Hello everyone, welcome to our first livestream of 2026.
And, we're talking about, Q&A about future of science and technology, and I see a number of questions, saved up here.
Let me see… First, I have one from Gary.
For AI-generated videos and pictures, do you think there should be legal requirements, like, say, a watermark, to denote this is AI-generated content?
Yeah, I think that's a… a slippery and hopeless slope.
What is AI generated?
If you, were taking a photograph.
And you did red eye removal.
just, which people have done for forever and ever, well, particularly in the days when one actually had to use flash before there were high dynamic range CCD cameras and so on, you would get these, you know, people's eyes would look red because you're seeing the reflection from,
the visual purple on the retina coming back at the camera, and that was a sort of standard thing to remove red eye. You could argue that that's kind of an AI operation.
how do you then, if you then say, well, let's now start sort of blurring the face. Oh, let's make the face look a little better by, you know, changing the skin tone. Let's, try and, you know, make somebody look more lively, or whatever.
Or let's do this or that thing. At what point is it, sort of.
AI versus is it something which is kind of a… an actual reflection of the thing that was there?
Now, I suppose you could say, if it's a video, for example, you could say, was
the… was there sort of a… is there a real-time puppeteer who's driving the video? But even that's not going to work, because, you know, you could take the human face and map it onto, you know, a dragon or something, and then that's presumably AI-generated.
The fact that there was a human who was sort of moving around, and that was what was moving the dragon head around.
I don't know whether that even matters, or whether it was dragon head all the way down. So I think it's sort of a hopeless task to try and define even what you mean by AI-generated versus not.
I think there's, there's other interesting questions. I mean, like, for example, if you…
have something which is an AI version of a person, like a person's voice, or something like this.
Is there some sense in which… could you imagine some way of saying, if it is intended to be presenting somebody's, individual existence, does that
does that have to be signaled in some way? So, for example, that's a… there's sort of precedence, I suppose, in the law for things like that. For example, you would have something like,
you know, in a trademark or something. It's like, is this trademark that isn't quite
the name of this, but it's embellished, or the logo isn't quite the same thing, but there's a notion of it being sort of, being likelihood… likely to cause confusion. I could imagine a setup where one said, okay, if you have a thing that is kind of,
The likeness of some person, or could reasonably be confused by people as being that person.
then there might be some kinds of, requirements about how that's presented. You know, if you, if you have face-swapped,
you know, person A and person B, and there is a sort of likelihood of confusion, that what you're seeing is actually the person whose face was swapped on to somebody else.

that you could say, you know, that's a thing that has to be labeled in some way. I can see that being a thing that would be reasonable and enforceable and so on, and have a real definition. But just saying, is it AI or not? I think that's sort of a hopeless story. Now, you can also ask, if I see this thing, and it's, you know, a news report or something like this, do I know that this was real, or was it something that was just completely made up?

and generated by AI, or for that matter, generated by a special effects team using traditional, kind of, special effects techniques. You know, was it the real thing or not? And that then gets one into questions like.

Cameras that kind of record the geolocation and put some kind of signature onto the video they're generating, say, yes, this was really taken by this camera at this geolocation and this time. And one can imagine throwing some of that data onto a blockchain, for example, to be able to validate that, yes, it really was… well, at least you could validate

That this video was created, before the time when it was put on the blockchain.

It couldn't have been… because it was there on the blockchain at a certain time, in this ledger that's created on the blockchain, where you know this block was put on there at that time, and it couldn't have been… if somebody says, well, actually, that video wasn't made last week, it was just made today, you can say, well, actually.

You know, au contraire, it was already on the blockchain in that block that was in the blockchain from last week, so to speak.

So that's sort of one way to validate these kinds of things. You can certainly imagine some much more ornate forms of validation of, kind of, was it, is it…

really, sort of, from the original person. It's a kind of a… it's like these kinds of tests of liveness and so on that happen when you're trying to validate that you actually have a human being trying to, I don't know.

I don't know, validate their driving license or something for some app, and the thing will say, you know, look left, look right, whatever else. And it's making use of the fact that you actually are performing that action, and you actually seem to have sort of a three-dimensional structure. I mean, there are also plenty of things you can do, so you can imagine things where

I mean, the question is, if you were just presented with a video, can you tell if it was a video that was genuinely a representation of what it said it was a representation of, or was it something that had sort of AI in the middle, or for that matter, AI all the way down? And I think that's…

a… I mean, that's… without, sort of, external metadata, that's a challenging thing to establish, I think.

So, a few thoughts about that. I think that, this whole question of, sort of, can you tell

First question is, what do you mean by AI-generated?

For text, for example, it tends to be a lot more obvious, because it's…

you know, you give a prompt, and then the AI generates lots of text. And it's not like that text was just a small tweak. I mean, I don't think one would expect that if a spell checker was used that uses AI in some sense, you wouldn't

that is a much smaller change to a piece of text than typical video redressing, restyling is to video. So I think it's easier to have a notion of, was it AI-made text or not?

Because you probably can exclude those minor, sort of… you… I suppose the criterion would be, if a human reading the text would just say, that person can't spell, but I understand what they were trying to say.

Versus, in the case of a video, oh, that's person A speaking rather than person B, and I'm confused, type thing.

So… Anyway, a few, few thoughts on that, on that topic.

Let's see…

Well, Brady is asking, do you think AI data centers should soon be in orbit and solar-powered? What if we add superconductors to the computation in the cold of space?

I mean, I think… the, It is…

I don't know the current calculations, but I think it will certainly, as the cost of sending things into orbit goes lower, the, you know, it becomes an easier thing to just say, stick all those things in orbit.

There's a question, you know, when sort of limited by the speed of light.

the one has to deal with the transmission delay of getting things to orbit, and sort of the way to avoid that is to have so many things in orbit, in low Earth orbit, that there's always something more or less straight ahead… up ahead, and it's only, you know.

I don't know, 100 miles or something to it, as opposed to something where there's a… where you're looking close to the horizon and there's a big slant distance to get to the thing.

So, you know, that's one limitation, is that to be in orbit, you have to be above the atmosphere, and that has a minimum distance. It's a lot smaller of a distance if there's so many things orbiting around.

that there's something sort of above you, directly above you at all times. Needless to say, if you were actually having, it will be a kind of a strange thing, there are plenty of services which are stateless in the sense that, you know, you… you go to Wolfram Alpha, you ask a question.

Wolf Malpha computes the answer.

The answer comes back.

then there's nothing maintained by Wolfram Alpha about that question.

it's ready for the next question immediately. But if, for example, you were doing something like editing a document.

then the server that is providing that document has to maintain the state of the document between the time when you press one key and when you press another key. And if the satellite that you were talking to had sort of gone out of range between the time when you pressed one key and when you press another key, that's a tricky thing. Then, so what you would then have to do is hand off

from one satellite to another, that's something which is readily done by lots of things like Starlink and so on, even by old-fashioned sat phones, had kind of handoffs, so you could typically hear the handoffs, as, as,

Different satellites went below the horizon, and so on.

So you'd have to have some kind of somewhat elaborate software infrastructure to be able to say, yup, this person who was communicating with satellite 1 was editing this page, now satellite 2 is going to go into range, and you have to transfer the data

to Satellite 2, to be able to go on editing that page. I mean, I suppose, to some extent.

So long as the satellite is not below the horizon for the other satellite, well, still, you have… you have issues of latency as you start having to sort of send a radio signal, sort of a significant fraction of the way around the Earth, so to speak, to get to the other satellite.

But… You know, I do think that the economics of launching things into Earth orbit

will, people say, get even better if you can launch them from the moon.

Because the Moon has one-sixth the gravity of the Earth, and it has no atmosphere, it's a lot easier to launch a rocket from there, and to send its payload into Earth orbit than it is to launch something from the surface of the Earth to get it into Earth orbit. So that requires that you have to be able to somehow build the satellites on the Moon, and that's a whole giant adventure of infrastructure to make that possible, but presumably

Eventually it will be, and so it'll be pretty easy to throw things into Earth orbit.

Again, I don't completely know the economics of doing that, but my guess is that there'll come a time when it makes sense to have servers there, and certainly powering them from photovoltaics, I think, makes sense. I don't think that…

the servers of the Earth, it's an interesting calculation, we could probably try and estimate it right now. But, if you wanted to power all the servers of the Earth from

from, photovoltaics, solar cells in Earth orbit, how big an area of solar cells do you need, and is that area at all significant when it comes to, sort of, the, the, the whole sky, so to speak? I mean, will we find that, oh, we are, will we, we,

You know, we put all these things into Earth orbit, and we're powering all this stuff there, and oops, it's, it's turning out that we're blotting out the sun some part of the time because, because we're doing that. I mean, it's a… if you want to cool the planet, it's a good way to do that.

But, I think, probably not something people would generally want.

In terms of superconductors,

the, certainly, one of the issues, you know, why does it take so much power to run servers? Well, it's because they display a lot of energy, in the way that they're currently, in the way the physics of semiconductor devices is currently set up.

If you were using superconducting devices, you could presumably dissipate a lot less energy.

I think that, there was a big enthusiasm for sort of making superconducting computers back in the 1980s, using Joseph's injunctions and other such things. It didn't really work out that well at that time.

It's sort of come back again because one of the approaches to quantum computing involves using superconducting devices, and the very same companies, particularly IBM, have jumped into doing that using the technology that they had originally developed in the 1980s.

I don't know the… the, level of reality of getting fast.

electronics of the kind we have with semiconductors, with superconducting components. My guess is there's a pretty long technology development path

that you need to go through to get those kinds of devices to, sort of, semiconductor, current semiconductor level. In fact, we kind of know that, because we know that quantum computers based on superconductors

have a hard time getting, you know, it's a… it's a big deal if you can get 500 qubits. And in a microprocessor, you might have a billion, gates on, on your… on your chip, so…

That's the time when there were only 500 gates on a trip. Might have been sometime in the beginning of the 1970s for semiconductors, and so you've got a solid 50 years of development.

at least in the semiconductor case, to get to the point where you can have smaller components and so on. So, I think it's a kind of a long,

A long story.

Let's see… Trich is asking, what's something you're excited about in the future that's unrelated to AI?

Well, there are lots of things.

Lots of things in basic science.

that I think, well, I've been working on a lot, where it feels like

There was a lot of basic science that got done about 100 years ago, and then in physics and mathematics and biology and in other areas that then kind of got stuck

for lack of, kind of, the right formalism for thinking about things. I think we finally have that formalism, and I'm certainly having a great time, kind of, steaming forward, trying to make, sort of, foundational progress in lots of fields of science.
I think that's going to be pretty… I mean, I think it already is pretty exciting. I think it's not as widely understood as it could be yet. It will be eventually, presumably. I think that's going to seem like one of the big things that got figured out at, sort of, this… this
Period in history. So that's… that's one thing in basic science. A lot of, kind of, foundational questions about, sort of, what's underneath physics, what really is mathematics, how does biology really work?
These are things that I think we are sort of in the years where that's going to… a lot is going to get figured out along those lines.
Another kind of general area has to do with biomedicine. It's not unrelated to these questions about the foundational science of biology.
But there are things that are sort of separate from that and are more kind of engineering kinds of things. My guess is that there will be… there's a fair amount of just sort of decoding biology and replacing biology with
With technology in ways that will have, sort of, dramatic effects in biomedicine.
Biomedicine is always kind of an area where there's… everything has a footnote, and a footnote to a footnote. There's always all sorts of details that you didn't sort of know up front.
It's a little bit easier, often, if you're outright replacing things, than if you're trying to, sort of. reverse engineer and fix what's already there. You know, if you're printing a new organ and inserting it, rather than trying to fix an organ that's already there, or if you're doing something where you create a new line of cells or something that you can make use of, rather than having to sort of fix things that have gone wrong with cells that are already there.
I think there's, there's lots of things where
Where one tries something, and there's a kind of gotcha, there's a footnote to a footnote to a footnote, which is a horrible gotcha, where you didn't know, oh, but that's fine, but it interacts poorly with, you know, some
something about some piece of the innate immune system, or something like this, and then it doesn't work. Even though the technology stack looked great, there's some kind of awkwardness in plugging it in. But I'm guessing that there will be some things that will really be
sort of substantial improvements. It's worth understanding that…
what tends to happen in the development of science, technology, etc, there's sort of two kinds of things that happen. One is breakthroughs, and the other is sort of systematic incremental development.
In engineering, we were just talking about superconductors and semiconductors and so on. In engineering, there's often just sort of systematic, kind of progressive improvement.
And that can go on for 50 years, 100 years, just things getting more efficient, faster, smaller, easier to manufacture, cheaper, whatever else.
And it's incremental engineering. It's lots of details, lots of clever ideas that all build this sort of tower of capability.
And we've seen that in lots of different engineering domains.
Then, there's things which is a bit more common in science, where there's just an outright breakthrough. There's a methodology one didn't have before, now one has it, now one can figure something out.

It's an interesting question in biomedicine, to what extent it's about breakthroughs, and to what extent it's about incremental development. You know, if you don't know that germs exist. Finding out that germs exist and acting accordingly is sort of a breakthrough kind of thing. Knowing that, you know, DNA stores digital… stores genomic information digitally. That's kind of a breakthrough that leads to many kinds of things. But sort of being able to sequence a piece of DNA efficiently, that's more of an engineering story. And what can happen with these engineering stories is that things that at first

At first, it's a sort of quantitative difference. Yes, it goes faster, yes, it's cheaper.

But eventually, it becomes a qualitative story, because it's like, well, okay, there's this point at which you can have a sophisticated computer in a thing that's battery-powered and small enough to hold in your hand.

And then you have the possibility of a smartphone. And, you know, while it was still just, oh, there's a computer, and you can have it, but it's the size of a desk.

you don't even have the idea of having something like a smartphone. And I think in biomedicine, there is… it's a case where, you know, whereas technology, a lot of it is incremental engineering development.

that then eventually passes thresholds that allow a new, sort of, ideas about use cases to emerge. That's one thing. And in science, I think it tends to be much more driven by, sort of.

sort of point breakthroughs. In biomedicine, I think it's a mixture of those two things. And it's sort of interesting to see how that progresses. And I think we can expect, in both dimensions, that kind of thing. So, for example, let's take genome sequencing.

you know, right now it costs a few hundred dollars, can take a few weeks, it involves lots of computation, et cetera, et cetera, et cetera. What if you could do it… what if you could do it for a dollar?

whole, you know, human genome. What if you could do it on a device the size of a smartphone without, you know, what would you then do?

you would then start saying, well, you know, I'm going to… I'm eating this piece of food. You know, what is this food? You know, where did that… where did that piece of salmon actually come from? Let me… let me have my, you know, pocket sequencer and just figure that out.

Or let me, take, you know, something which is continuously measuring, you know, some continuous device that's measuring my bloodstream, and, looking at different,

looking at, I don't know, the antibodies that are floating by, and sequencing those antibodies, and saying, you know, what can we conclude from that about… insofar as the immune system is kind of surveilling your whole body, what, what can we conclude from the fact that, oh, there's a jump in this type, particular kind of antibody, or something?

These are things where, sort of, the engineering development of those kinds of things

could lead one to a really qualitative change in the way that one thinks about how to do, for example, medical kinds of things. I mean, I do think that the tendency in medicine will be you measure

Just tons and tons of things. And you know, sort of, directly from the measurement what you should do.

rather than going from, oh, I measure, you know, this thing, and is it a yes or a no in the, you know, do you have…

you know, elevated level of blah, or whatever else. Oh, if you have elevated level of blah, then there's this diagnosis, and it's one of 100,000 possible diagnoses, and given that diagnosis, then

there's this treatment. I think it's much more that one will go from, oh, here's a gigabyte of data, and then from that data, here's sort of the optimal thing to do next.
you know, that's a thing once seen in endless kinds of systems, whether it's autopilots, things like this, where it's like, ingest lots and lots of data and decide what to do, rather than say, is it an A or a B, and then decide what to do after you've gone through this kind of thin.
Level of, of, kind of, sort of, diagnostic, representation.
Let's see, other things that, I'm kind of, expecting, well… on…
On the technology side, one of the things that's always interesting with technology is there are ideas that have been had a long time ago, but they just weren't terribly practical, and they don't get used. I mean, a classic example is kind of virtual reality, augmented reality.
You know, there's some use of that, but the whole, it's going to be really big, and everybody's going to be using automatic reality all the time, hasn't yet happened.
Presumably, it eventually will. That will happen as a result of, sort of, incremental, progressive, incremental engineering improvement.
The,
the thing that, and when that happens, there'll be a bunch of, sort of, new kinds of things that become possible. I mean, it's often very hard to predict how that's going to play out.
You know, for example, I don't know, something like social networks, enabled by the fact that you could have, you know, images on a computer where it was connected to a global network, etc, etc, etc.
You know, it wasn't obvious from, oh, we're going to connect it to this global network, that there would be this use case, which, arguably is a use case which is sort of a reflection of kind of general human tendencies that have existed from, you know, your average village on, so to speak.
But joining those dots and seeing what's going to happen, it's often very hard to know… it's very hard to know what it's going to feel like when a particular kind of technology gets to a certain level of usability. I mean, I notice this all the time in building Wolfram Language and our other technology, that
There are things where we think about building that technology, and then once it's built, One realizes, gosh.
With this technology, this thing is now possible, which I had never even really imagined beforehand.
I mean, we have seen that a bunch with AI and LLMs, that there are things where one just didn't sort of think that that will be possible, to sort of, I don't know, write some document for some compliance purpose.
Where it was just sort of all synthesized by AI from a small number of points, or something like this. That's something one just didn't really imagine that was possible. One didn't really think through those use cases, because it just didn't seem that that was going to be possible.
So, those are a few things, at least.
Let's see…
Samian is commenting.
Biology will be…
will be good, big time, yeah? Longevity, making biology programmable, understanding the origins of life and how it could change.

organelles could potentially be replaced. Yes, all of those things are fine directions. I mean, I think, you know, the big thing that, well, at least ancient people like me are big on is longevity. How do you,
how do you increase it? Well, I think, there are sort of the… the big picture, I would say, is, you know, can you increase longevity? Can you increase longevity even to the point where
Sort of immortality is there, and you're sort of swapping out components, or you're… or you're, sort of making
things better in place. There's… there's kind of hit the pause button, use cryonics, get frozen for a while until you can, figure out how to do things better. That doesn't work yet.
But…
That's the kind of thing where I can imagine that there might actually be just a straight breakthrough. I mean, that's a kind of place where, like mammalian cloning, for example, where people had said for years, I'd ask people for ages, you know, when one can clone some kinds of amphibians and so on, why can't one clone mammals? Oh, because mammals are more complicated, blah blah blah.
Was there a fundamental reason? No, there wasn't.
And in the end, it was figured out how to do it, and it was kind of a weird set of things to be done. Just like, you know, was it possible to get stem cells?
Not to reprogram a cell to be a stem cell, to reprogram, you know, a skin cell or something to be a cell that could differentiate into any other kind of cell, just like, sort of, the original egg cells that us mammals come from, so to speak.
And it was, again, like, oh, that's not going to be possible. And then it was figured out that, yes, you just sort of apply these particular chemicals.
And you can reprogram a cell, more or less, I mean, there may be some glitches to that, but more or less you can reprogram it to be a stem cell. So there will be breakthroughs like that, and I think cryonics is one that is very much, sort of, crying out for a breakthrough. I think it's one of these places where not enough work has been done on it to even know whether it… how hard it is. And it's a place where I'm expecting there'll just be a moment
Well, there's a breakthrough in that.
I don't think that the kind of elixir of eternal life
type breakthrough, you know, just take this pill every morning and you'll live forever type thing. I don't think that's going to work. I think biology, the fundamental nature of biology is such that it's not sort of a quick fix like that.
I mean, it could be a quick fix if it turned out that our limited lifespan is an evolutionarily determined thing.
That is, that one doesn't live forever, not because it isn't possible, given the components we have, but because it was bad for the species to have all those old fogies hanging around forever.
And that it was sort of better for the species to let the young'uns come in and do whatever they're going to do, and that there was sort of almost an evolutionary switch
that said, you know, when your time is up, you should go to make room for the young'un, so to speak. It is conceivable that there is such a thing. I don't think that's likely to be the case. I think that what's happened, as so often happens in biology, maybe at one time it was evolutionarily kind of beneficial to have the old fogies go, you know, disappear.
But then, that fact…
If that was, in fact, a thing that was evolutionarily desirable, that fact got tons of things built on top of it.

To the point where it's like, oh, well, if you're only going to live for a finite time, you don't need this and that and the other, let's economize on this feature and that feature, and pretty soon the whole organism is built with that assumption. I mean, it's like… something like sleep.
Which, you know, presumably, its prime function, presumably, is to clear out the gunk that happens in brains as they operate during the day.
But given that you're gonna shut the organism down, more or less, for some number of hours every day, it's like, well, let's pile a whole bunch of other functionality on top of that. That's a typical kind of thing that biology does.
And so if you say, well, let's have this other way of cleaning out whatever it is, whatever the gunk is that develops in brains from all those nerve firings and so on, the, you know, we've got another way to do that.
Let's say… It's still not going to… that's not going to mean you don't need any sleep.
Because biology has just piled, you know, a dozen other pieces of functionality on top of the fact that sleep is there, so to speak. So I fully expect the same thing to be true in the case of longevity, and it's, I think it's one of these things where
Where even if you could unswitch the switch, you'd have all kinds of other things you have to deal with.
I think, though, that… and the whole setup is probably such that the components just end up sort of… if some component is wearing out, another one's going to wear out, too, because why wouldn't it? There's no point in making that one last forever if other critical components don't last forever.
And I think some of the things I've done in trying to understand, kind of, the foundations of biology and kind of what it is that biological systems do. I mean, biological systems are a unique example of, kind of, molecular-scale computation that actually works.
Where, sort of, there's a whole orchestration. There's kind of a bulk orchestration of all these molecular processes, of all these different proteins that are moving around and fitting into these other crevices and other proteins, and doing these things which sort of microscopically represent mechanisms, but they represent this whole pile of different mechanisms all put together, kind of orchestrated in this way that makes
Makes Living Systems. And how you, kind of.
fix things that have gone awry in that whole elaborate orchestration is not clear. I mean, I tried to study that a bit, last year, the sort of computational foundations of medicine, and what's really involved in kind of, once you have this kind of evolved
Mechanisms, evolved organism that has,
that has been sort of successfully evolved to do what it does, typically in a very complicated way, and you poke it in some way, how do you sort of fix the poke? How do you… how do you restore that very complicated way that the thing actually did what it was going to do?
Simeon here also mentioned, things like organelles being replaced. I think that's a… looks like a very promising direction. I mean, mitochondria being one of the… one of the prime examples. I mean, we got a lot of mitochondria in us, and they're pretty important, and as we go through life. the genomes, which are separate from our main nuclear genomes, progressively degrade. I say they're separate from nuclear genomes, but as is always the case in biology.
It's full of footnotes. A mitochondrion has a bunch of proteins in it, and you would think the mitochondrial DNA, which is quite small, would contain the proteins that a mitochondria needs, but you'd be wrong.

You know, some point in the very distant past, a couple of billion years ago, mitochondria were probably separate organisms, and

They, at some moment, a mitochondrion got ingested into a cell, and then it became a thing that was a part of a cell, and it's a very critical part of a cell.

For, for us, kinds of, for… it's… for the eukaryotes, at least, it's a, it's a… it's the source of energy that we have in cells.

And, but it's an independent, in a sense, it's an… it seems like it's an independent, quotes, parasitic organism or something, symbiotic organism.

Because it, you know, mitochondria replicate separately, but

Mitochondria also use proteins that come from nuclear DNA, so they're not really completely separate. Given that they've been hanging out with us for 2 billion years, or whatever it is, they have sort of interoperated in all kinds of ways.

But nevertheless, It's probably the case that if you could, sort of.

get your mitochondria and restore, kind of, perfect genetics to your mitochondria, they'd be more vigorous than ever before. What's become clear in the last few years is that mitochondria do migrate from one place to another through the bloodstream, and so it does suggest that you just, you know, you take a mitochondrion shot

And then your mitochondria will… those mitochondria will migrate to tissues and organs and so on that have need for energy at that moment in time.

That may be a thing that's sort of, what can I say, end-user serviceable, so to speak. I mean, it's just like when you have a machine, you know, sometimes there are no user serviceable parts. You have to send it back to the factory if you want to get anything fixed. But… and… and, you know, with

With us humans, the kind of, unfortunately, the only kind of send-it-back-to-the-factory thing we have is, you know, have kids and pass on your genetic material and go to the next generation.

Because we don't have a, that's kind of the send-it-back-to-the-factory scheme.

But, you know, then we also have the question, is there stuff that is sort of end-user, you know, serviceable? And if it is the case that you can just inject more mitochondria and have them go to the right places, that's a pretty convenient thing. I mean, it was hoped that

You could have stem cells, or differentiated stem cells that were differentiated into, for example, heart muscle cells or something. Just inject those, and they would somehow magically end up going to the right place and integrating into, let's say, heart muscle.

My impression is that has not worked. That was a thing people thought might work maybe a decade ago, and it hasn't been, so far as I know, has not been as successful as was hoped.

But, you know, that's…

that's the question of what can you replace, and so on. And I don't know, I think the mitochondria are probably the highest kind of value target in terms of organelles. I mean, I don't think

I think most of the organelles, first of all, they don't independently replicate, well, except for chloroplasts and plants, but, you know, we don't have chloroplasts. You know, there's always the question of, can you, like, just start putting other stuff into our cells.

My guess is that's a pretty difficult road. For plants, it's an easier road. You can start putting even additional DNA into plants and so on, and plants somehow seem to respond okay to that.

And even will produce, sort of, will be bigger, stronger, and so on, if they have more… if they… if you put, you know, for us.

I don't think it works to just add more chromosomes and just say, okay, there's more places to get genes from, to get proteins from. In plants, it does seem to work.

up to a point, you can just add in more genetic material, and the plant will just make more proteins, and it'll be a bigger, juicier piece of fruit or whatever as a result of that. But I don't think that works for us, us animals.

So… That's a, that's on that story.

there's the question about making biology programmable. I think, you know, biology sort of is programmable itself. The question is, can we go in and modify it? And one way we do that is with various kinds of,

gene editing and so on, where we say, okay, you know, we've got… we've got this particular piece of DNA that is or isn't making this protein in the right way, let's go in and send in some gene editing package

that will go and snip our DNA in the right place, and make a change.

Now, again, one of the things that is just a bizarre thing in biology is that

biology really uses combinatorics in a very serious way, in the sense that there's just tons of things in biology where you say, I've got this sequence of, you know.

12 base pairs or something, and there's this thing, maybe it's an immune system cell, maybe it's some DNA repair mechanism or something that's got this thing which is exposing some sequence of base pairs, and the idea is that

that will just manage to find its way to a complementary sequence of, let's say, 12 base pairs. And even though there are billions of base pairs in DNA, somehow it will bubble around and knock around, and within some number of minutes, it'll find its match.

It's kind of bizarre that that thing like that works, but that's a lot of what biology is based on, is that kind of combinatorial matching mechanism.

And so, it is not implausible to say that you can go and just go edit that piece of DNA, because you can just, like, find that piece, and that there are lots of… lots of techniques these days where you're kind of finding that piece of DNA using sort of flanking sequences that have a particular form.

And then say, well, okay, now we're gonna make a change to the DNA.

Right now, the changes that can be made by CRISPR and so on are rather limited. It's just, like, snip out this base pair, things like this. One imagines that, sort of, the engineering will get better, and it will become possible to actually say, let's really make an edit. Let's change, you know, this sequence of five base pairs to this other sequence of base pairs.

And that's… that will be, you know, if that really works, that will be a big deal, because there are lots of things where it's like, oh, you have a genetic predisposition to this or that, let's just edit your DNA.

Now, whether you'll be able to get into the cells that are the most important ones, I'm not sure. If it's things for… that produce tumors and so on, there's a chance you will be able to, because the ones that produce tumors tend to be cells that are replicating a bunch, and those are ones where you're more likely to sort of have access to them than cells like nerve cells that have just been sitting there for a long time.

without replication, and presumably, they're not set up to have, kind of, genomic changes made. I'm not sure, but I think that's a difficult thing to do. I mean, you can go, you know, viruses do manage to inject themselves into cells.

And that's a path for doing this.

But, I'm not sure how well, you know, I think that's more difficult.

The other thing that one can make programmable is things that we build that kind of interact with biology. So, for example, drugs right now, there are only a few thousand drugs that are known, and they're all molecules that

Essentially operate, in essentially all cases, operate by finding some shape correspondence in some protein or something

in us, and fitting into that shape, like a kind of a lock and key. It's just like, this thing fits into that, binds to that active site in that protein, and it does that because it's this particular shape

And it manages to stick itself in there. There's increasing knowledge that proteins are quite floppy, and it's not really just a question of it's this shape or that shape. There's a more dynamic story that is only partly being understood at this point. But that's been sort of the traditional thing, is do you fit? Does the drug fit?

In this particular molecular crevice, so to speak, in this molecule.

Well, you can imagine much more elaborate setups where the molecule that is doing that docking has its own little computer on board.

And, in fact, this happens in biological systems in there are plenty of really very elaborate, sort of, computational-like devices, whether it's in ribosomes or spliceosomes, or other kinds of, other kinds of devices that are constructed from proteins, and that do identifiable things in cells. I mean, more and more of those are getting discovered, and they are things where, in a sense, they compute as well as just fitting things.

And we haven't built drugs yet.

that have the sophistication of a ribosome or a spliceosome or something like this, but that's something one can imagine doing. And, you know, the question is, it's sort of a competition, then, between biological evolution

Sort of discovering how to do these things, and us sitting down with computer-aided design, so to speak, and saying, we're going to design a ribosome.

So far, at the molecular level, biology and biological evolution has absolutely won that race so far. So far, you know, with all the engineering cleverness in the world, we can't build things like that yet.

I suspect we will be able to at some point. I think that,

you know, we have to compare. There are about 10 to the 40th organisms that have lived in the history of life on Earth.

That's a big number, but the process of biological evolution is incredibly inefficient, and partly because what matters is the survival of the whole organism, but if what you're trying to modify is something down at the level of the operation of an individual ribosome or something, ribosomes take, take genetic code and convert that into proteins. That's just one of the things that's needed in, sort of, the machinery of life.

The, it's… if you're trying to design a ribosome, then sort of knowing whether the whole organism has more children or not.

is a very long lever arm relative to the properties of the ribosome, and it seems much more likely that if one is at the point of being able to say, I've got this

giant CAD drawing of my artificial ribosome, and I can compute this or that thing about some piece of it, and do some simulation of this subsystem, whatever else, that you'll be able to more efficiently create it. Now, there's a lot of work that's going on that's sort of very AI-themed these days, it's been differently themed in the past.

of create a protein that, that will have this or that property, that will have this or that shape, that will have this or that behavior, that will have this or that bind in this or that way. That's a thing that's very much today's kind of drug discovery story.
One has to be a little bit cynical about this, because there have been 5 previous, sort of, there's going to be a magic way to design drug stories, and they haven't worked out in the past, at least on a large scale. There have been particular successes, but on a broad scale, it hasn't worked out.
I think one of the challenges when you're doing these kinds of things is, well, okay, let's say you want to make a molecule that does something, what do you want that something to be? In other words, it's like, you know, imagine that we could search through programs to find one that will do a thing we want.
We have to know what it is that we want before we start to search through programs to find one that does that.
And I think in biology, we have rather little understanding of sort of the big picture of sort of this bulk orchestration of biological activity, and we might say, well, we know this one particular mechanism, this one particular circuit, and we can do something about that.
But that doesn't really fit into the whole story, and it may be difficult to fit into the whole story.
Anyway, a few thoughts about,
Future of biology there.
Jh is asking, could biology ever make humans photosynthesize? I think yes.
you know, I'd always been waiting for the moment
When there would be a fashion of getting jellyfish protein kind of integrated into your genome, maybe for your children, maybe for you, if you have some retrovirus going and reinstalling something, or some gene editing package doing that.
The, you know, could we make glow-in-the-dark humans? I think, yes, you could make glow-in-the-dark frogs and so on, I don't… zebrafish and whatever else, even plants. I don't think there's a reason you can't make glow-in-the-dark humans.
Now the question is, what about…
using photosynthesis and chloroplasts to, to generate energy. It's an interesting thought. I'm, I'm, I mean, plants…
have leaves, which, as you may notice, are quite thin. In other words, they have a big area where they have, you know, where they do photosynthesis, and there's not lots of inside of the leaf there. You know, there's a… there's the inside of the trunk, or whatever, the inside of the stem of the plant, but then the actual leaf
It's very, it's kind of large surface area for small volume.
For us, we're big blobs, and there's a lot of inside relative to the surface. So we're not as well set up to do photosynthesis, unless you also want the humans to grow leaves and so on. We're not so well set up to do it.
I think that, you know, what's the maximum energy we could produce by being green? You know, we'd be, like, you know, there'd be a fashion for, you know.
Don't,
everybody going around, you know, shirtless or whatever, because they're all green, and, they're, you know, ingesting green photons from the sun, and, generating energy that way, and it's like, I don't need to eat as much, I'm getting,
And getting my energy directly from the sun. It's always interesting to see how these kind of fashions about things might develop. I'd always thought

if, well, if augmented reality really comes in big time, everybody's going to want to wear glasses, unless maybe there's some engineering solution where you have, you know, contact lens AR or something, but that's yet another level of challenge.

And I think, so, you know, the fashion, it seems like, people are like, you know, for lots of people, it's like wearing glasses is a shocking and terrible thing to do,

And, so… but it'll be… if augmented reality is delivered that way, it'll be back. I mean, I think…

The same thing is true for all kinds of people wearing, you know, the fitness tracking ring, or the fitness tracking, you know, device on their wrist, or whatever. There are all these things where… where sort of utility overrules fashion. So I don't know, I… I don't think that will, necessarily happen for the… for the green people anytime soon.

Let's see…

Samson is asking, do you think intelligence itself is a biological process, or a computational one that happens to run on biology? Oh, most definitely a computational one that happens to run on biology. But the computation we can do in biology

Well, I think one of the things we've learned from LLMs and the whole sort of modern AI story is that a lot of things that we thought might require some special molecular-scale biological thing that we didn't understand yet is just computation that you can run on a digital computer.

The fact is, we have kind of a lot of molecular-scale computation going on in brains.

There are things that brains can do that, you know, and that's…

so far can't do, like… like learn at the same time as they are running, things like that. They also have very good power consumption characteristics, all kinds of things like that. They're also incredibly slow, relative brains are, relative to digital electronics. But I think that what we've learned

Is that, sort of, the… the…

phenomenon of intelligence is really a disembodied, abstract computational phenomenon. It's really the story of souls all over again, that there is a disembodied thing that is the essence of minds, independent of the biological substrate on which they're running.

what that means for, sort of, the human experience and the future of it is a whole different issue, and whether, sort of, the inner experience, you know, if you were just a simulation of you in digital hardware, how would your inner experience differ from the inner experience that you have with biological… in a biological form? My guess is that at the level of the mind, the inner experience would not differ.

There's questions about, sort of, what consequences does that have for the way that the mind lives its life?

And that will be very different, you know, if you're an immortal soul that's disembodied in digital electronics, there are lots of different priorities you might have from ones if you're a finite biological, mortality-ridden kind of thing.

Let's see… JH is asking, is it possible that diseases could be debugged instead of treated?

Well, I mean, that is the story of lots of, kind of, high-end medical diagnosis now, in the sense that, it's…

You know, sort of a treatment in the most basic sense would be just like, you know, take this pill and call in the morning type thing.

You know, there's an awful lot more, particularly in areas like cancer and so on, there's an awful lot more, and also, I suppose, to some degree in infections and so on, although that's not developed quite as far.

Of, like, well, you know, there's this thing, and it's going to mutate in this way, and we predict it's going to do this, and we watch what it's doing, and we kind of block it by going ahead of it and putting in some kind of agent that will block that particular pathway that it might follow, etc, etc, etc.

I mean, it feels more like kind of,

it feels more like knowing more about the mechanism inside. I suppose that's the distinction between, you know, what is treatment for a computer?

Versus what is debugging for a computer. I suppose debugging kind of has some implication that you understand the mechanism of what went wrong, and you can go and fix things at the level of mechanism, whereas treatment is just like, let's, let's bypass that piece of code and make it work differently, so to speak.

So I think that,

there's sort of in… okay, it's an interesting question, to what extent we can actually find mechanisms for things, and to what extent we just have to say, it's like a computation that's running, and here's the result it gets. So that we really don't have a choice. We can't go into the… there isn't a mechanism that we can go in and say, let's reverse engineer this or that piece. We really just have to, sort of, bypass the whole thing.

And say, let's use, you know, an artificial, you know, kidney or something like this, rather than trying to fix the issues with the actual kidney, and so on.

Let's see… Gosh, lots of questions here. Hunter's asking,

Well, a couple of questions here. Will AI ever be credited as a co-author on scientific papers?

That's already started to happen, for better or worse, meaningless or not. I mean, you know, it's sort of a shame

Well, people have done this to some extent. You know, our Mathematica, more from language technology stack has been used as a very core element in lots of kinds of research for nearly 40 years.

And there have been moments when people sort of said, I'm going to put it on as a co-author, because, in a sense, it discovered a lot of what is in this paper. It's, you know, I gave it the seed, and it went and searched a trillion cases and found all these kinds of things. I mean, I've done that a lot.

You know, what does it mean to be a co-author?

And, you know, in what sense, you know, do you just say, okay, if you're using a tool, like you're using Wolfman Language, or you're using an AI, those are both examples of tools, you know, to what extent do you mention that in writing your paper?

It's something where, you know, people use Wolfram language all the time, and it's a little disappointing for us, but they don't mention it.

It's just, that was a tool I used, and I got an answer. People might use, you know, other tools. They might use a spell checker. They might use a, these days, they might use an LLM to, you know, find something in the literature, or something like that. They might use, sort of, a citation-searching thing.

You know, do they mention that in their paper? Do they, first, do they mention it in the, oh, I use this tool, and mention that at the end of what they're writing?

In some fields, there is a tradition of you should mention every tool you use. That's true in a lot of biomedical life sciences-type fields. It's not true in the physical sciences, typically. That… and it's not true in more abstract areas, like, like, like mathematics and so on.

The,

There's… there's then a, you know, there's… there's different levels. There's what do you cite as a… what do you say you used as a tool?
There's what do you cite as kind of an academic reference, and there's what co-authors do you put on your paper. Each of these things has different sort of practical and social consequences.
I mean, citations sort of became, starting in the 1960s and 70s, they became this giant sort of network of connectivity in science that is both useful and sometimes damaging.
I mean, I think that the sloppiness of citations is incredible. I mean, I know it's kind of amusing, there are some things… I've written, some major things I've written, which
Where somehow, in some citation database, completely wrong citations to these things got generated, and you can kind of see the epidemic spreading through the scientific literature, as people just copy the citation from somewhere else, not even bothering to figure out whether it was correct.
You know, what's the role of those things? For myself.
I think an important thing in science is to contextualize ideas by explaining where they came from.
But just saying, reference 7, blah, I never read it, you're not going to read it either, reference 8, blah blah blah, that's really a lazy way
to answer the question, where did this science come from? How does it fit in? And what I've tried to do for the last, I don't know, 30 years or so.
is really figure out the history. Like, these ideas, they were the evolution of this thing that so-and-so did 40 years ago, or whatever, or 2 years ago, and this is how they are now being incorporated here.
And to kind of be able to write the narrative history of what happened. I've also tended, in recent times particularly, to write the narrative history of how I figured out what I've written.
Because I think that's useful. When I try to understand what other people have done, it's like something that just sort of arrives as a kind of deus ex machina of, you know, here's the result. It's like, where did this come from? I have no idea.
You know, back in the 1970s, when I was a kid and doing physics and publishing academic papers and so on, one of the things that I did was I sort of had discovered this secret weapon of using computers to do things like mathematical computations, and I would regularly, in my papers, have these incredibly complicated formulas.
without any comment at all, I would just go, here's the formula.
It wasn't, you know, I didn't have the thought at that time of, I should explain this was made by a computer, and this, et cetera, et cetera, et cetera. And so people were very confused by, you know, how did you possibly get this result?
You know, because it just… it just seems to sort of deliver… it dropped in there, you know, it was a… it was just a… it fell off a tree or something and showed up in the paper.
the,
Anyway, I think then there's the question of, sort of, so there's this kind of network of connectivity of things, and there's some sort of structure of that network from academic citations.
I think it's much more useful when you're reading a document to be able to see the narrative history. I think that's much more valuable to the actual reader of a document, if… but the sort of citation network has become a thing that is widely used
in figuring out, you know, did this person really write a cool paper or not? Oh, all these people cited it, it must be cool, type thing.

And it's become, well, for many years, it's become a thing which, in the academic world, is part of the dynamics of, you know, do you get your promotion?

And, you know, oh, do you have enough papers that are cited in enough ways with enough kind of impact factor, because they're cited by enough people in enough high-profile journals and all those kinds of things?

Even more complex… so, in different fields, there are different conventions about what you cite, what you don't.

Authorship is even more complicated, and there are all sorts of different conventions in different fields.

There has been a tendency to have more and more authors on papers, in almost every field. I mean, it used to be the case that there were many fields, particularly theoretical fields, where most papers were single-author papers. In mathematics, for example, that tended to be the case.

In theoretical physics, that was often the case.

In, in something like philosophy or humanities, it was usually the case.

In areas that are more kind of experimental, engineering-oriented, and so on, there started to be just more and more people who'd made similar contributions. You know, I built the such-and-such detector, I did the this part, I did the that part, and in the end, it's 200 people, or it's 500 people, or it's 1,000 people.

In modern machine learning and AI, there's become sort of a convention of, well, there are lots of people who contributed in some way or another, put all their names as authors on the paper.

No machines so far, for most of the AI papers, but lots of humans. And

you know, I think in different fields, there's sort of a decoding of authorship on papers. Like, I think, in many fields, it's like the first author

The first few authors are the people who really did the work.

The last authors are the people whose lab it was done in, or who were the sort of producers who got funding for it, or whatever else, and the ones in the middle sort of did different little pieces of what's in the paper. Now, sometimes there are footnotes, and sometimes there's even narrative at the end that says who did what.

But exactly how this works, I'm not sure. And, you know, at some point, it becomes like movie credits, where there are lots of different pieces, you know, including the folks who made the food and all this kind of thing, that, that get, by convention included. I don't know whether science papers are going to go that way.

But the question of whether you, it's sort of interesting, actually, for movie credits.

there are some branded technologies, like, you know, what kind of lenses did you use? What kind of color development technique did you use? There are some sort of branded technologies that do get mentioned, similarly sound systems and so on. Whereas, I think

interesting question, whether the… I mean, they don't… you know, it doesn't tend to get mentioned. In modern times.

when there's sort of AI, updating of movies and things like this, that technology doesn't tend to get mentioned. I… you know, I think it's a… it's a matter of certain kinds of technologies that were sort of… it's in technicolor or something. It's a… it's a cool technology, and we want to mention it.

Those things, sort of, by convention from 100 years ago, started to get mentioned. But it's an interesting thing, that movie credits include lots of humans and a few technology systems.

And the interesting thing, actually, it would be… it would be very nice if scientific papers also had, you know, people as authors and technology systems mentioned, because we made a lot of

the technology that's used for these things, and it'd be cool to see that get mentioned more. So I… that would be a lovely thing, but that hasn't been the social convention up to this time.
And I think the, the question of, you know, this thing… now, I have to say, in terms of the more amateur avocational side of science, somebody like me gets
a very large number of, kind of science theories that I get sent every day, actually, many, many per day. And it is a little frustrating that in modern times, a lot of those are, I did this in collaboration with
Grok, or something, or I did this in collaboration with Claude, or something like this.
And you know, that's not a big inducement to… it's not a big advertisement, it's not a big positive advertisement.
Actually, the other thing that's happened that I've started to see, which is kind of a sort of a very bizarre thing, is people saying, I wrote up this theory, and here are the comments on it from 5 different AIs.
And I've even seen now book blurbs, where, you know, in a book, you'll often… on the back cover, there'll often be a bunch of sort of pithy quotes from a bunch of people who, one might have heard of or something, who are commenting on this book and saying how great it is.
Well, now I'm starting to see people, feed their book to an AI and have a quote from the AI saying how wonderful the book is.
I don't know whether this is a great kind of human story. I mean, I think it's one of these things where it's kind of a human feedback mechanism. You're talking to your AI, and the AI is telling you, this stuff you've written is just fantastic.
It's kind of nice, up to some point, but it certainly gets one off in a kind of a… peels off from reality in some sense, as the AI just sort of tells you how great the things you're doing are, and I know that has caused a certain amount of sort of psychological difficulty for people, and, you know, that's a thing that is… it's not surprising that's going to happen.
But I think it's a thing for science
That, it's like, look, I have this great… this theory, and it's great, because the AI told me it was great. I don't think that's a,
That's a terrific direction.
Lee.
Let's see…
David is commenting, These days, some scientific publishers forbid AI co-authors. Recently, they say, may be tempted to give credit for discoveries that would have taken years longer.
Yeah, I mean I don't know, the concept
That, you know, don't use this tool.
Doesn't seem like a winning story.
Now the question is, if you use that tool, do you tell people that you use that tool, or not?
Certainly, sort of, the, the, the bootleg, sort of, black market,
AI. I kind of am thinking, you know, if AI is outlawed, you know.
which it won't be, but, you know, the concept of, or if it's outlawed in some particular sort of thing in science, it's kind of like lots of bootleg AI is there. Now, I do know that for the folks who are putting up, you know, preprints, for example, like the big preprint servers like Archive and so on.
One of the things that's happened in the last few years is a rapidly increasing number of AI-generated preprints that get sent in.
They're terrible.

They're, you know, why are people sending them in? Because it's been noticed that, you know, if you can get a, you know, a preprint on there, and you're a high school student or something, you're applying to college, it looks really cool if you can say, I wrote this paper. Well, or at least I put my name on this paper.

This is clearly not a very useful thing, and it's sort of a challenge to rebuff that, and that's kind of a not very useful kind of case.

But I think the idea that, you know, you can't outlaw the tools, but it's still sort of got to be the case, I think, that for it to be meaningful.

for these kinds of human, kind of, processes and things like, sort of, papers people read, it's like the human has to be in charge. The human can use whatever tools they want, but the human has to be responsible for what's there. If it's just like, well, my AI did it, but I didn't notice what it did, that's really not a thing.

Now, I have to say, as I'm saying that to you, I'm realizing that one of the things that is an interesting point is, you know, I'm constantly generating pictures by computer that have tons and tons of detail in them.

I haven't looked at every one of those details. I'm just saying, here's this thing. I can tell you what made it. Now.

kind of, like, it's… I'm passing on this thing for everybody. And I suppose you could argue that in some sense, a sort of an AI-generated piece of text is a bit like that. Here's what this AI said. Make of it what you will.

it feels a lot less useful, and it feels… than saying, here's this precise thing that was generated in this precise way, and now you can conclude something from it. I mean, I suppose you could say the same thing about if you've got, you know, your paper about the new endangered, or the new, very obscure species of lizard that you just discovered under a rock somewhere. You take a photograph of the lizard.

and you put it in your paper, and somebody points out, didn't you notice that that lizard has a word written in some strange script, you know, painted on it type thing? Well, you know, because you communicated that whole picture.

But it's a different… I think it's a very different kind of thing, because I think you're not kind of swapping it in for something where there's an expectation that there's kind of a human, behind the scenes kind of putting the text into your paper. So I think that's a, you know, it's a thing that doesn't, doesn't seem useful.

Memes is commenting, giving the quote, the easiest person to fool is yourself.

says, often attributed to Dick Feynman, I… I… since I knew Dick Feynman fairly well, I can tell you he actually did say that. It's, in lots of different circumstances, so I don't know whether he was the originator of that quote, but it was certainly a statement that he used.

Let's see… MP is…

Saying public discussion around AI was mostly around LLMs during 2025. For 2026, what other approaches to AI would you highlight as showing the most promise to make significant gains in capabilities?

You know… I think… the thing…

that LLMs do what they do. It's interesting, it's useful, it can be tuned up.

But the harnesses you can put around the LLMs, kind of, the LLM is like the wild horse that was discovered. Now you have to kind of put harnesses on the horse to connect it to a plow or whatever else.

that's where I think a lot of the value in the next few years will come from. You can think about it as particular verticals.

of particular domains where one can apply AI. One can think about it as connecting to tools, like our technology, for example, for actually doing computation kind of under the control of the LLM.

But making that

kind of, those kinds of things work better, and really putting this sort of infrastructure around, kind of, the wild horse LLM, I think that's where lots of the value is going to come from, and sort of injecting, sort of, LLM AI into lots of things

which… where it now becomes sort of possible to do. Now, there will be sort of a qualitative change as

LLM AI becomes cheaper and more runnable on local machines. I don't know, I mean, that's definitely an engineering thing, it's in process, hasn't quite happened yet. There are some economic forces that kind of push

to have, you know, oh, we just want a centralized thing. I mean, like, search engines, a lot of that could have been localized, but wasn't for largely economic reasons of particular companies and things like this.

I think there's enough, kind of, diversity of activity around LLMs that there will be significant economic pressure to make LLMs be runnable on local computers.

And, you know, yes, it'll drive having more storage on local computers and things like this, and more, kind of, GPU-like capabilities.

But when that happens, there's lots of kind of LLM-in-the-loop kinds of uses that haven't really been possible when you're going to a cloud and you're spending money on each inference and so on.

And so I expect a bunch of things there where LLMs get more, in a more fine-grained way, integrated into, kind of, computational experience. In terms of

kind of core technology. I mean, there are things like, can you make a system that will learn at the same time as it's running?

Hasn't happened yet. Can you make incremental learning systems? I mean, something that did happen was distributed training, being able to do training in parallel rather than just sequentially. But at least some of that has happened.

then there are things like being able to have sort of fine-grained mixing of kind of precise symbolic computation with LLM technology. I've certainly thought about that a fair amount. I think that's quite difficult. I think that may even be almost theoretically almost undoable.

Just because of the nature of these different kinds of things, and there'll always have to be this kind of interface. I think it's very much related to the question of in brains, when you have kind of direct neural connections, can you kind of do things digitally.

Right there at these… at the level of direct neural connections to individual neurons.

Or is it better to do things digitally when you've got the whole, kind of, output mechanism of your typing with your fingers and sort of… or talking to your computer and getting the result back into your eyes, and so on? Is it… do you need, kind of, the I.O, the input-output packaging of eyes and fingers and ears and so on, or can you sort of pick things up right at the level of what's happening

Down in the individual neurons. And that's kind of the same question of whether you can do sort of a fine-grained connection of kind of these sort of actual computation and symbolic-level computation with kind of neural net technology.

And I think that's a… so, you know, there are other kinds of things. There are, modalities for… Okay, so another big area is, being able to

make use of things you know about data that you don't have to learn from training. So, for example, if you're trying to train a robot.

And you already know about the laws of mechanics and the rotations of objects, and things like this.

Potentially, you need a lot less training data to get to the point where you have a fluent robot than you do if you have to, you know, record every possible configuration of the robot manipulators and et cetera, et cetera, et cetera.

And, in fact, it is a challenge, for example, for robotics, that the amount of training data that's easy to get is not that large. It's expensive to get a lot of training data for most areas of robotics.

And so it's much better if you can make use of, kind of, a physical model of the world.

To, to sort of scaffold what you're doing.

And that's the thing where I think it is going to be possible, and there's a bunch of efforts, even I've been involved in some, that try to kind of burn into the structure of a neural net, or an AI, or the way you encode data for the AI, or whatever else, certain things about the nature of the physical world.

that means the air doesn't have to learn that stuff. I mean, it's a little bit like the very simple thing for humans, that, you know, it could be the case that all our neurons are sort of connected to everything and so on.

But the primary visual cortex, that's the first levels of neurons that are connected to our eyes, those neurons are laid out in a two-dimensional array that mirrors the two-dimensionality of the actual scene that we're seeing with our eyes. And so that's a place where using the kind of geometrical structure of visual scenes

We're kind of simplifying the neural connections to the point where we can do a lot of things kind of for free without having to learn everything at the level of connections of neurons.

MP is also asking, with the current investment in AI research, we might hope for proliferation of alternative schools of thought, for example, more mathematically rigorous methods with proofs that show which problems are solvable by the methods, and which problems are fundamentally hard for the method.

Well, yeah, maybe. I mean, I have to say that the progress on, kind of, foundational understanding of what LLMs are doing has not been fast.

I mean, I've tried to do some of that myself, and connecting, kind of, the things that happen in LLMs to phenomena like computational irreducibility. I think I've managed to do some of that, and the main conclusion is, what goes on is

Is very, sort of, irreducibly complicated.

And, you know, the analogy I've been using, my 2025 analogy at least, has been, you know, machine learning is like building a wall out of rocks.

It's like there are these lumps of computation that just exist

are makeable with neural nets, and the question is, when you train the neural net, can you find those lumps of computation and fit them together in such a way that you build a wall, or achieve your objective of recognizing cats from dogs, or whatever else it is? And I think that, the

The kind of saying that

that we're going to have a sort of a proof of what's going on, I think that's kind of a hopeless thing.

Now, people go to a lot of effort to find, you know, these little mechanisms and little circuits that do things. They do the exact same thing in neuroscience. It's worked a bit in neuroscience, but it's very challenging overall. What's been done with LLMs is somehow technically easier than neuroscience.

But I would say I consider the results extremely unconvincing.

I mean, I think that they're hard to interpret, they're very fragile, they… in many cases, I think they're… they're shockingly kind of unconvincing.

And, you know, this idea that you're going to get sort of mechanistic interpretability, you're going to be able to say, why did the neural net do that? I don't think you're going to be able to say that any more than you can say, why did that stone wall have a little bump up in that place? Well, it's because these rocks happen to be lying around on the ground when the stone wall was made. It's not going to be the case that there's sort of a principled explanation.

Now, you know, one thing you might ask is, if you bash the neural net hard enough.

Will it learn to do arithmetic the way we learn to do arithmetic? Will it learn algorithms by being sort of bashed hard enough? Will it fit into this, this sort of, this niche in the space of all possible neural nets or whatever, that basically just is the circuit for doing that algorithm?

I think that's pretty hopeless. And we've had plenty of examples in the past

Classic example is sorting networks, where people have done exhaustive searches for sorting networks, found efficient ones, and so on, but the efficient, the most efficient sorting networks, even if you are successful.

in finding the most efficient algorithm, it will not be interpretable. The most efficient sorting networks look like random, random little transpositions of things all over the place.

You can prove, not difficult to prove, that they're correct, that they will always sort things, but it's not as if they are understandable, in the sense that there's kind of a narrative that we can understand

The, you know, what's, sort of step-by-step what's happening in that sorting network.

And so, I think… I think the idea that you're going to have, kind of… okay, so let's take the following thought experiment. I've got a neural net, it tells cats from dogs. Let's prove that when shown an elephant, the neural net will never say it's a dog.

Well, I have no idea how to do that. I think nobody has any idea how to do that.

It's, the question of what it considers a cat, what it considers a dog, is very fuzzy. The question of, kind of, you know, the proof that you go to this side rather than that side is,

It doesn't feel doable. It feels like a story of computational irreducibility, where the only thing you'll say is, well, just try it and see what happens.

Now, this notion, let's take the cat-dog example, really what's happening there is you're asking, does the neural net evolve to this attractor or that attractor? So it's the same kind of thing as if you've got a ball rolling around on a landscape, and there are sort of two depressions in the landscape.

Well, you put the ball in one place and it kind of rolls around. Eventually, it ends up in one basin, or if you put it in another place, sort of over the watershed to the other place, it'll roll into the other basin.

And you can say, well, why did it roll into one basin rather than another? In that particular very simple case, you can draw this sort of vector field of in which way is the local gradient going

That, you know, goes down the mountain from that particular point. And you can say, sort of, from which places, you can map out, sort of, the basin of attraction to each of these… each of these different, endpoints.

And so you can say, if it's in this place, it goes to that. If it's in that place, it goes to that.

even the very simple mathematical examples of this, the basins of attraction are incredibly complicated. So, for example, a classic example is, if you have an equation, like X cubed plus 5x minus 7 equals 0 or something, and you say, what's the value of X that satisfies that equation?

So there's a method, well, one method is so-called Newton's method, where you start from some value of X, like 1, and at each successive iteration, you, use derivatives from the equation to go closer to the correct solution.

Now, that particular equation has three solutions, so depending on where you start off, you're going to end up in one solution or another by applying this iterative method. Well, the question is, depending on where you start, which solution do you end up with?

That picture of the basins of attraction for different solutions is this complicated fractal mess.

And that's in the case of just a simple polynomial. So in the cats and dogs case, it's a… it's a hugely challenging thing to kind of imagine how you would map out

even… how do you define the space of possible images? How do you map out where you go to?

Actually, in some things I've done, sort of explaining neural nets, like the thing I wrote about, how ChatGPT works and so on, I did show some pictures

of what the basins of attraction look like for very, very, very simple neural nets as a function of numerical parameters. And what you see there is, well, the basins of attractions are kind of complicated. It's not easy to give, sort of, a narrative description and to give a proof

that… I mean, again, with images, for example, to say the things that are in this part of image space, whatever that even means, will go to this, the ones in that part of image space will go to that.

We just don't have good mathematical technology to even talk about that, to even imagine what it would look like to give a proof of those kinds of results.

So I think that's… it's very far away. I mean, in… and some of the things that I've done, some of the kind of foundational breakthroughs we've been able to make in some of these areas of science, I think have the best chance we have right now to develop a formalism

to try to make these foundational understandings of what's happening in AI systems, but we're pretty far away from it right now, and I would say that the,

the sort of motivation isn't completely there, because some people are just doing engineering. A lot of people are just doing engineering, and it's like, we just want to make a thing that works.

We don't really care why it works, how it works, we just want to make a thing that works.

And then there are people who want to do something academic in this area, and they've had a very hard time, sort of, getting into these foundational questions, and a lot of what gets done is almost the psychology of AIs. If you feed the AI this or that, what will it do?

Rather than these more foundational kinds of questions. There are things at the periphery of AIs, like the tokenization of inputs and things like this, where you can start saying things a little bit more, kind of, abstractly, but in the main part of it, that hasn't really been probed, and that's a… that will be a super interesting thing if it happened in a big way.

in, in 2026, it needs a breakthrough. It needs a science breakthrough.

you know, I think it's a really interesting thing to work on. As I say, I think the methods that I've kind of been developing the last few years are our best chance to get that breakthrough right now, but I don't know how to do it… yet.

All right, I should probably return to my day job here, but thanks very much for joining me, and I will

For the next few weeks, at least, I should be doing my regular schedule of livestreams on a variety of different topics. We have been planning some new kinds of livestreams for the new year. Probably starting next month or so, I'll be adding in
some livestreams about technology and about our technology in particular.
kind of more, how do you use this technology to do things? I may also add in some live streams that are,
actually doing some live experiments, I've done those before. Also, one of the things that,
is something that we are now doing at our Wolfram Institute, is we have a number of programs that are trying to kind of pull in
various kinds of people, researchers, academics, and so on, in existing areas, and we have kind of a need to educate some of those people so that they can make the kinds of contributions that we're able to sort of support at the Institute.
And we may end up livestreaming some of the, kind of, my efforts to provide some of that educational content. One of the areas we're going to be doing that in is philosophy, trying to, really… we're really trying to
Creates some material that connects, sort of, the,
the kinds of things that we've discovered in foundational science to classic ideas and conclusions, questions in philosophy. And to do that, we need to sort of educate folks who know about that classical philosophy, and we may end up livestreaming some of those kinds of things as well. But, anyway,
That's it for today. Thanks very much for, joining me, and, I see a comment here from Prabh. Give us a tour of books and artifacts. Yes, next month I plan to do that. It's just a bunch of, it's…
a little more difficult to orchestrate. We just… we practiced it, actually, a few weeks ago, and I think… I think we can do it. It's a question of if you're… if you're going up to bookshelves and pulling books down and showing people what's in the book, it's like, how do you have a camera that can look at the book and also look at the shelf and do this and that and the other?
Without a giant movie production effort.
But I think we may have debugged how to do that.
Anyway, the, I'm seeing another comment here. Computational service overview.
We're good here.
I'm not quite sure what that means. I'd like a little bit of clarification about that. And, another comment from Mark. So, plus one for classical philosophy lessons. Okay, well.
My main thing is trying to bridge, what we've done in foundational science with classic philosophy. I am… I am…
Not confident. I think I do know a decent amount about classical philosophy, but…
I don't think I'm confident of my knowledge there, in the same way that I feel fairly confident in my knowledge in lots of areas of science. I mean, if somebody says, you know, what did Heidegger say about this?
And I say such and such. And somebody says, no, no, he didn't say that, he really said something quite different.
I don't feel confident kind of…
to know enough about what was said. You know, if somebody says, you know, if I say, I think Turing said this, somebody says, no, he didn't, I'm really quite confident that I know what Turing said.

Because I've been able to understand a decent fraction of that corpus. That is not true for a lot of classic philosophy. I think I understand the big picture, but I don't think I can… I know enough about the whole corpus to be able to defend that, and I'm kind of hoping to bridge this
To people who really do know those, those whole corpora, And,
And perhaps we'll have some way of having sort of an interactive session in which we can involve those folks, so we can all learn those kinds of things.
Anyway, thanks for joining me, and bye for today.