focus on scenario modeling rather than individual forecasts due to the unpredictability of economic outcomes.*
- *The labor share of income has remained surprisingly stable despite technological advancements, raising questions about future trends.*
- *The elasticity of demand for goods and services plays a crucial role in determining how automation impacts employment and wages.*
- *Developing countries face risks of being left behind in the AI economy, necessitating strategies for wealth indexing and investment in technology.*
- *The future of wealth distribution may depend on whether AI becomes a broadly accessible resource or remains concentrated among a few companies.*
- *The narrative around AI's impact on jobs is often negative, overshadowing potential benefits and the transformative power of technology.*

Today I'm chatting with Alex Imas, who
is Director of AGI Economics at Google DeepMind and Professor of Economics at the
University of Chicago, and Phil Trammell, who is Head of Economics at Epoch
and research scholar at Stanford. In general what I want to understand in this
interview is what economics tells us about what we can expect in a world with more
and more automation and more advanced AI. I want to understand what that tells us about
what will happen to wages and the labor share, what the best way to tax and redistribute
the wealth generated by AGI will be, and what kinds of things will be scarce.
What is scarce tells you where the value will accrue.
I want to start there.

What are some plausible
candidates of what will be scarce? Something like the relational
sector, which is defined as services and goods where the fact that a human was
in the loop is part of the value of that product. Because humans are naturally scarce, if we have
automation where a lot of other things stop being scarce, we will still have scarcity in the things
that humans are involved in and in the loop for. I'm curious to understand whether humans
doing services for other humans can ever be a big part of the economy.
Here's maybe one intuition pump.

In a world where AI can physically do anything
humans can do, there's this whole machine economy where they're building factories and
doing research and coming up with new ideas. Humans may or may not be involved in the physical
production of those things, but probably not in the ultimate limit, if robotics is solved.
If you don't care about humans being involved in that process, why would they be?
But then there are these other things you point out where we actually do want
the ballerina or the barista to be a human.

That's part of the value of
going to a cafe or a performance. But only humans have that preference.
So there's this human economy where humans are doing services for each other, and part
of their wealth is flowing to other humans. But part of their wealth is also flowing out,
because they will want some of the automated goods this machine-only economy is creating.
This is not a closed loop. A lot of things in the machine-only economy are a
closed loop because the machines don't care about getting the human barista to make them a coffee.
Within that model, isn't it intrinsic that the human-only economy will become
a smaller and smaller share?

I would like to pitch a
rephrasing of that question. My view is that the individual forecasts
economists like us would make, as individual forecasts, are not necessarily very useful.
There was a blog post by Andrey Fradkin, Brian Jabarian, and Andrew Koh that came
out yesterday looking at economists' forecasts about the labor market.
What they found is that there's a ton of disagreement in every single direction.
What they advocate for, and I'm in agreement here, is that rather than thinking about
individual forecasts, we should be generating prediction markets where you get aggregate
forecasts and wisdom-of-the-crowd effects.

The reason I think this is because we have
been famously terrible at forecasting. Let's go all the way back to 1820.
This debate we've been having is actually 200 years old.
David Ricardo is one of the classical economists, not neoclassical.

When the Industrial Revolution started happening, he wrote a bunch of stuff saying, "This is going to be great for everybody. Prices are going to come down."
But then he turned around and said, "Wait, I can see all these jobs that are creating value are going to be automated by these machines.

This is going to be really bad.
Everybody's going to become unemployed, and there's going to be political unrest."
And if you look at Ricardo's predictions, they're actually right.
All those jobs that made money in Ricardo's time got automated.
If David Ricardo woke up and somebody told him all those jobs did get automated, and then asked him, "What do you think the prime-age employment rate is in 2026?", I think he'd be surprised to be told it was the highest it's ever been other than 2000.

We have the highest number of employed people that could potentially be employed since 2000. That was the peak and now it's the second peak basically. What David Ricardo ended up missing is that you have these economics of structural change, where everything that got automated became cheap.
People had more money to spend, and then they started spending it on services.
This is the lump-of-labor fallacy. David Ricardo didn't consider that new jobs would be created.

But it's not obvious that money would go to services. Why wouldn't it go to more automated goods and something like that? I'm not using this anecdote to say this is what's going to happen now and that we're going to have full employment.
I'm using it to say it's really hard to make predictions.
What may be a really useful tool that economists have is to instead start with a premise.
Maybe we start today: labor share is zero.

Labor share has gone down.
What could possibly explain this? Let's write down an economic model of what happened. Phil will talk about this later today.
Or you can write down a model that asks, "What if labor share just stays the same? What can make that happen?" If you don't take anything else out of this conversation from me: We don't have any data. I've been saying we need a Manhattan Project for data. We don't have data on consumer demand elasticities.

We don't know what they are.
We're not really tracking what jobs are getting created or destroyed.
The O*NET database, with all of the tasks and different jobs, has been rarely

updated and is super low quality. What is really useful is to think about the potential scenarios, map them out, and say what dimension of scarcity will generate each scenario. If there's full employment, we can talk about the relational sector. If the labor share collapses, we can talk about other sorts of scenarios. That will tell us what data we should be collecting. It's probably worth defining labor share and capital share real quick. The whole economy, the total sum of goods and services sold, is either paid out to people in wages or it's paid out to capital, which is to say, there's rent on buildings and shareholders of companies that get paid out.

For many hundreds of years, ~60% of the economy basically gets paid out to humans in wages, and the other 30-40% gets paid out to people who own machines and land and claims on companies. The question is, if 60% is going to wages right now, does that shrink as AIs get smarter and better? This is a Kaldor fact.

We should stress this. It's incredibly surprising that it's over 60% after the Industrial Revolution and all of the automation we've ever seen. Some people are worried it's an accounting error that it's been so constant. There's even a controversy right now. Some might say labor share has been falling in the last 20 to 30 years. But there have been a lot of accounting changes in the last 30 to 40 years.

For example, Atkinson has a paper showing that if you keep the accounting constant over the years, labor share hasn't even fallen ever. But it's not that surprising, right? Phil, you made this point that if labor and capital are complements, you need both to do anything. It would make sense that you'd need to pay both of them to get something done. You have had stuff be completely automated. There's a sense in which nothing has yet been completely automated.

Look at the network-adjusted factor shares of a good. Look down the supply chain and not just the final step and how much of that is done by capital and labor, but what went into the machines that can automate that final step. You'll find that labor is adding a lot of value down the supply chain. Computer and electronic products in the US have a very stable network-adjusted capital share of around 50%. It's not 100%. I do think there's this qualitative shift that I think we agree is coming, which is that there will be at least some goods whose network-adjusted capital share goes to one. The whole supply chain can be automated, and there's no part in it that we care intrinsically about having a human do.

That will be a qualitative shift.
Interestingly, the implications of that shift for the overall capital share are ambiguous.
Let's say we've got two sectors: the human-intrinsic sector with the
ballerinas, and everything else. Right now, everything else has been
scarce because of the lack of labor in it. But if we fully automate the supply chains for
everything else, and we satiate in everything else really fast, then the quantity of everything
that's not a ballerina goes to infinity, but the marginal utility in that stuff goes
to zero faster than the quantity is rising.

I also want to move away
from the ballerina example. The point I was trying to make in my
post—working backwards from a particular scenario—was that the ballerina and the
performer are the wrong reference class. Right now we have a lot of jobs
where you have different tasks. This is the task-based model of jobs.
Take a doctor, what is their job? They're filling out insurance documents.
They're going and calling different pharmaceutical companies.
One of their tasks is to see the patient and talk to them, but that's not the main part of the job.
You could have a job and a service or a good be a product of different types of tasks, and
you can automate a ton of those tasks.

If the consumer is willing to pay more
for a product or service where every single task is automated except for that
one part where the doctor is delivering the diagnosis and providing support, we would
call that job part of the relational sector. People are willing to pay more for the
human to stay in the loop in the job. We don't have data to say, "Here
are relational jobs, here are not."

You literally need to collect
data of the following sort. Do a conjoint analysis of your willingness
to pay for this service or good. Here's the counterfactual where
everything is produced by machine. Here's the counterfactual where this
one task is not produced by a machine. What is your willingness to pay?
What is your elasticity for the human to not be in the loop?
If I don't have that data, what prediction am I going to make in this story?
Isn't there another point, which is that there are a lot of fully automated
goods that don't even exist yet?

And you can't collect any data right now
about, say, how much people will want to keep buying more and more of some drug that makes
you healthier that is fully produced by the AIs. Absolutely. That's kind of Phil's point.
You could have an increase in variety in capital where you don't get the satiation.

You're increasing variety, so you're not hitting that diminishing marginal utility point where most of your income is going to the human sector.

If that increasing variety is fast enough,
and there is no such increasing variety in the human sector, then you can get all of the relational goods you want, but it doesn't matter for labor share.
It goes to zero. Phil, I liked your analogy to some
Mongolian economist sitting around in 1400 thinking about what will be scarce and the limits of that kind of analysis. I think you should talk about that.
Just look at the goods available to a Mongolian of the distant past.
I'm no expert on this society, but I know that they didn't have
nearly the variety that we have now.

Look at the jobs that were intrinsically
human, like being a singer. And then you look at the things that
were not intrinsically human, like the transportation services provided by their horses or the different kinds of food they had. If they just held the varieties fixed in both categories and asked, "What will happen once we have a lot more automation?", they might have said, "We'll just satiate in horse-like transportation and in yogurt and in yurts.
Those shares will all go to zero, and we'll be left spending all of our money on singers."
But of course, that's not what happened.

As we've accumulated more wealth and more advanced
machines, we've expanded the range of things other than singers to spend our money on, and the share spent on singers has stayed negligible. Likewise, that's my central prediction about how the future unfolds, though it could go either way. I was going to make a point
and I realize it's a fallacy, but the reason it's a fallacy is interesting.
It's just hard to imagine a world where there are trillions upon trillions of robots, but there's only some billion-odd humans, and the cumulative amount we're spending on robots is less than what we're spending to pay Magnus Carlsen or— Financial advisors or doctors or tutors.
Right, or podcasters or whatever.

But then I realized why it's a fallacy.
The number of transistors in the world has literally trillion-X'd, maybe quadrillion-X'd.
Your colleague Chad Jones has a very interesting result about how the share of the economy that is going towards paying for computing, paying for the transistors, has been decreasing.
The point you made is that one way to think about Moore's law is... What sets price?
Supply and demand. So not only are we producing more transistors more cheaply, but also the value of the marginal transistor is decreasing.

As you were saying, another

way of saying Moore's law is... I like the pessimistic framing of Moore's law:
every 18 months, the value of computation halves. We're running out of uses for computation
so fast that it's sustaining Moore's law. This is relevant to a conversation about AI where
maybe for the first time, this is no longer true. The famous fact here is that an H100 costs
more to rent now than it did three years ago, even though we have much superior technology
and much more compute in the world.

Because as models get smarter, the
opportunity cost of compute gets higher. This is Phil's point about increasing variety.
What we have done is increased the types of things that people demand from capital.
Now all of a sudden you have a new variety that you could be using
capital for, and you jump back up. You could imagine we just never
satiate demand for compute. As long as that stays the case, the
share of the economy that is going towards compute would keep increasing.
That's the big question. That is the ultimate question that we need to be looking at.
What number of new uses are we finding for that compute where you have the demand for these uses?
What I want to emphasize is that a lot of models in economics, especially in the space that we're
talking about, take demand as almost exogenous.

They don't unpack the psychology
of what people actually want. What got me thinking about the idea of
the relational sector is work that I was doing on the fact that there does
seem to be this intrinsic value. It's not just because it's scarce; it's
because there's some intrinsic preference that people have for empathy, connection,
and interacting with another person. One of the experiments that
we ran involved an art print.

We have an incentive-compatible
way of asking, "How much are you willing to pay for this art print?"
People are actually paying real money for it. Then we say, "Look, there's
only one of those art prints, and it's either made by AI or by a person."
These are between-subject conditions. With one, you get the effect that the person-produced art
print is valued much higher than the AI version. Then, in a set of other conditions, we
say there's 500 of these being produced.

For the human-made one, the price goes down
a lot because it's no longer seen as making a connection with this one artist.
With AI there's no difference. AI is already viewed as a commodity.
We need to do a lot more research on this, but it seems that's the key difference
between this and something like a horse. A horse was an input into an output, where
you can replace the horse with something else. You only care about the output.

The only way this relational story works—and this is what we need more data on—is if a human is not a horse in the sense that they are providing value from the output, where if you replace the human, the value of the output decreases.

If that's not strong enough, and if it doesn't hold for enough sectors or enough jobs, then this story doesn't work anymore. There aren't that many institutions that have thought as hard as Jane Street about how to turn smart people into some of the most competent researchers and engineers in the world. This relies in part on an apprenticeship model, where new hires are paired with senior mentors. But Jane Street also runs a bunch of classroom-style lectures and hands-on bootcamps. These courses cover a range of topics and they go pretty deep. There's one lecture that focuses on reverse engineering systems with tools like strace and gdb and another that teaches you how to profile code down to the cache hierarchy level. Importantly, Jane Street designs these courses not just to teach the relevant object level skills. but also to impart the relevant tacit knowledge.

For example, their week-long neural net bootcamp starts with general theory, but then quickly progresses to how to apply neural networks to trading. And here they cover the specific obstacles that Jane Streeters tend to encounter and the workarounds they've come up with to get around them. Jane Street takes this sort of learning incredibly seriously. Every office has dedicated classroom space and courses are prioritized as part of regular work. If you'd like to work at a place like this, Jane Street is hiring.

You can check out their open roles at janestreet.com/dwarkesh. There's one possibility which Molly Kinder has written about, this "Messy Middle" scenario. That possibility made me think about whether it might be better to have—at least as far as wealth distribution and redistribution go—a much faster AI takeoff. I want to ask you whether the following possibility is at all likely, or if there's any set of assumptions that can make it so. AI makes it possible to automate jobs such that many people are losing their jobs, but it doesn't create enough wealth, while the process of automation is happening, to basically pay off the people who are getting laid off and create a Pareto improvement, where everybody's getting better as a result of AI automation.

Of course, there's a trivial sense in which that must be true. Whatever money the company is saving by not paying the humans instead of just paying the AIs, those resources still exist in the economy and can just be paid out to people. But there's going to be some allocative inefficiency. The government doesn't know exactly

who got laid off because of AI. There's a political problem. If the
Meta worker gets laid off first and they were making $200,000 a year, is there a
politically sustainable situation where you give them a $200,000 check a year when there
are many working people making much less?

Do you find this scenario plausible, where AI is
automating a bunch of things, but there isn't as much wealth creation as there is automation?
I think it's possible. To me, it does seem like a pretty narrow window.
My guess is that if we have the technology to automate so many jobs that
it becomes a new kind of political problem, then the pie will also be growing really fast.
Well, unless in all of those professions it's automating, it's just a hair more productive.
So the cost of all the capital to replace all the software engineers is just a
hair less than the cost of what we've been paying the software engineers.
Why is it implausible that a company can save money by laying off a bunch of software engineers?
And in the long run, there's a Jevons paradox, and we can't anticipate in advance what we'd do with
more software, and surely there will be more uses.

But in the short run, the effect is
just that a lot of people are laid off, and they still need to figure out how they can
use a million times more JavaScript tokens. Phil and I have been writing about these
things, and we have mathematical models in the back of these things.
We don't have any political economy in any of our models.
Andy Hall wrote a really nice blog post about the politics of AGI, and
he made a really interesting observation.

If there's a 2% increase in unemployment,
the political winds completely change. Unemployment has a huge effect
on what happens politically. Referring to Molly's excellent essay, I think in
some ways one of the worst scenarios is a drip scenario because of the political economy piece.
What you might see is people not really being unemployed en masse, but moving into
sectors that pay them less money.

This is what happened with phone
operators between 1920 and 1940. Phone operators were completely automated, but it
took 20 years, even though the technology existed. There was this drip. It wasn't like
this giant sector just disappeared. There's a really nice QJE paper on this
showing that they got reabsorbed into the economy, but at lower salaries,
and they were mostly underemployed.

That's the scenario Molly was writing about,
this messy middle where things aren't a disaster. We saw with COVID that the fiscal response

can move quickly if there's an emergency. An emergency is a quick uptick in unemployment, which could even look like 2-3%. That becomes a national emergency if it happens fast. The concern is that whatever you're saving on those white-collar workers, if that's not growing the economy but just creating saved resources that can be allocated elsewhere, is that enough to do a broad-based redistribution scheme? You have the money you've saved off a couple of people.
Unless you can figure out exactly how to get it to them specifically, you have the problem of, "Can I do a UBI off the money I saved by laying off...?"

You're basically saying the pie did not grow that much. You're just displacing a bunch of people, but that didn't grow the technological frontier of what the economy can produce.
Then there's a question of whether every time this has happened in history, the technological frontier has expanded a bunch. I think that's the case.
Simply in history, the technological frontier has expanded.
I think Phil made the same point. It's hard to imagine that sort of scenario where you are getting intelligence that's just enough to replace the software engineer but still costs a lot of money.

It's just a hair less expensive than the software engineer, so you're not getting this abundance effect.
Where is the redistribution going to happen because the pie didn't grow?
This is very helpful. Many different things have to be true for this scenario to come to pass, each of which seem unlikely. One, it has to be the case that it is possible to automate entire white-collar jobs, but only in a piecemeal way.
That is to say that you can only automate software engineers, but that same program can't also automate an accountant and an analyst and whatever.
My model of intelligence is such that—both the breadth of tasks it requires to do something like software engineering and what intelligence is—if you can really just lay off all the software engineers, you've got enough in the bucket there that you could automate all kinds of white-collar work.

There are huge amounts of potential savings that have happened as a result of these layoffs, and also AI is going to be cheaper than human labor.
If both of those things are true, this messy middle scenario where we literally don't have the wealth to go around seems unlikely. Then the question is, what is the best way to tax it and redistribute it? I have some thoughts. I think it's really important to outline the costs and benefits.

First, there's differential complexity in implementing these things. Two, they differ in the timeline

of being actually helpful. Something like universal basic capital is not
going to generate returns for something that happens in six months.
You probably are going to end up with a layer of things.
Take a negative income tax, for example. You implement it, and the day it turns into
law, you already have this insurance that there's a floor where everybody gets a certain
amount of money, and if you earn more money, you get taxed more.
But there are positives and negatives to a negative income tax.
With UBI, for example, I worry a lot about the political economy implications.
If people are just dependent on a check, it really matters who's in power.
Right now, we're endowed with labor that can turn into income.
When that is no longer the case and we are at the mercy of the elected official for
basic needs, that feels like a power-sharing arrangement that's really dangerous.
But wouldn't that be true of any sort of government redistribution program?
With something like universal basic capital, where you have an ownership share and property
rights for capital, you just have a share.

You're a normal shareholder.
You're just a normal person. But this goes back to the question of
indexing, because if indexing is hard, then universal basic capital is hard.
That's the problem of universal basic capital: targeting.
What do you target to put into people's portfolios?
Like, what if Anthropic goes to zero, but some random robotics company takes all this over?
Exactly. That's the risk of universal basic capital.
With a negative income tax, you have the same sort of issues as with UBI, where somebody
comes into power and says, "We're not going to do that anymore," and people can't work, and
then you have the issue of the floor being gone.

One concern with the wealth tax
is that there's no politically sustainable equilibrium at a 0.5% wealth tax.
This happened with the income tax, of course. It starts low, it’s for war or something, and then
it slowly escalates until the marginal income tax rate in the US is on the order of 40%,
and in certain states, upwards of 50%. With a capital tax, is there a reason to
worry that it would distort investment? Would people just say, "Why would
I invest in Anthropic or Intel?

The government is going to take larger and
larger shares of it and dilute my share." Hold on. It's worth separating how the revenue is
raised, what's taxed, and how it's distributed. It could be that the government hands
out shares of Anthropic to everyone by a broad-based tax and then buying Anthropic.
Which would probably be the right thing to do. Hopefully, some populist proposal
doesn't interfere with that and expropriate some particular company
that everyone happens to know about.

You're suggesting there could
be some sort of optimal tax. We're taxing externalities or we're taxing land.
I guess we probably need to tax something other than just those two things.
Or consumption. Ok, a consumption tax, like a European
value-added tax, allows the government to go buy a bunch of stocks, and then they
just distribute those stocks to everybody. That's David Autor's...
That's not going to be that different from just redistributing the
stocks, but it will be a little different.

That was the proposal for
Social Security, by the way. That was privatizing Social Security.
It's worked so far, but there are questions about how long it's going to keep working.
Privatizing Social Security was basically giving everybody a basket of stocks.
People talk about whether there's a white-collar apocalypse already.
Is there any evidence that suggests there is mass automation or
unemployment as a result of AI already?

A lot of people are looking at it. This is an area where there's a lot of
eyes and a lot of data being produced. The Budget Lab over at Yale is
doing really good analysis on this. They just recently released a report, and you
really have to squint to see anything happening. If you want to take an approach across the entire
economy, even looking at software engineering, the most exposed sectors, there's
just not really anything going on.

There might be a little bit of a signal about
junior developers getting jobs less than before. But that's a "less than before" rather than a
level shift, as in there's actually an increased demand for senior software engineers, if anything.
If you look at the trend, for junior developers, it's a bit below trend.
So you're saying the growth is slower than before, but there is still
growth even for entry-level software engineers. What do you think is going on with the anecdotal
evidence of graduating college students saying that they're finding it harder to find CS jobs?
I think that's anecdotal evidence.

You think it's always been hard to get jobs
for some people, and now it's getting turned into an AI narrative?
Same with the layoffs, where it's probably just a normal layoff,
and they turned it into an AI layoff. You have to be careful with all of this.
There are these public coordination devices. Let's say we get into a narrative where if
you're a firm and you're not laying people off, then you're seen as not adapting AI enough.
Then you're going to just get a cascade effect of firms needing to keep up with the
Joneses in terms of starting to lay people off.

That's super worrying, where the firm
might actually be worse off after the layoffs than before, but it's just doing
the layoffs to have the perception that, "Look, we're not behind the times.
We're using AI." You probably heard these anecdotal stories of token counters, where
you have to maximize tokens and things like that. Right now, we don't really have any
evidence of a white-collar bloodbath. Is that surprising at all, given
all these things AI can do?

This is a story as old as time.
If you automate some complementary task, the overall bucket of things—the human labor which
complements the automation—will increase in value. One of the statistics that's really important
for that argument is elasticity of demand. Take the O-ring model of jobs.
A job is a series of tasks. Let's say the AI automates nine out of ten tasks.
One task is not automated.

If that person can now focus in on that
task, the job will become more productive. If that translates into a price effect where
the product is actually cheaper, and if demand responds enough where it's being bought
more and the service is being used more, that could actually lead to more hiring.
A lot of people on the internet have been making that argument very generally, saying,
"Look, if anything in the data, we're seeing an uptick in software engineering demand."
Which suggests that at least for now, given the way that jobs work, it might be elastic enough.
I think this elasticity of demand argument is incredibly important for a lot
of arguments that people make, or just a lot of labels that people use without
understanding what the underlying causation is.

People often talk about Jevons paradox.
This is the idea that as something gets cheaper, you will want so much more of it that the
total amount you spend on the thing increases. Famously, this happened to
coal in Britain ~200 years ago. But really this only happens if the
demand for something is highly elastic. There are many things for which
there is not super elastic demand. If oil, for example, gets super
cheap, it's not like magically— Or insulin.

Exactly. It's not like magically there's going
to be so many more cars that now we're going to be using way more oil than before.
At least not in the short run. Exactly. The long-run elasticity is
higher than short-run elasticity. But even in the long run, agriculture famously is
the example where we can produce way more food if we dedicated the same portion of the economy
that we dedicated to agriculture in the past. We're already producing more food regardless, but
we could produce even more if the same portion of the economy that was producing food 100

years ago was currently producing food.

But you eat enough, and then you're done.
The claim with software is that it is not some inherent property of markets that as it gets cheaper, you'll just keep wanting more of it. The thing about software is this is a particular kind of good where as it gets cheaper, we'll want more and more of it.
It is also highly relevant, and you wrote an essay about this—a lot of this podcast is me summarizing your essays back to you.

There's this very viral scenario
planning about the future by Citrini, predicting that as a result of automation and very powerful AI, there will be a recession. White-collar workers will get automated, their salaries will no longer be available, and so there will be a slump.
Do you want to recapitulate why this might be implausible?
Part of it is plausible, part of it's not. The part that we started the conversation with is the idea that there could be a lot of unemployment. If the speed of automation is quick, people could get laid off, and they may not find work very quickly.

We can quibble about the unemployment part of
the Citrini essay, but that's not the issue. The issue is that they talked about negative economic growth. What I did in the piece, that Phil and I had a back and forth on, was to say, let's start with the proposition that there's negative economic growth. What conditions do you need in the economy to get negative economic growth? It turns out the conditions are pretty improbable.
One thing that you need is for the holders of capital, rich people basically... Basically what you have in those sorts of scenarios is a reallocation of wealth and income from lower-income people who are using their labor towards tech capital owners.

So you need demand to be bounded, like a hard
bound, not even a soft diminishing sensitivity. You need for them to eventually say, "I've had enough. I don't want to spend any more money."
And for that money to not enter as investment. Then you can get negative growth.
The crucial thing is, even if we don't want more shit, the world in which there's a singularity and we don't want to invest more money is crazy. We're not saying, "Let's build more data centers. Let's build more fabs." Even though we have AGI, we're not investing in more data centers to run the AGI and that's driving more economic growth.

I sent the essay to Phil, and Phil wrote back
being like, "This is pretty dumb," like my essay. He said, "You're trying to say that there's going to be negative economic growth, but these are very implausible conditions."

And I was like, "That's the point of the essay. These are very implausible economic conditions." That's where scenario planning really shines. You have the Citrini essay, which was great that it was written because it started a conversation. It's so intuitive, this idea that if there's demand collapse, we can get the economy to shrink.

You could get that with a depression.
In the Depression, the technological frontier didn't expand.
Here, the technological frontier is expanding. You actually have abundance. For abundance to generate negative economic growth, that's really hard to get.
Google recently announced Gemini Omni and its video editing capabilities are incredible.
You can upload a video and then tell Omni to do things like change the background or adjust the lighting or add or remove elements. All while keeping everything else consistent.
But Omni isn't just a video editor.

I got a chance to sit down with the
research and product team behind Omni and I learned that it's a preview of how future Frontier models will be trained. It can take in any kind of input, whether that's text or audio or video. And while it doesn't currently do so, architecturally it's capable of just seamlessly outputting images or text.
So it's really a bet on the multimodal data transfer hypothesis.
The model becomes better at predicting one data type by seeing the others.
For example, Omni is really good at accurately rendering text on video, even though Google didn't specifically target that capability in this model.

And Omni is the next step towards
more accurate world models. Because in order to predict the next frame of a video, you have to have a deep understanding of physics and spatial dynamics.
As Omni progresses, it'll be interesting to see whether it can close a Sim2Real gap.
Because it's much harder to collect data in the real world than it is in simulation, robotics progress has lagged other applications of AI. But if you have really good video models that can simulate reality, maybe that stops being the case.

In the meantime, if you want to try Omni,
you can check it out in the Gemini app at gemini.google or use it in Google's AI Creative Studio, Flow, at flow.google. We were talking a second ago about why there isn't more automation as a result of LLMs. One plausible mechanism could be, as you were saying with the O-ring theory... O-ring theory refers to the fact that the Challenger shuttle blew up because one component malfunctioned, and it destroyed the whole thing.
Maybe that's a more general model of how goods are produced in the economy.
You have to make sure everything is reliable and works well.
So you can't automate an entire job to an AI right now.

Even though it might be able to perform it at some probability, you need extreme reliability in order for it to not destroy the finished good.

This might explain why there's a lot less automation now than there otherwise could be. But I think it works in the other direction once AIs get advanced enough. Integrating humans into the production flow of future goods will become difficult. Even beyond the arguments about how humans will be more expensive or less capable, there will be whole production flows organized for AI labor. They're talking in neuralese. They're thinking many thousands of times faster. So even if there's some comparative advantage where it makes sense to hire a human, there will be transaction costs and worries of reliability that will actually make it hard to integrate humans into future production flows.

That seems right to me.
In particular, I just want to distinguish between the point that if you automate nine-tenths of a job, people might shift over to the last tenth, but there might be ten times more work demanded of them. Compare that to the model of O-ring automation from Gans and Goldfarb recently. If you can only automate nine-tenths of the job, but you do it to a lower standard of quality than the human could, you might not want to automate even those nine-tenths.

That's the thing that could totally port over.
Symmetrically, it could be a reason why we don't use a human for one-tenth of the job anymore, because a human just can't perform it to the level of quality that the AI can perform the other parts of the job, or the level of speed. They end up pulling down the quality or speed of the finished product. By the way, the model you're talking about seems extremely plausible to me for why more lawyers, accountants, or even software engineers are not automated.

There are cases where there's a pretty good probability that the thing worked as you expect, but the thing you're paying the lawyer for is: "No, really, my company's not going to go under because—"
You're also paying for a lot of regulation-type stuff.
With lawyers particularly, you need some entity to back up the product.
You need ownership of the product. You need somebody to be able to fire or hire, and there are licensing issues. There's a lot of regulatory layers that are also going to be keeping—even if there's no relational element—humans in the loop that have nothing to do with the ability of the human to actually perform the service.
All of these frictions on the political-type decisions that we are accustomed to only trusting humans for—legislation, being a judge, being a juror, or all the licensing that keeps certain professions human—that all strikes me as transitional.

What we expect to come from a human and how we organize our politics has changed so many times throughout history, from little hunter-gatherer bands to empires to whatnot. Once an AI-run political system is much more efficient than the alternatives, those will probably tend to out-compete the others.

Speaking of which, we've been talking about what preferences humans currently have and what impact that has on what kinds of goods will be scarce in the future. But of course, we'll have different kinds of entities in the future: AIs. There was a time when there were no humans on Earth, but evolution selected for agents that have specific drives and preferences because those tend to survive the most, and those preferences now determine what a hundred-trillion-dollar world economy produces.

Why not expect the same thing from AIs in the future? This is not even a world with catastrophic misalignment, where they just kill everybody. But there will be evolution of, even if not individual AIs, firms which have AIs as part of them. What will that evolution favor? It will probably favor firms or agents that grow. There's a selection argument that things which grow will be more prevalent. Maybe just based on that, you can make some predictions about what their preferences will be. Is the kind of entity which prefers to have human-intrinsic goods going to be the kind of entity that accumulates resources the most?

Probably not. Probably it saves more and has unsatisfiable demand for whatever the relevant resource happens to be. Compute is an obvious one. Can we use that to make some predictions about the non-human preferences that will be guiding the future? If there's an AI that has its own welfare, is fully autonomous, and is making its own decisions that are welfare-relevant, to be honest, I have absolutely no prior that it would prefer to deal with humans.

There's no reason. But let me take the other side of that argument. Humans' preferences to be interacting with one another, to trust and empathize with other humans versus a simulated AI, I think it's a really important question whether those will change. I've heard a lot of arguments saying, "Look, right now we're just not used to the technology. What you're thinking of as relational... At some point, people are just going to see an AI therapist as a superior product, and they're not going to need the empathy that the human is providing."

I think this is actually a

really complicated question. Here's one argument for why it's not going
to go away, and it has to do with evolution. Let's say there are two types of people.
One person doesn't really have this preference. They can just interact with an AI,
whatever can simulate it better. The other one has almost a moral emotion—using
Jonathan Haidt's framework—against offloading those sorts of social interactions to an AI.
Which of those two people are going to reproduce, find a mate, all of these sorts of things?
I think the answer is clear.

It's the second one that has
the preference for other people. Depends on how the reproduction is happening.
Fair. But if we're in the world where reproduction is still happening the way that it's
happening, I think... And this is a big question, I'm not making a prediction.
You had David Reich on the show. His point on the last podcast was that
we're buzzing with natural selection. So even if you get some sort of indifference now,
you might get selection to point into an even stronger preference for other humans.
Here's one way to think about it.

How is the wealth of the richest
people in the world instantiated? We were having a call earlier, and you
made the point that their consumption is more geared towards relational goods.
Like Mark Zuckerberg is hiring MMA instructors and dancers for his wife's birthday, and so forth.
But most of his wealth is just stock in Meta. As a controlling shareholder, he could say,
"Meta, turn all this wealth into dividend income, and I will just spend that on consumption."
Instead, he would rather have his wealth compound and have Meta build more data centers.
So you don't even have to change humans for this to be the case.
The humans who are wealthiest—and growing wealthier because their wealth
is compounding—just have this almost Nick Landian preference for accelerating capital.
That does seem to suggest that this is an important determinant of what kinds
of things are produced in the future.

There are two ways you could get the two kinds
of people, one of whom prefers a human therapist and one of whom is fine interacting with the AI.
If they both satiate equally quickly in capital but the one who likes the human therapist also
just likes having some human-intrinsic services, then the marginal value of capital in the future,
compared to the marginal value of capital today, for each of them if they start out equally
rich, should be basically the same.

There could be interactions and whatnot,
but basically, that should be the same. If what's driving the difference is that one
person just doesn't satiate in capital because they're engaged by the prospect of exploring
the universe and turning their head into a galaxy brain or whatever, and the other

one satiates, then the person who doesn't satiate in capital is going to, if they're being rational, have a higher savings rate. So in the long run, they're going to have most of the wealth, and the overall capital share will basically be the capital share of that person's spending, which is going to be one.

It's important that we're not talking about a hypothetical future. Elon Musk is talking about mass drivers on the moon. He's by far the wealthiest person in the world. Obviously, currently his investments are going towards humans as well as machines, but I don't think he cares particularly that his future researchers and engineers are humans versus AI. And he manages to reproduce fast as well. So I just think it's worth drawing that distinction. There are currently some rich people that don't seem to satiate quickly in capital, and so maybe in the long run they'll save the most. That does seem right to me.

I would also say, even if they do reproduce more slowly biologically, that might just not matter that much in the long run if they can live forever. The living forever is key. Again, we're scenario-building here. If you could live forever, a lot of stuff changes for my story as well. To your point about rich people not consuming a lot and investing, this will all depend on the returns to capital. Right now, the returns to data centers are super high, but if we get into a situation where people are satiated with capital, then the returns to accumulating capital are going to be lower. Then these rich people are going to be consuming more, because the incentive to invest is smaller. Basically, you think about the general equilibrium of this sort of process... We have gotten tremendously richer since 1820.

Many more people are investing, but you're still getting a consumption response which keeps people employed and labor share high. That's because— Hold on. Wait, not necessarily. I think you're probably making the same point. It could be that their investment has to be titrated through actual laborers who have to do things for their investment to work. In the future, only the consumption is human-mediated, right? Because the investment can just be done by the robots.

So we're in the scenario of how you can keep high labor share. Let's take that scenario. In the scenario with high labor share, for whatever reason, the returns to capital are going to be lower. That's right. To the earlier thing where we were saying why the messy middle is implausible, I feel like we can do a similar thing here. For our returns to capital to be lower,

the growth rate has to be lower, right? It certainly has to be lower than what we're expecting through the period of transformative AI.

If there's explosive growth...
Yes and no. The capital stock could grow quickly, but the price of capital goods relative to consumption goods could be falling faster than the capital stock is growing. It's the difference between the potential frontier of technology and the realized prices of these things. Because you have relative prices.
So you're saying I could be putting my money towards earning 30% interest and investing in data centers, or whatever. There will be something in the future, if the growth rate is high, that earns high returns.

Or, as a result of all these technological breakthroughs, there's some cool product that I really want to buy right now, and both of those will be compelling options. Yeah. It doesn't have to be a new product. It could be a human-intrinsic product. Although, if it's a human-intrinsic product, we would want to have it much more in the future than we want it now, because the thing it compares against is— We might want it the same as we want it now in the sense that the marginal utility in a ballerina performance is exactly the same as now. But the marginal utility in a robot might just be a lot lower than now.
So in units of robots, we want it a lot more than we want it now.
Would the interest rate be 30%?

It depends what you mean by the real interest rate. It might be that every robot now can turn into 100 robots next year. So in units of robots, the interest rate's 10,000%. But if the price of robots is falling really fast... Prices adjust. I think that's the whole point. Here prices are adjusting in this interesting way that too many macro models don't allow for. What's happening is what would be called investment-specific technical change.

The price of capital is falling relative to the price of consumption, instead of doing the standard macro thing of saying there's just output, this chimera of a thing called output, which one for one can be allocated to capital or consumption. That's not going to be true in this world. Every unit of capital next year is giving up way less consumption than each unit of capital this year. One robot now turns into many robots next year, but the number of ballerinas is the same.

Again, we're going to go back to the increasing varieties thing. If all of those extra robots next year

are actually different varieties of robots and I'm not getting satiated on those robots, then it's a very different story. But now we're talking about the consumption world. For the investment side of things, there could be just some greedy titan of industry who keeps wanting more and more robots. That alone would be enough to increase the marginal value of robots and therefore decrease labor share?
Yes.

Why are we not expecting greedy titans of industry to keep existing? Greedy titans of industry historically have built libraries and— But that's because they die, and they're like— Oh, they all die. Everybody dies. Well, we'll see.
Conditional on people dying... You had a guest on the show who said to understand the future, you should think about the past. You could have new types of titans being born whose entire reason for accumulating wealth is just to accumulate wealth. But a lot of the time, at least historically, the wealth accumulation process is part of a large social interaction amongst peers, amongst the community, where you want to be admired in some way.

The stylized fact of titans of industry is you accumulate the capital, and then you buy a bunch of stuff.
I guess this is a historical question, but it does seem to me that in a lot of cases what is happening is that as they near the end of their life, they either hand it off to their children, who are worse stewards of capital than they are. They don't even manage to grow their wealth at the rate the economy grows, much less faster than the economy grows, which their parents were doing. They're like, "Well, I care less about my children having it than me playing this game of accumulating wealth.

So I'm just going to give it to some trust."
If people are living longer or if they can figure out some way to align their trust to this wealth accumulation process... It just feels like the evolution here is so strong. You just need a couple of agents that think this way for this to be the dominant thing determining the preferences of the whole economy, because this part is growing much faster than the other parts of the economy.

The part about satiation and diminishing marginal utility keeps coming up, I think it's really important.
If a person has an intrinsic preference for accumulation, that's just what they want, then I think your story is totally right. But that's just not how preferences usually work. You have enough hedonics in your life, and then the social status... Rousseau wrote about this, St. Augustine wrote about this. This is a basic part of preferences.
Now, you guys are arguing about something else, where you could have such high concentration that

you could just have a couple of exceptions to the rule, and that's going to be enough.
I have nothing to say about that.

I think the claim is a little stronger, not just
that you could have some exceptions, but that historically and today we see the exceptions.
They just haven't really taken over the economy historically because there have been
these dissipation shocks, as they're called. They've given it to their kids who squandered
it, or they put it in foundations which spent it. It's not really a shock, but...
People might have liked to fill the universe with monuments to
themselves and live forever, very wealthy.

It's a weird preference, but it's
not a hypothetical preference. I think that's the claim.
But who knows what's going on in their heads? Even without the intrinsic preference
for accumulation, there are some instrumental reasons why some people might value
accumulation, which is also worth bringing up. There's a desire for political,
philosophical, or religious influence.

People get into an arms race over what
society looks like and what people believe. Similarly but differently, because
it's not an arms race, there's just total utilitarian philanthropy.
When I think about why it might be good to have a lot of wealth in the future as a
good classical utilitarian, to me, the value—or at least one way you could have an almost
unsatiating utility function in having wealth in the future—is to create new happy beings.
They just add to the total welfare of the world.

This idea goes at least as far back as Bostrom's
astronomical waste point, that we could put Dyson spheres around the stars and turn all the
energy into really happy simulations and whatnot. I think the particular greediness of this
optimizer doesn't matter, what they're greedy for. Forgetting about utilitarian philosophy or
whatever, a pure von Neumann probe has... I don't know, is this an accurate way to say it?
They just have high marginal value for the random solar system they'll occupy because
that turns into more solar systems.

A von Neumann probe is a thing that can exist.
That's a very greedy optimizer. If we're talking about whether they'll dominate
the economy, maybe this is a technicality. But we only count final consumption
goods and investment goods as GDP. If there's just this phenomenon—
How does a von Neumann probe show up in GDP? Exactly. If we recognize it as a person that
owns itself, and it's optimizing on the margin between spending a bit more on a baby von
Neumann probe that colonizes another star system or a ballerina or something, and it

just doesn't value the ballerina very much... When we're talking about AI beings,
it just completely depends on how we're doing the accounting there.
What does the world look like in a world where von Neumann probes are possible?
Is it possible labor share is high?

I think it's possible the labor share
is high the way we usually count it. One of the biggest problems in RL right now
is credit assignment because you have these extremely long rollouts and you need
to know why they succeeded or failed. One of Cursor's researchers, Sasha Rush,
gave me a blackboard lecture on how they use targeted RL with textual feedback to deal
with this problem and train Composer 2.5. I filmed on my iPhone, so
apologies for the camera work.

So we've generated this output.
It's just a sequence of tokens. We're gonna send those sequence of tokens
to this model that's gonna read it, then it's gonna isolate a specific
turn that it says is problematic. Then we're just gonna do text manipulation.
We're just gonna take that trajectory and we're literally just gonna smash in some extra tokens.
After Cursor injects these hint tokens, they run another forward pass.
The trajectory itself doesn't change, but the hint causes the model to assign
lower probability to the error tokens.

Cursor then trains the original model
to match those probabilities, basically teaching it to downweight these specific mistakes. There's a lot more nuance that we
couldn't include in this mid-roll. If you want to watch the full
thing, I posted it on my Twitter. And if you want to try out Composer
2.5, head to cursor.com/dwarkesh. Do economists have any advice for countries
which are not in the AI production chain? If you're not either producing the AI models,
you're not producing the hardware that goes into AI models, if you're not Korea making HBM or
Taiwan with the fabs or the Netherlands with ASML.

India or Nigeria, what should
they be doing right now? If you're talking to Modi
right now, what do you say? I think the biggest lack of resources
that we have allocated in the economics profession is thinking about middle-income
developing countries in the age of AI. This is something I fault myself with as well.
There's not enough people thinking about this question.
There are scenarios where you get AI technology being allocated and
dissipating to Nigeria and developing countries, leveling the playing field, essentially
giving them a level up as far as capabilities.

But there's another world where, because they
don't have enough resources, they're not training the models, they don't have the hardware,
and they just completely get left behind. And because of automation, we can produce
commodities in developed countries now. Then we don't even have the consumer market.
That world looks pretty bad. This seems to me like an extension
of the messy middle case. One of the ways in which the messy middle might
only be bad in a narrow range of scenarios isn't just that it would be easy to redistribute because
the pie would be bigger, but because the interest rate would be way higher, and/or, equivalently,
the price of everything except human-intrinsic goods would be falling really rapidly.
They're sort of two sides of the same coin.

A little bit of savings would turn
into a lot of consumption next year. Things have to go really wrong for
us to just get over the threshold of capital being productive enough to automate
lots of work, but not be productive enough that the interest rate is high and the price
of capital-produced goods is falling a lot. Even without redistribution, a little
bit of savings will save a lot of people. You're saying if the developing countries
have some savings in the developed world, that will be enough to produce a
lot of surplus that they can then— They will now be able to consume
a lot using their savings.

But the messy middle could be wider in this case.
They're starting from such a lower level in terms of how much they have and how much it's
actually indexed to the global economy. I think it's important for them to get on it now.
I don't have strong feelings about whether it should take the form of sovereign wealth funds
that invest in the right supply chains or just subsidies to their own
citizens to buy a little bit of— This is actually a crucial point.
We were talking earlier about why the Rockefellers of the world, why their
descendants don't control everything, if our argument about the selection
of these greedy optimizers holds.

One argument is just that it's
very hard to index the economy. Maybe they would've just decided to have
their heirs index the economy and have their wealth grow at the rate of economic growth,
and their heirs would be trillionaires by now. Before index funds existed, it was just very hard.
A very small fraction of the economy, going back 100 years, accounts for
a majority of the value created now. If you missed those particular things,
your wealth would've just stagnated.

Maybe there was a brief golden window
from the creation of index funds up until five years ago where you could

actually index the economy and have your wealth grow at the rate the economy grows.
But now we're in this world with very concentrated returns, especially to private companies.
As we were making the point in our blog post, this is capital that the average person
has disproportionately less access to. Most of their capital is having a
random house, at least in the US.

Or a part of a house.
Which, as we were saying, is capital that is uniquely ill-suited to be
complementary to the production of AI or the serving of AI or to robots.
Or the kinds of goods that the rich will bid up the prices of.
Exactly. What is the value of a house currently? It's that the land is close to other humans and
modulo relational stuff that is just not going to be the main factor of production in the future.
This would be why a Georgist tax would not raise enough money for the sort of
programs that we will be discussing.

Right. But stepping back, the point I was trying
to make is, if it gets harder to index the economy now, and that is the main way in which
normal people are supposed to—modulo some sort of universal basic income—
In the developed world. —are supposed to have some
purchase on the wealth from AI. And it's also the way that developing
countries are supposed to have some purchase on the wealth gains from AI.
But it's very hard. Does Nigeria own a lot of SK Hynix and Anthropic?
I'm guessing not. It's not enough for them to just own the S&P 500.
This brings up a really important point.

Is AI going to be like
electricity or social media? Think about ConEd, or whatever
the electricity provider here is. It's a monopoly. It provides a
resource that everybody uses. But do we think about electricity as
creating a concentration of power? Does ConEd have this huge amount of political
power, social power, or something like that? No, because with electricity, a lot of
the downstream benefits actually came to the users of the electricity rather
than the actual entity producing it.

On the other hand, with social
media, it was the opposite case. Social media was everywhere. Everybody uses
social media, but the rents went to the platform. That's a really interesting point.
I don't endorse this take yet, I'm going to talk out loud.
The more you think our economy is going to be run on AGI the way our economy
currently runs on electricity—that is, there's a broad fundamental transformation
of the entire economy—the more it looks like electricity... Every company in the S&P of the

future, if it's going to make it to the S&P 500, it is because it has leveraged AI.
Exactly. And then you're indexed again.

But then again, I guess if you just look
at how concentrated the S&P is over time, just these big tech companies much more so... I guess this goes to a fundamental point that it's hard to reason about how much of the gains from AI these individual private companies will be able to control.
I think the open model thing is going to be a big point here.
If we're indeed in a world where the open models are six months behind the frontier—or nine months—then we'll hit AGI, we'll hit whatever, and in six months, everybody has access to this resource.

This goes to show you that every
question is connected to every other. That question about whether there's runaway gains connects to questions about recursive self-improvement.
Even if not recursive self-improvement, then continual learning, or online learning, which lets a model learn on the job. So if it's deployed, it gets to learn more.
These are just forecasting technical questions which then impact whether Uganda will have any purchase on the returns of AGI.

The reason I'm emphasizing the question
is I think both for the messy middle and for developing countries, a recommendation that is often made naively is that you've got to do some kind of retraining.
You've got to do some kind of jobs program, or you've got to have them
build data centers in your country. I think you guys are suggesting something closer to just buying the index of AGI. That's probably a much cleaner strategy and much more likely to succeed.

These are the two scenarios.
I think there is a world where it is concentrated, in which case it's
going to be really hard to index AGI. There is another world where it is electricity.
Basically every company has access to AGI. So you just buy the index.
Nigeria just needs to buy the index, and Nigeria has access to AGI because of the open models.
Just to get back to the question about whether to go with retraining or trying to index.
I would prioritize trying to index, just given how fast AI could hit the world.
But I definitely wouldn't just rely on that.

In the messy middle cases or the long-timeline
cases where we don't get anything like AGI all that soon, it would be leaving a lot of value on the table if you could have retrained to be a bit better educated on how to use the latest wave of computing. I don't think there's that

much of an either/or there. Maybe the reason to be pessimistic about this is because one of the reasons a country is poor is that it has a bad education system, so becoming the best in the world at retraining people at using AI doesn't seem like a particularly promising strategy for that poor country.

Although there are cases where, in developing countries, you had this leapfrogging effect with, for example, mobile banking. It's much more prevalent in Nigeria than it is in Germany. Everybody is doing mobile banking. They have it on their phones, and they're constantly doing this sort of thing. Again, I'm not putting probabilities on this, but with a transformative technology like AI, you could get leapfrogging where you skip the step in the middle and get really astronomical growth.

Just about the ease of indexing, I think it's definitely something to worry about a bit and keep an eye on. But as discussed in our own essay, and as other people have pointed out, it's already not that hard to index. There's been a bit of an increase in the privatization of returns, but still, well under 20% of the total market cap of non-tiny companies in the US is private.

Everyone thinks about OpenAI and Anthropic. If that's where all the wealth will accrue, then all these questions about whether open models will stay only a little bit behind, those are important. But even they look like they're going public before too long, probably. The frictions that have been keeping companies from going public might themselves be alleviated by AI a lot, just all of the disclosure requirements and whatnot. They want to get access to more potential investors, too. If I had to guess, I would guess that the long general trend of lowering those frictions and making it easier for more and more people to index will continue, despite the recent bump in the other direction.

This actually makes me hope even more so than before that the labs do get commoditized, or at the very least they go public as soon as possible. But hopefully they just get totally commoditized. I think AI will be much more popular and, more importantly, will be much more likely to lead to broad increases in prosperity if it is as hard to capture the gains of AI as it is to capture the gains of electrification. Exactly. There's no anti-electricity people out there. I mean electricity doesn't take your job, but— Well, it takes some people's jobs. It took some people's jobs, yeah.

This is maybe tangential to the

conversation but I think narratives matter. There's this really negative narrative around AI right now, but that's because people are not putting out the positive narrative. There's a reason. It's more difficult to imagine a good thing that doesn't exist than losing something that exists. It's much easier for somebody to go on a podcast and say, "These jobs that you like, they're going away," than for somebody to spin up a utopia which doesn't exist yet.

I hope this isn't too out of left field, but I would be remiss if I didn't point out one big cost of having commoditized frontier AI models, which is the tech race dynamic. For safety purposes, you might want fewer frontier companies so that each one has a buffer in case they want to slow things down to make things safer. The way this relates to our point before about the widespread access of the returns, is that I think there's a lot less of a trade-off there than some people imagine.

Some people think either frontier AI gets commoditized and we all enjoy the benefits—but there might be some risk, because the market's really competitive and cutthroat—or things are safer because there's a big gap between the leader and the laggard. But that means the leaders get fantastically wealthy? No. You could just have a relatively big gap, but it's a public company, and ownership in it is widely distributed. More recently, I have been thinking that the risk of commodification—which is that it diffuses the ability to use AI to harmful ends—is worth the benefit. I worry that having these concentrated labs not only makes it so that the surplus isn't as widely distributed through society, but also creates a very tangible, clear political target for the government.

We saw this with the Defense Production Act threat against Anthropic. If there wasn't one lab, or a couple of labs, that are clearly ahead of others, this kind of threat would be much harder to make. Thank you guys for doing this.
I feel like there's a lot of unresolved questions, but it is helpful to know what the first branch is along all these important dimensions. Great.
Thank you.