Hello, everyone. Welcome to another episode of Q&A about future of science and technology.
So I see I have a bunch of questions here.
George asks, as AI becomes able to generate knowledge faster than humans can absorb it, do you think the future bottleneck of science will become discovery itself, or human understanding?
I would say that,
there are many areas of science already. Forget AI.
There are many areas of science where the ability to generate data, to do experiments, computational or otherwise, far outstrips the ability to, quotes, understand what's going on. It's worth understanding, you know, what is science? What is the purpose of science? You know, the world is doing what the world is doing, you know, natural systems are doing what they're doing, abstract computational systems are doing
what they're doing. When we think of science, we think of it as being some way to take all those things that are going on and form a narrative in our minds that allows us to think about those things in a useful way. So, in other words, science is the sort of bridge between what's actually out there, which can be… which is many, many different things, and that sort of thin set of things that we can understand in our minds.
I mean, I would say that, for example, my effort in building Wolfram language as a computational language
The effort there is to sort of define a language which allows you to think about things in the world in computational terms. We have, you know, the thing… the ways we think about things, we think about things using kind of mathematical formalization. We think about using, kind of, logical formalization, and so on. And computational formalization is another big part of that. But the main point of science is take those things that are out there and make them kind of something about which we can think through a human narrative. So, for example, in things I've done in exploring the computational universe of simple programs, it's very easy to do experiments that generate tons and tons of interesting things. You look at them, you say, that's a picture, that looks pretty interesting.
interesting. Do I, quote, understand it? Do I have a way of sort of connecting that to kind of a main, sort of the main trunk of
kind of scientific things I think about, not necessarily. That is already a bottleneck. Now, in terms of AI,
AI can certainly make use of
that data and can make commentary about it, it can write things based on, sort of, using as data the scientific literature already produced by humans, and yes, it can produce tons of stuff. I mean, I saw that even today, some people working at our Wolfram Institute had produced a whole bunch of material that was mostly written by AI,
And it's like, you know, what is this? Some of it looks kind of interesting, but it takes a lot of effort to read through it, and some of it is useless and uninteresting and off track, so to speak, but some of it isn't. And, you know, it's really the same story as looking at life sciences, where there's just a drowning in data and not much in the way of theory, and so on.
So I think it is sort of the fundamental problem of science, in a sense, is
to sort of… to be able to get to the point where there's human understanding. That's sort of the purpose of science is the human understanding, and it's challenging to get to that point. You know, can AI help kind of take this sort of large amount of data that's out there and kind of grind it down to something that's human understandable.
to some extent, I think that, you know, it can do some of that. It can also just generate more stuff.

Which, as I say, is rather easy to generate by doing physical experiments or computational experiments.

So, yes, I think that is the bottle… the bottleneck of science is the understanding part of it, not the mechanical grinding. I mean, in my own work in building, you know, Mathematica, Wolfram language, Wolfram Alpha, all these kinds of things, they're all about

being able to automate the mechanical aspects of doing things like science, and really getting to the point where, sort of, the picture of what one does in science is you have an idea, you have a kind of pattern of understanding, and the sort of technical details can get done automatically.

And I feel like I've been sort of at that point for a long time. The bottleneck really is kind of the…

human, in my case, my understanding of things. And so, yes, I think that's a… that's sort of a big issue there.

Okay, there's a question here from Alyssa.

If the scientific method were invented today in the age of AI and massive computation, do you think it would look fundamentally different from the version created centuries ago? It's an interesting question, I'm not sure I've thought about that. Let's, let's unpack that a bit. So, first of all, what is the scientific method? And then, does science actually use the scientific method?

You know, the scientific method tends to be you have a hypothesis, you do an experiment relative to the real world, typically, and you see, is your hypothesis true or not? If it is.

validated by what you observe in the real world, then that's good. If it's not, if you sort of falsify your hypothesis, then you,

Kind of go back and come up with something different.

I mean, that's kind of the, the sort of the typical pattern of, of the scientific method, so to speak.

Now, in practice.

an awful lot of science doesn't really work that way. I mean, many of the great theories of science don't have that pattern. They are more, kind of, ways of understanding things. Famous example being the idea of natural selection.

You know, given any weird thing that you observe in the plumage of a bird, you can always come up with a story for why the bird has that kind of plumage. But that may or may not be… and validating that, proving that that's the why, is…

essentially undoable.

But having that as a sort of pattern of explanation is useful. So it's kind of not… and if you look at areas, well, like mathematics, for example, they just don't have the same pattern of kind of… I mean, in mathematics, you're not sort of doing experiments in the real world to see what theory is true, you're… you're trying most… in most pure mathematics, you're trying to sort of put together these proofs, explicitly in that formal system, so to speak.

You might also be doing, sort of.

computational experiments, and I suppose there you could say there's a sort of… a version of the scientific method where you try and do all these computer experiments, and maybe you find a counterexample to the thing you thought was true, in which case you're kind of back to square one, and not… and you've sort of falsified your theory, but that's not the typical pattern of the way people think about, sort of, the scientific method.

I suppose, you know, I will say that on the ground, in doing science.

Theories are hard to make, experiments are hard to do. I've certainly had many experiences in my, sort of, early life in science where I came up with theories, the experiments seemed to say those theories couldn't be correct, but it turned out the experiments were wrong and the theories were just fine.

And, so it's a… it's always a complicated sort of balancing act of the experiments are hard, the theories are hard, it becomes a complicated, almost sociological issue. You know, do people really want the experiment to come out a particular way, et cetera, et cetera, et cetera. But I guess the question here is, in…

Modern times of, of kind of,

Well, okay, I'll say one thing about massive data and so on. You know, a lot of areas of, particularly life sciences, social sciences, things like this, let's say particularly life sciences. Were kind of… the first quantitative work done on them was done 100, 150 years ago, things like this.

And there were sort of small amounts of data, and people had to kind of squeeze out of that data everything they possibly could. And the methods of statistics, hypothesis tests, all the various, you know, the so-and-so tests, you know, the, I don't know, various kinds of experimental design where you're trying out different possibilities and trying to get the most data.

you know, analysis of variance types of things. All these kinds of methods were invented typically in the 1930s and so on, often in the context of agriculture, of figuring out, you know, if you have a… if you're trying to test 5 different possible things that might affect whether your crops grow well or not, there's the question, sort of, how do you do that? Because, sort of, there are two to the 5 possibilities if you're on and off for all those things. Do you really do all the 32 cases? Do you do fewer cases? How do you… how do you organize

sort of doing the experiments, those kinds of things. But most of all, you're trying to get, sort of, out of fairly small amounts of data, you're trying to pull conclusions which are just sort of on the edge, where there's some kind of statistical distribution, and is it… are you in the tail? Are you too far in the tail? Are you so much in the tail that the thing isn't really what you thought it was, and so on.

Lots of methods invented for those purposes.

In many of the things I've done, mostly computational experiments rather than physical ones, but physical experiments where there's automated data taking have the same feature, there's just oodles of data.

It's… there's no lack of data, and the effects you're looking for, if they're real effects, you typically can make a well-designed visualization, and the effect that you're looking for is either staring you in the face, or it's not there.

And all of the kind of statistical, you know, trying to tweak it and reel it in and so on, doesn't really get you anywhere.

And I think… so that's a… that's a difference in…

You know, and some of the methodology, particularly in biomedical research, one is looking for, sort of, it's all about small effects and so on, and it's all about, kind of, small statistics. Now, when you do clinical trials and there are people involved and so on, again, you're back to that small sample type of situation, but in more, sort of, foundational biological research, for example.

You're typically doing automated data taking, and you have huge amounts of data.

So, you know, that's probably one difference, is sort of the importance of visualization and the importance of going sort of straight from the data to something that you can internalize, without going through layers of kind of statistical interpretation and so on, is probably one… one difference.

In terms of…
of what can you do with AI that's different. I mean, one big difference is that AI sort of allows you to make quantitative things which only sort of had a qualitative way of being talked about before, whether it's, you know, how many of these images are really images of turtles versus images of something else.
That's something that you can be fairly quantitative about now. Things about, sort of, textual data, you know, did the, did these people really,
You know, are they really saying this or saying that?
Now, sometimes the AI's discrimination of these things won't be sort of everything that a human would do to discriminate these things, but it'll… it's probably fairly good and fairly close to what humans would do, and it sort of opens up a whole bunch of other kinds of things that become quantitative that were not quantitative before.
So I think that's another kind of,
kind of thing. Now, in terms of this, kind of, the loop of,
Hypothesis, experimental test, you know, was it right or not?
I'm not sure. I think that is a poor model of most of science as science is actually done, both in terms of the pattern of the way that theories are built.
and in terms of the sort of sociology of the way experiments are done, and so on. It's a very kind of idealized view of science that I think has often been very misleading, and I'm not sure the extent to which, sort of the modern AI world, how much… you know, I think it was already a misleading and bad model, so I'm not sure that, you know, maybe something about AI makes it really obvious that
a bad model. I mean, for example, the idea of sort of units of scientific progress being scientific papers that people put certain amounts of effort into, that's kind of all being thrown into disarray, because it's sort of easy to generate something which is kind of like a scientific paper.
just by pushing a button with your AI, and what that means in terms of how, sort of, science should be absorbed and disseminated is not clear. I mean, I tend to think
that a lot of the value of what gets disseminated is more, sort of, computable kinds of things. I mean, it's clear the dissemination of software is clearly a worthwhile thing.
And it's clear, you know, software is more immediately usable than something that's written in a book.
And insofar as, sort of, things about science can be encoded in computational terms, they become immediately buildable on in a way that isn't true if you're just writing some paper. I mean, I've been pushing now for
Oh, 45 years, for sort of, computational, scientific publishing, so to speak.
The idea that you can make, kind of, your average paper be something that doesn't just have a bunch of words and maybe formulas in it, but has actual
computational language code that expresses the ideas that are in the paper, but allows one to kind of immediately build on that. I mean, I've been a… I've been kind of taking action where I've been talking about these things, because the things that I write
every picture in them, there's Wolfram language code behind them that you can run to get that picture, and that's been super valuable in a lot of our educational programs, where people have started from the code that's in the things I've written, and then they can kind of go from there and get much further than they could if they had to sort of think about re-implementing things.
So I suppose in terms of the, kind of, the arc of how science works, what one can hope for is that the AIs will make it a bit easier to get computational language code out

a wider range of people will be able to write that code. We'll see, I mean, we're doing lots of experiments on getting AIs to kind of make that code as clear as possible, but that provides kind of a different picture for things like reproducibility of science and so on. It's like, well, it was, you know, here's the piece of code and anybody can run it.

Even you can go further, when you're dealing with cloud labs, automated data taking, and so on, to say, well, this is the script.

for the cloud lab, written in computational language, you want to know whether the world works this way at a level of actual chemistry or something, just run this script in the cloud lab, and it will come out the same way it came out before. You're kind of turning, sort of, the doing of experimental science into something that's all driven by software, rather than being something where it's kind of like you have to get, sort of, by hearsay.

here's how much you have to shake the test tube type thing. I think,

So this idea… of,

Yeah, I think, you know, to some extent, as I say, the scientific method, as often presented, is kind of not a good model of science, and

I think that some of the things that it tries to idealize away, like the sociological aspects of science and so on, are probably still alive and well and don't really get affected that much by AIs. So far as I can see, right? Right? Immediately.

Let's see… There's a question here…

about, another thing from Nuno, talking about several paralyzed people, Have an injection of something,

And we're able to move again.

Do you think in the future, paralyzed people will be able to walk again?

My guess is yes. It depends on what's gone wrong. You know, if it's, kind of some break in the spinal cord, there are, well, a bunch of companies and efforts

For figuring out how to somehow repair that.

either… I mean, the challenge is either it's a question of getting, you know, there are many fibers in the spinal cord, and, you know, nerves grow at a slow rate, and I don't know how, you know, how each fiber finds the, you know, if you break it, how the fibers sort of find their way to the right fibers on the other side. But I think one thing that sort of one can hope for is that you can kind of put digital electronics in the middle of that.

And you can be kind of, rather than having to wait for the, you know, incredibly slow growth of nerves at, you know, fractions of millimeters per year and things like this, you can just say, you know, have a piece of digital electronics, which is kind of the switching station. That's one approach.

Another approach is, you know, perhaps less satisfactory, is forget about the human muscles, you know, have the exoskeleton or something like this, that is controlled by nerve signals from the spinal cord, from the brain.

And the challenge there is, well, what do the… you have this whole bundle of nerve fibers, and sort of each one is connected to a definite place in the brain, a definite part of the motor area of the brain, and each one is kind of like, you know, it's the wiggle your toe

You know, wiggle your left toe kind of fiber, and another one might be, you know, move, move your knee kind of fiber, and the question is, which one is which?

And one might hope, and I know there are a bunch of experiments on this, that sort of one could essentially machine learn from saying to somebody, you know, try wiggling your knee, even

though you can't actually do that because the nerve fibers are broken, for example. But if you think about doing that.
Then, can you get some digital electronics system to kind of
see that signal and figure out what to do from that. And I know there… I've seen,
well, I've seen videos, and I've seen… I've talked to people who do the work, but I have to say I've not seen actual, patients, in… in the flesh doing this, of being able to, to… to get, sort of, the reprogrammed, kind of,
Ability to… to do that.
I think, you know, then you can ask questions like, are you directly stimulating the neuromuscular junctions and things, and are you, you know, stimulating nerves that are kind of far up from that, or are you going all the way down to those things, and so on? I don't know the details of how that will work, but my strong impression is that
Over time, there will be, sort of more and more ability to sort of have some digital electronics in the loop
that help out things that are broken in into the connection of nerves and so on. I think the other one that is sort of a hoped-for thing is that if you
are trying to speak, but you don't have, you know, there's something wrong with the mechanics of the vocal tract or whatever, that one will be able to kind of pick up the signals that would be going to the vocal tract from the appropriate area of the brain, which turns out to be, I gather, fairly near the surface of the brain, so you can expect to put a little electrode array right there without sort of having to go deep into the brain.
put it on the surface, apparently, and start just sort of thinking what you're going to say, and then have some piece of digital electronics that actually does the saying, so to speak. So I'm sort of… my impression is that those things are, you know, it's always the case with these medical things that they sort of take longer than you expect, and there are more kind of gotchas and details and problems of, you know, oh, if you do that, you know, infection will creep in, or if you do this, the thing will habituate, and the effect will no longer be there, and etc, etc, etc. But my impression is that's definitely the direction things are going.
A question here…
from Duncan, how will we teach the scientific method to students if real science is done with huge computational systems?
There are… You know, the kinds of experiments that Kids, students, whatever, can do.
There's a lot of automation that makes in… well, okay, when it comes to computational experiments.
those are sort of easy for anybody to do. Yes, there are big, crunchy ones that take more effort, but you can do a lot of interesting ones just with your average laptop. And,
that's an area where, sort of, the scientific method of, you know, trying to understand, sort of, making hypotheses, seeing what happens, and so on, it's absolutely alive and well. I would say that's not taught
as adequately as it should be, or perhaps even really effectively at all, in sort of typical K-12 education, for example. I would say that the whole methodology of computer experimentation, which I've been
you know, kind of myself been sort of pushing for half a century now, is, is still sort of a coming soon kind of thing for many people. When it comes to physical experiments.
you know, I think… It's…

Well, I don't know. I don't know to what extent… I mean, I don't know, I'm trying to think about even experiments I did when I was a kid, for example, where, sort of, the genuine scientific method was usefully used, as opposed to it all being a bit of a fake.

And I'm… I'm having a little bit of a hard time. I mean, I… I am,

You know, did, you know, you do, kind of, little natural history type experiments,

Yeah, I'm not sure, I'm not sure that that was ever a really convincing thing. So I think,

You know, the idea that you do experiments and things happen that you don't expect.

that's an all-the-time kind of phenomenon in computational experiments, and that's probably the best way to place to see that. I mean, the thing you're teaching in physical experiments is usually it's just hard to get stuff to work in the world.

And, you know, it's so hard to do that little biology experiment, the chemistry experiment.

Not because there's something intrinsically difficult about it, but just because you have to hold the test tube at exactly the right angle, and, you know, it's got to be just this height above the Bunsen burner, and it's got to do this, and, you know, what if the thing, you know, got some little impurity in it, and etc, etc, etc, which is a different thing that you're teaching than something about, kind of, the overall arc of how you understand whether something is true about the world.

Let's see…

Luke is asking, could there be a future where most important scientific experiments happen entirely in computational universes rather than in physical labs?

I'd like to think that a bunch of science I've done over the last few decades has had some important features, and it's all computational.

In terms of…

when, you know, when can you have kind of a digital twin of what happens in the real world that's accurate enough that you can really do scientific experiments on it? In some areas, that is the case, it's already the case. In other areas.

you know, that there are… we just can't do the computations fast enough. I mean, we're bitten by the phenomenon, essentially, of computational irreducibility. Even though we might know the underlying rules, the laws of physics that govern some particular system, to actually work out what the system does, you have to kind of follow each step

And to do that in a computer is hopelessly inefficient compared to just doing it in real life with real molecules and things like this.

I would say that, I'm trying to think what areas,

I mean, there's a slow progression of some areas being sort of simulatable, sort of effectively enough that you don't actually have to do the physical experiments.

In some cases where the physical experiments are just more efficient to do, it's like having an analog computer. Instead of having a digital computer where everything's been turned into bits, you just have something where you have an analog of the system, so to speak, or it could actually be the system.

sort of put in a harness of some automated data-taking, automated, sort of preparation-type, type setup. So I, I kind of think,

There are plenty of discoveries to make in the computational universe, but there's also a bunch of things that are computationally irreducible, where you kind of have to do them in real life. There are also things, although they're much rarer.

where there's sort of a… where we don't know the underlying… the sort of the ground truth of what's underneath all these systems. We don't… you know, if we're trying to study fundamental physics, we've got to actually use the particle accelerator, because we don't… we don't have it… we don't know what the underlying

structure is where we could just use computation to build up to what we see. I mean, I actually think with our physics project, we do have an idea of what the underlying structure is, but it's a very long road from what we know about the underlying structure to something you can actually observe in an experiment.

Let's see…

an interesting one from Brady. Will there be endangered forms of human thinking that we need to preserve the way we preserve languages?

have not been that good at preserving languages. I mean, there are probably about 7,000 languages spoken around the world, and there are plenty of endangered languages where the number of speakers is small, and where there are no kids who are learning the language, and it's kind of only people getting older and older who speak the language.

I mean, I think that

all sorts of forms of globalization have definitely endangered lots of languages. All sorts of, kind of, uniformization brought on by whatever, you know, by both economic pressures, political pressures, etc, etc, etc, have… have tended to make it unattractive for people to have, kind of, some… some small-scale language. I mean, in other words, if you,

If you had,

a language where… where there are only, you know, 2,000 speakers of this language, nobody's going to produce material for that language, or very few people are going to do that. It's not really economically viable.

So… but, in any case, there are efforts, although they're rather small in number, disappointingly small in number in some ways, to preserve endangered languages.

And, there could be more done on that. Now, in terms of ways of thinking, that's an interesting question. I mean, if we look at history, there have certainly been

times in history where, sort of, there's been lots of pressure put on don't use that way of thinking. I mean, in,

In kind of the… the sort of 1600s, there was a kind of certain pressure of, don't…

talk about science that goes against, sort of, fundamentalist religious doctrine, so to speak. You know, that's not a way of thinking one should pursue, so to speak. And then in more recent times, it's kind of gone the other way. It's kind of, don't…

have that kind of, religion-based thinking, you know, science is, at least in many places, kind of… science knows the way that… the things that are true. Now, you know, I would say that the… the,

Perhaps… One could say that the most endangered and concerningly endangered form of thinking is thinking at all.

In the sense that there are certain, you know, when one's looking at, sort of, the way, sort of, education works, sort of, media works, and so on, it's very much like, let us tell you what you should think. Here it is.

And it's kind of like, if you don't let us tell you how, you know, how you should think, you won't do well on the test type thing. And, you know, I think it would be the question of, you know, are you actually thinking for yourself

That's… that's a pretty worthwhile thing for humans to do. It may not be… sometimes that's very inconvenient.

whether for a teacher running a big classroom, or whether for society trying to sort of have everything kind of run harmoniously, people sort of thinking for themselves and doing things that, you know, if we were a school of fish, it's… if there are too many fish that think for themselves, you don't have a school of fish anymore. You have fish doing random stuff.

And, you know, so there's a certain dynamic of societal pressure, and maybe, maybe, you know, if everybody's

too creative and doing their own thing, sort of, there isn't even a coherent society that forms the sort of regularity that, for most of us, lets us sort of, you know, not spend our whole life kind of dealing with the irregularity of the world, but spend it doing other things that we might want to do. So I think… but, I mean, this question of is there

Are there things that people,

the ways of thinking about things that get lost. I would say another… another perhaps relevant thing.

So… Most… sort of…

sophisticated ways of thinking, things where there's been a big tower of you think this, and then that allows you to think this, and then that does this. You know, those towers are tallest in areas like mathematics,

But there are plenty of towers in areas where there's formalization of thinking. What tends to be the case is that there are certain kind of guilds of people who understand that way of thinking. And every different field has a different way of thinking. Physics has a different way of thinking from chemistry, very different from sociology, let's say, very different from pure mathematics, let's say. They all have different patterns of thinking, and the fact is that

the… most of the time, the people who sort of adopt those modes of thinking are sort of deeply bought into those, in effect, guilds. They, you know, went, they did their PhD in that area, they learned all kinds of things in that area, they have that way of thinking sort of understood.

So it's an interesting question, whether, you know, there are… I'll give you an example. I mean, I don't know, let's say in biology, there's a lot of, kind of, natural selection-based evolutionary thinking. It is that way because of some advantage that the organism gets. Not always correct, but that's a common way of thinking.

or let's say in economics. There are all kinds of ways of thinking where, oh, it has to be this way because it increases the utility for this, you know, agent in the market, and so on. It's always, to me, interesting, as I try and learn different fields, that when you're first exposed to the field, there are people where you… who are experts in the field, who, given a question, will just be able to, you know, boom, boom, boom.

they'll think it through with the paradigm of that field, and one is just left kind of like, I didn't, you know, I didn't see that.

And… but if… once you get into the field and you really understand the paradigm, you do get to sort of follow along with that and see… and see how that pattern of thinking works. So, it's a reasonable question. Are there patterns of thinking that go extinct? And the answer is surely yes. Just as there are, sort of, ways of,

what's a good example of a way of thinking about the world that's kind of gone extinct?

Let me think… Well… Thinking about different areas of science, and where,

Where people feel like, sort of, that was,

I mean, there are plenty of things where people would figure stuff out. Yeah, I mean, here's a good example.

The natural philosophy way of figuring out things in physics.

Can one just think?
kind of philosophically about the way the world works, and figure out something in physics. That was the thing one did. Before the late 1600s, that was kind of the only way to figure out things about physics, was just by pure thought, sort of informed by obvious experiments.
But then, kind of the idea of mathematicizing physics came in, and these sort of tall towers of conclusions were developed.
And and then, to some extent, people were like, if it isn't math, it isn't physics, so to speak. If it's not the result of some formula and calculation and so on, that can't be a piece of physics.
Now, interestingly, there were some moments of great progress in physics that were made by something much closer to natural philosophy-style thinking. Famous example being the 1905 Einstein's invention of special relativity.
Which you can really think of as being
sort of a deduction in natural philosophy, with certain assumptions about the world, the speed of light is constant in all frames, and, you know, that there's invariance, things look the same in different frames, and so on. And therefore, as a matter of pure thinking and natural philosophy, it must be the case that, you know, lengths are foreshortened if you're going faster, and things like this.
That's a pretty good example of a form of thinking that was all but extinct at that time in physics. I mean, Boltzmann had also taken a similar approach.
But people were very skeptical. Some people were very skeptical about relativity at the beginning, because it did seem to be a thinking, a kind of thinking that was coming from a different place, and from a much earlier time, of centuries earlier, so to speak. So that's an example of a piece of thinking, a type of thinking that sort of almost went extinct. I think that there are surely other examples of that.
And I would say that the… the general, sort of,
way of thinking about things… well, okay, so there's probably other ones. There's probably ones in computer science.
Where… Let me think… Well…
Kind of early ideas about, sort of, simple models of computation Gave way, in computer science in the practice of computer science and what it means academically to, you know, build big software systems and do things like that, rather than, except in the theoretical computer science area, rather than think about these sort of simple models of computation. And then perhaps an area one might say is definitely not
as prominent as it could be, of sort of thinking abstractly and in a sophisticated way about, sort of, the foundational ideas of computing, rather than about, sort of, the details of software engineering. A few thoughts there.
Let's see… Okay, so is it… Question here from Enrique.
Saying, heard that if there were aliens a certain number of light-years away who looked at Earth, they would see the Roman Empire. Is it possible to use this theory to create a technology that does this for humans to look back on Earth's history and confirm ideas that are only theories?
Well, first thing to say is that, yes, if you're a star 10 light-years away, the… the… you will see light coming from the Earth
only, sort of, 10 years later, but there isn't really… there isn't really a meaning to saying 10 years later, because how do you synchronize the clocks? In other words, the light started from the Earth now.

10 years by Earth time, 10 years later it will arrive at the star 10 light years away, but how would the folks on the 10 light-year-away star have set their calendars?

that's the thing that is sort of arbitrary, and that's kind of a key idea of relativity, that this sort of relativity of simultaneity, this question of sort of what counts as now, what's the, what's the… sort of, how do you… how do you decide, for example, on Mars, which is often, like, 20 light minutes away from the Earth.

So it takes, like, 20 light minutes to get

put a signal to get to Mars.

if and when there's a Mars colony, will that Mars colony be operating on, for example, GMT, on Greenwich Time or something? That would actually be very funky, because the distance from Mars to the Earth varies through the Earth year and the Mars year, and so the question was, would you… would you kind of reset

the time to be such that it has a constant delay from the Earth, or how would that work, and so on. This is all to do with, kind of, in relativity, the idea of making reference frames where you're kind of defining what counts as simultaneity. It's a very practical issue if you're dealing with spacecraft, deep spacecraft in the solar system, for example. What is the, kind of, what counts, you know, when you say.

when, I'm sure in, communicating to the,

Pioneer spacecraft, or… or whatever, Voyager spacecraft, the most distant things still being communicated with, I guess that's, I think, the… I'm trying to remember, is it Voyager or Pioneer that's still in radio contact? But in any case, it's,

I think several light days away from the Earth.

And it has to know, because the way that communication is done with deep spacecraft, is there are a bunch of antennas on the Earth, and at a certain time of the week, those antennas will point at a particular spacecraft, and kind of… that's its moment. And the spacecraft kind of has to know when that moment is supposed to be, and so when you set the clock.

to determine, you know, when should the spacecraft be waiting for the transmission from Earth, you better have figured out, kind of, what the reference frame is, what counts as simultaneity. It's like, we're gonna… we're gonna transmit to that spacecraft at 12 noon on Fridays. But the problem is, our kind of, you know, what… what that means for the spacecraft is kind of a different time, because it's… it's… you can…

kind of, you can only synchronize things by sending light signals back and forth, and so you're always… you could always say, well, it's the thing that comes with the, you know, whatever the light signal delay is from the time the thing was emitted at 12 noon on Friday, or whatever.

Okay, so in any case, the very notion of

Kind of, sort of, do you get to see things which were happening in the past?

and so on, is it depends very much what you mean by past, how you set up these simultaneity surfaces, and so on. There's a different issue, which is, could you have something where, for example, light from the Earth

It's kind of like the, all the, sort of, the, the, I don't know what the, it's a good example, you know, the Great Fire of London, or something, from whatever the 16…

60s, maybe? You know, probably was to a modern satellite orbiting the Earth, that would have been very visible, as wildfires are visible today.

So, the photons that came from the Great Fire of London have gone off into space, and they are… they've now gone 300 light-years, or whatever, away, 350 light-years away from… from the Earth.

And so one question you might ask is, well, could you somehow reflect those photons from something 350 light-years away from the Earth, send them back to the Earth, and in 350 years say, oh yeah, that's what the Great Fire of London looked like?

Okay, so there are a few practical problems with this.

the number of photons… well, the number of photons produced by the Great Fire of London will be… have been quite large. But those photons, by the time they're 350 light years away, they're incredibly dispersed, they're incredibly dilute.

I mean, in general, the kind of the… let's assume that they go out sort of in a spherical shell. This is kind of the inverse square law story, because the area of a sphere, the surface of a sphere of radius r is 4 pi r squared, and that means that if the photons are kind of equally distributed in direction, that means the number of

photons in a particular direction is going… a little patch in a particular direction, is going to decrease, like, 1 over R squared, and that's the same thing as saying if the radius of that sphere is expanding at the speed of light. So that means the number, the density of photons is going to get very, very small. Now, how many photons do you get? Is it enough photons

That you could kind of reconstruct an image.

well, it's really hard to collect those photons. I mean, in the technology we have today, for… there are situations in which you can observe single photons, even single photons of visible light. But that would be a challenge to collect, you know, to have some big collecting dish that's collecting all these photons, and I suppose what you would want to do is, if you had a pure mirror.

then I think you're… you're sort of pretty out of luck, because the… then, you know, you're trying to… you're reflecting these photons back where they came from, and you're doing another inverse square law, and the number of photons that you'll get back is really, really small.

But can you imagine, conceptually, that, the,

That, you could kind of reflect back photons and see the past.

You know, it's funny, in the very early days of computers.

People wanted to make memory systems, and they wanted to have kind of digital data that could be remembered for a short while while the computer was doing its calculation.

Well, a technology that was used a little bit before my time was mercury delay lines. And so you would have a tube of mercury, liquid mercury, and you would have a loudspeaker at one end. And, you would basically put, sound waves into the… into the tube of mercury, and then you would kind of pick them up at the other end. But those sound waves, because mercury is very dense and so on, those, and I guess, now that I think about it, I'm…

I'm thinking, how would you compute the speed of sound? Let's assume the speed of sound is quite low.

which I'm not quite sure why it is, but let's assume the speed of sound is low in Mercury, which I guess must be.

for this to have worked. That means you're putting sound in at one end, you're kind of waiting for the sound to go through this whole thing of mercury, then you've got that signal at the other end, then you sort of feed it back to the beginning again, and you keep kind of recycling it through this mercury delay line, and it's kind of like you're keeping on remembering it. You're remembering what you're remembering, and so on. I mean, in,

you know, it's quite possible that in brains, there's a certain amount of that that goes on, too, of kind of these loops of remembering what you remember and so on. But that's a very practical case where you literally would send a signal, you'd kind of have it just go through a delay, and

then you get the signal back again. So an extreme version of that would be having those photons from the Earth go out into space and sort of reflect them back with a mirror.
not, I think, directly practical. Now, it's worth saying that in terms of, well, what happened in the past? How much do we know about the past? One of the things that's been surprising in the course of circular archaeological history is that
we're getting to be able to say more and more detailed things about the past. As we can get to make molecular-scale analyses of things, we can pick up DNA from that, you know, that bowl that somebody used 4,000 years ago, and we can find out this is the, sort of breed of sheep that they were, kind of eating at that time, and all that kind of thing, which seemed inconceivable. But now we can pick up individual molecules
and DNA is a stable enough molecule that it will have survived for those 4,000 years. We can do things like that. We can sort of get into studying the molecular scale structure of things, and we can know things about what happened in the past vastly more detail than we could have ever imagined being able to know.
And it is one feature of solid objects that solids are kind of solid. They don't change much. And so, it's a good question to what extent there are sort of traces of everything that's happened to a solid object in the course of a long period of time. Like, you know, could it be the case that sort of, if you look down at an individual molecular scale, that, you know, when I type on my computer keyboard, that every… every piece of typing that I ever did is somehow encoded in the motion of one or two atoms in the, you know, in the… in the U key or something on my keyboard.
And, you know, when it comes to archaeological kinds of things, is it the case, you know, you can look at, kind of, wear marks on stone tools from the Stone Age, for example, and know a little bit about how things happened there, but if you look down at a molecular scale, could you tell every single motion that was made in, kind of, napping a flint, or every single thing where, you know, every single use of an arrowhead.
Could you tell it by looking at it sufficiently precisely at an atomic scale, at that piece of material?
I think the answer is quite possibly yes.
I mean, with an atomic force microscope, you can kind of pick out where all the atoms are, but it's a very slow process.
we may very well eventually have technology where we can just have an atomic… a solid, and we can just find out where every atom is, and then we have to sort of figure out, well, why are the atoms that way? Oh, there's this little trace of a sequence of atoms that are displaced along this line. Oh, that must come from the fact that there was some particular scratch that was made on this material, or something.
So I think, you know, that's one… one of the things to say is that, sort of, one always gets to have more and more precise information about the past by doing these kind of atomic-scale things.
Now, presumably, there's a limit to that, because the presence of heat, temperature, will jiggle around molecules and atoms.
And probably there's some sort of lower limit to how much, you know, how precisely you can, you can reconstruct things just because of that.
I mean, certainly, if you say, well, this thing that was a solid that was very, you know, we've got this wonderful, you know, gold coin that was minted in the time of Croesius or something, then But it got melted down. Well, you're kind of out of luck if it got melted down, because in liquid, there's kind of all the information associated with the configuration of atoms in a solid is all

gone. And presumably, the same kind of thing happens at a slower rate for solids. I mean, we know, for example, if you look at magnetic tape.

Where what's happening is not a sort of physical change of, of things, but it's more that the, that the magnetization of, little pieces of ferrite, and of the, the,

The atoms and molecules that make up the ferrite.

Those were pointed in particular directions, encoding particular bits on the tape, and over the course of time, sort of random thermal fluctuations, random heat, will sort of kick those atoms around, and you lose the magnetization that encoded that data for the magnetic tape.

things where you've actually etched something out, like you've etched out a piece of silicon to make a microprocessor, you've etched out an old-fashioned CD to make the channel for the laser to track, you know, those things seem like they will last a lot longer.

And there's a very slow rate at which atoms kind of just sort of randomly fill in the gaps between things there and such like. So I think that's the better hope, is kind of look small rather than look large in terms of being able to recover history. I mean, there are certainly funky things that could happen, they're not going to happen for our planet and so on, but

There are… there's the phenomenon of gravitational lensing, where gravitational fields, sort of make light.

deflect light, just like a lens, a glass lens, or something like this. And so, it's now a completely routine thing to see that the gravity of a galaxy will cause the light from things sort of behind that galaxy to be deflected, so you can kind of see things that were absolutely behind the galaxy, but the light was coming sort of at an angle, and then it got deflected in, just like if you have a convex lens, the same thing happens.

So, then the question would be, well, are there such strong gravitational lenses that you could actually have a photon that was going out

towards this gravitational lens, and it's kind of like a gravitational mirror, because the photon will come right back at you. The answer is yes, that can happen. There are actually photons that can be in orbit around black holes, at least spinning black holes, and

the, but that's a phenomenon where the photon has to be really close to the black hole, but you could imagine, it's very kind of science fiction-y, but you could imagine that there's a photon that's tooling out from the Earth, and it goes so close to a black hole that it has so much gravity pulling it around that it's actually pulled around the black hole and made to come right back to the Earth.

So that would be a way of kind of making a very generalized mirror for things, but it's not something that… we don't have a lot of black holes, it seems, in our vicinity, which is probably just for the best.

Let's see…

Someone asked here… Ruthie asks, will future people spend more time in real worlds or virtual ones?

Oh, I don't know. I mean, I think…

You know, we humans are really pretty wired to interact with other humans.

How close do the other humans have to be before we kind of feel satisfied with those interactions?

It probably varies with, I mean, certainly, sort of the neurotypical of us, there's one answer. The not-so-neurotypical, there's a different answer. Sometimes it's easier

for some folks to interact with machines than with people.

I think, you know, I tend to be something of a people person, although I've spent a lot of my life interacting with machines, but somehow I think I wouldn't do that well, I think, if I only interacted with machines.

I think that, in, and this kind of live streaming is an intermediate case.

So, but, you know, I think… in… my… you know.

it is an interesting question whether there's sort of layers of one's interaction with reality that get thicker and thicker. I mean, you know, some of us wear glasses, or contact lenses, or something like that. That's a sense in which we don't have sort of real reality coming to us. It's slightly modified.

If I could imagine that if augmented reality finally becomes real, and they're finally a kind of glasses-style, you know, ways of getting it, I can imagine that many of us will routinely have augmented reality in our interaction with the world.

And, so, you know, so then the world that we're interacting with is annotated or modified relative to the actual sort of world out there if we… if we sort of took our AR glasses off.

So, you know, there's a… there's sort of this different levels of how much is it real reality, and so on. And then the question is, are the things that you're interacting with, are there actual humans you're trying to affect and communicate with, or is it a machine that you're trying to affect or communicate with?

And, I think…

It's, you know, as I say, we humans are very wired to interact with other humans, but what's good enough?

I mean, you know, it… people had imagined before telephones were widespread, nobody would ever, quotes, do business on the telephone.

nobody would ever be able to get, sort of, enough of a feeling of, sort of, of communicating with a person by telephone to, for example, be able to transact business. But of course.

Telephones are good enough, or the modern versions of telephones are good enough, that we do quite well with that.

You know, it's the same kind of thing with video conferencing. You know, how real is video conferencing?

It's sort of funny, because I'm not a big fan of video conferencing. I do sort of meetings with people all the time, and I find video conferencing very distracting, and I prefer to just have audio and screen sharing. I find it both tiring and distracting to have the video. I'm always surprised how tiring I find it, that, I don't know what it is. I suppose it's… it's sort of concentrating on that,

That image in a way that one doesn't have to do if one's… if one's just using audio.

But I think, you know, there's this sort of question, to what extent does, The,

to what extent does, kind of, video conferencing make up for, you know, how good is that relative to the sort of human face-to-face interaction? And then, well, what happens if the thing at the other end of the video conference is a fake human, so to speak?

you know, I'm realizing that I must have been…

When did I first do video conferencing? Must be…

1992, perhaps? It was a funky different time when you would rent a video phone from the phone company.

I did work, more or less. It's a bit… bit kind of jerky, but there was… there was something there.

I mean, video phones had been kind of a… a routine sort of aspect of the future since the 1950s, but they didn't… they didn't… weren't at all sort of consumer usable until the… until the early 1990s.

And I, I, I certainly know that, yeah, I… it, that's right, I can date some of these things very precisely, because I remember that,
when…
that I was supposed to go to some conference, this was a few years later than that, at a time when one of my children was born, and, so I wasn't going to go, and I set up a video conference instead, and that was by that point.
It was mid-1990s. That was, you know, that was something you could readily do, although it was exotic to people.
But okay, so then the question is, you know, do we have, sort of, as good a time sort of interacting with humans by video conference? Then, could the humans be AI humans, and how would we feel about that?
Well, I don't know. I mean, I think, people will… Kind of…
I really don't know. I think that there are some purposes for which, kind of, having that chat with the AI, will be…
sort of useful, and I mean, it's really… when you're having a chat with an AI, what's it… what's in it for you? In a sense, you're really hoping to reflect back
on you, things that you might not otherwise have figured out, or whatever. I mean, in a sense, you know, when you have chats with other humans.
part of it, I suppose, is the concept that you're sort of planting things in that other human, and I suppose the same will be true of AIs. You're planting something in that AI. Hey, AI, let me educate you about this or that thing, so that at some point in the future, you'll be able to come back to me and remind me about this thing.
Or, I'm sort of talking to the AI so that I better understand myself.
By having that kind of interaction.
But I… I don't know. I think, there's sort of been…
Moments when it seems like, oh, everybody's gonna live in virtual reality, and then moments when everybody rushes back, like after the pandemic, sort of everybody rushed back to in-person is really what we want.
I kind of think we're wired for in-person, so unless something very dramatic changes, I think that's the way we're going. Now, again, the layers of
glasses and AR glasses and so on that we have in interacting with what we ultimately believe is kind of the real world and sort of other humans out there. That's an interesting question, what layers there will be, and what layers one becomes kind of used to, or comes to expect.
I mean, it's almost kind of like, you know, it's some inverted version of wearing clothes, so to speak. You know, you come to expect certain kinds of things, and it's like, no, I'm not going to look at the world with my actual eyes, you know.
going to, you know, minimally, I'm going to wear sunglasses. But more than that, I'm going to wear these things that modify the world and change, you know, annotate things and distort things, and so on. Kind of the quintessential rose-tinted glasses. You can imagine the kind of rose-tinted glasses of the… of… you're looking at the world by just sort of blank out
the things, you know, if you're walking through some area, and there's, you know, some garbage dump on one side, and there's a beautiful piece of scenery on the other side, oh, my AR glasses will just blank out the garbage dump, and I'll only see the beautiful scenery, so to speak. And one can imagine that going deeper and deeper, and, you know, it's not clear where that ends.
Anyway, a few thoughts on that.
Let's see…

Well…
A few points here. Some Fem is mentioning, in terms of extinct modes of thinking, that the flat Earth idea is extinct for the most part. Yes. I mean, it is funny, in science.
That kind of forgotten ideas
you know, do come back with a certain frequency. I mean, I suppose in,
Flat Earth is probably… firmly gone.
But…
Although I think about map projections, I mean, in a sense, the Earth isn't flat, but the pictures you get of its surface sure are often flat.
And, so, if somebody had developed elaborate mathematics for the flat Earth, it probably would be being used to make map projections and so on. I don't think they did develop elaborate mathematics for that. But…
Yeah, I mean, there are…
what's an example? I mean, in, well, a lot of the things that are coming back in modern machine learning were things that were developed in feedback control in the 1940s, or were developed in, sort of language modeling in the 1950s and 1960s and so on, and they're back.
quite a lot of the, kind of, well, a lot of the doom thinking about, sort of, the effect of AI was exactly the same in the 1960s, and is now, sort of, back again. Maybe, you know, with the technology being much more real at this point.
Let's see, Oli asks, will made by humans become a luxury label someday? I think it already is, right? I mean, it's, you know, you buy a carpet, and it's kind of like this one was a hand-woven carpet, which has little defects in it, as opposed to a machine-generated one. I mean, this is having things that come from human effort.
Or the natural world, has been…
kind of a thing that makes them feel more meaningful. I mean, whether that's, you know, the artificial diamond versus the diamond that's actually been in the earth for a billion years. The molecular, you know, at the atomic level, they might be identical, but nevertheless, people feel better about the diamond that's been in the earth for a billion years. And similarly, one can imagine situations where… I mean, it's an interesting question for more
formal things. Like, this is a theorem that is a bunch of, you know, formal… a formal statement, and it's like, this was a theorem made by humans, versus this is a theorem generated by computer. Is it…
Is it… is there something sort of fundamentally more, important about the one made by humans? I think that…
It's…
you know, this idea of the provenance of the thing being important, where it came from, I absolutely do think that will be increasingly important. I mean, it's already somewhat important, but it will be increasingly important. And, you know, it's like when you have a piece of artwork. there's a piece of artwork, and it's, you know, it's three splodges of red paint or something. And, you know, to some people, it's like, oh, it looks like that was done by a kid with three splodges of red paint. But then you find out, oh, the artist had this whole elaborate backstory, and this represents the, you know, the moon the day they were born, or who knows what it is.
And that provenance then suddenly makes this otherwise sort of disembodied, irrelevant three splodges of paint.
now much more meaningful to people. And I think that's true of lots of these kinds of things, where you kind of know the thread that goes back to humans, so to speak, from the thing that has

been produced. And so, yes, I think that that's a… people, you know, people care about that. It contextualizes things for people. It makes it, since we humans are the ones sort of doing the observing of things in the world.
It is, you know, we kind of…
care about things that relate to us humans in many cases. You know, I think it's a,
it's an interesting question, when made by machine is chic versus made by hand. I mean, there are certainly some situations in which people are much happier that it was made by machine than that it was, kind of,
kind of done by, by hand. I, I'm, I'm noticing I, I, I, finally now…
have. I've sort of used them before, but I finally now have as, you know, a self-driving car as my main car, and it's kind of interesting to see the extent to which, you know.
My wife thinks the car drives better than I do. I think I'm not a bad driver. But it's sort of interesting how it's sort of surprising
that…
you could be, and I actually sort of feel the same way, the self-driving car does a pretty good job, and it's nice and smooth, and so on, and, you know, that's a place where I would have felt, oh, I want a human driver, you know, that's really the thing I want, and it's like, well, actually, no, I'm pretty happy to have this AI driver in almost all circumstances.
You know, there are bugs occasionally, and I always do the thing of it asks you, why did you disengage self-driving? And you can press a button and record the message, which I… I suppose it's my… from my life as doing software developments and technology development, I feel an obligation to… to make those kinds of bug reports.
Let's see…
Prabh is asking, what are my thoughts on data centers in space?
I haven't done the engineering calculations. My guess is that they are inevitable.
I think…
you know, it's a lot easier to get power in space. It's a little bit of an, you know, you get very hot and you get very cold in space. Whether you're in the sun or not in the sun, you know, whether you're getting radiant heat delivered to you or whether you're sort of radiating out into, into, Sort of dark space, is… can be different temperatures, but,
you know, you get all those… all that energy from the Sun and all those ultraviolet photons and so on that get blocked by the Earth's atmosphere that presumably allow you to have much more efficient power,
electric power generation than you can on the Earth with photovoltaics and so on, after you've kind of gone through the whole atmosphere. Now, of course, all of that assumes that the
cost to put things in orbit goes down and down, and there are probably other dynamics of how this works, and yes, it's a complicated engineering problem, but it feels to me like it's one of those engineering problems that sort of has a path to be solved. It doesn't feel like it's one of those engineering problems where it's, like, going to be, oh, we just can't figure this out for 50 years.
I think it's gonna get solved.
And at that point, you know, it's a detail of whether it's sort of cheaper to have the bulldozers, you know, sort of level the site and put a bunch of concrete down and so on, or to take that server, which
still needs all kinds of thermal insulation and all kinds of complicated things like that, and put it in space and not have to worry about power generation for it. I don't know, it'll be… and the

amounts of money that are being spent on data centers, you know, are large enough that you get to do an awful lot of engineering development for, you know, to make… if you…
If you're going to spend the equivalent of a trillion dollars on a bunch of data centers, then carving out a mere billion
To do a lot of development is, in order to save a factor of 2 in the final cost, is clearly going to make sense.
Let's see… the,
Crystal is commenting. One thing they realized
Was that when they had to deal with realtors, 9 out of 10
All started using text messaging. Only one of them ended up calling, leaving a voicemail.
So, I think people have gone away from verbally speaking, have gone towards communicating via text, although they could be speaking to their phones via speech-to-text. Yeah, so, I mean, these conventions about what people do, they change. I mean, I,
you know, for example, you know, I would consider it very… with…
you know, very few people, you just sort of, like, pick up the phone and call them. It's, you know, it used to be the case that in most kinds of business transactions and the business things and so on.
you just pick up the phone and call people. It's now, you know, you expect to sort of set up the time for the call and so on. You, in…
I mean, the extent to which people use text messaging, I mean, I'm not a big user of text messaging, I don't like it very much. I tend to be a kind of slower operative in the sense that I tend to use email mostly, partly because most of the things, I'm either… I'm either directly interacting with someone by voice.
Or, I am…
kind of, sort of doing a slow interaction by email with a timescale of days to weeks to whatever. And it's, it's… and the kind of… the text and many threads of texts, don't happen to fit very well into my particular form of kind of focusing on things. But, certainly many people are doing that all the time.
I, I think, you know, this idea that,
You know, I find there are certain kinds of interactions that you can have by email, and there are other ones where it's like, it's never going to converge unless you're actually talking to the person, and in even somewhere, it's going to converge a lot faster if you're talking to the person in person, so to speak.
But it's, I mean, it is interesting, the extent to which
these sort of conventions about how to communicate change, quite rapidly. I mean, it's, it's, I think,
And I don't know,
Yeah, I mean, I… it wasn't long ago in my past that, sort of, I would get these little pink sheet pieces of paper that were handwritten phone messages that were left. I mean, that was certainly still happening in the 1990s, and but that all changed.
Let's see, maybe, one more.
question, and then I need to go back to my…
the question here from Brad about,
Verifying computer code and using mathematics.
Oh, this is a can of worms.
I mean…

first point about verifying things is… the question is, what do you want it to do? That is, if you say, I'm going to verify that this does the thing I want it to do, it's like, how do you specify what you want it to do?

That's an important question.

And, you know, maybe if you have a cryptocurrency, you don't want people to double-spend transactions. Maybe if you have a computer system, you want to be able to keep certain data protected from some other… from some kind of access. Maybe you want to make sure that your program won't crash.

these are all reasonable things, but let's say your program won't crash. Well, what if you… but you might also say, but I want my program to be able to deal with data as large as I could specify. Well, eventually, it's going to become so large that it won't fit in the computer, and okay, it might do something more graceful than crashing, but it's not going to work.

And so, you know, one question is, well, you know, when you say we're going to verify software, it's like, what are you verifying it to do? That's the first problem. The second problem is, and when it comes to mathematics.

the idea that mathematics is about, sort of, formal proofs and so on is a bad model of pure mathematics says it's actually done by mathematicians. What mathematics says pure mathematicians, human pure mathematicians do, it's much more a kind of artistic.

kind of conceptual type of thing, where you're imagining different kinds of structures, seeing how they fit together, and so on. Occasionally, you have to roll up your sleeves and show, yes, you can make a formal proof of this, that, or the other. But most of the time, the most interesting mathematics is more a matter of intellectual structure than it is a matter of detailed proof.

Not always true. I mean, there are systems that do automated theorem proving, we have… we have one in Morphin Language, for example, and

you can, in principle, use those things. It is shocking how rarely they've been used for research mathematics. In fact, I believe it's the case that the only example of kind of a… an interesting new research result

That was explicitly found by automated theorem proving, rather than being found by, sort of, intuition, and then… and then just sort of tightened up, or even being known from an existing human proof.

The only example of some sort of research mathematics being found by pure automated theorem proving was a thing I found 26 years ago now, which is the simplest axiom for Boolean algebra, which I found by enumerating axioms and using automated theorem proving.

What is interesting about that is the proof is completely incomprehensible.

And I just did, last year, I did a kind of another study using sort of the latest AI technology and so on of that proof.

And it's still incomprehensible. It sort of dives into this incredibly detailed stuff. It has no kind of human handholds. It's all just pure detail, detail, detail.

And so what do we get from that proof? We find out the result is true, that's nice. We get no insight about mathematics, we get nothing of the kind of thing that pure mathematicians sort of like to and make progress thinking about. And so there are efforts now to formalize

Various kinds of,

of proofs that have been done in the past in mathematics, like the proof of Fermat's Law's theorem, I don't completely see the point of doing that.

I think that you're sort of grinding things down to a level where it's no longer really a human thing.

And it's… yes, I mean, I suppose if one found it wasn't true, it would be, you know, oops. But I think one knows the structure of the argument well enough to know, even if there was something, a little bit of a bug, it doesn't mean the whole system is bad.
So I think… and there are questions that you can ask about, sort of, foundations of mathematics, like, what are the axioms on which that theorem is really based?
But in fact, in these automated systems and so on, and, you know, the axioms are a little bit funky, and it's not… one isn't… I don't think one's going to know, oh, it's based on… just on the axioms of arithmetic, which it probably isn't, by the way.
And then you have the question of, you know, what axioms is it based on, and how many slightly exotic, very non-human, very arbitrary axioms are introduced? For example, in set theory, in the theory of infinite sets, there are axioms there which no human can really have intuition about. And in fact, one even knows that one can set up set theory where the axiom is there or not there. Like, the continuum hypothesis can either be set to be yes, or it's true, or no, it's not true, and both of those are consistent ways of setting up set theory. But we can't decide between them, because there isn't really… it's just… it's an axiom system, it's an arbitrary set of rules, and there's no
right version of those rules, or at least that's the normal way it's thought about. I think I tend to think that there is a kind of human-compatible version of those rules that probably has the hypothesis, the continuing hypothesis, be true, but that's… that's kind of a different view from the last century or so of views about the foundations of mathematics. So, you know, you could imagine, well, we'll get this automated proof
we might get… we might be able to formalize this… this theorem of mathematics. We, you know, what do we get from that? Not really clear to me.
You know, I have to say, there's been a sort of a big effort to make auto-formalization systems, where you can take a human-written proof, and you can try and tighten it up, and kind of ultimately lawyer it, so to speak, and make it so that everything is sort of precisely follows from everything else in a kind of proof-assistant, symbolic system.
I've tried doing this a bunch of times recently with kind of the latest tools. I've not had good experiences. The tools have a nasty habit of cheating, and you, you know, kind of… they say, oh, yes, yes, I've formalized this proof.
You look at it, and it's like, this doesn't look like it could possibly be right. It's too simple for what it's trying to prove. Or you look at it in more detail, and you realize, oh, there's a little piece in the middle here where it introduced another axiom. Oh, by the way, I introduced an axiom. Well, that axiom could be tantamount to just saying the thing you're trying to prove is true, which is, of course, a cheat.
So, I've had pretty bad experiences with this so far.
And I think… so that's… that's one kind of observation. Now, the next question is, if you were trying to do a formal proof
of some non-trivial thing about some computational system, for example, is there, in fact, a… you know, how big would that proof have to be? Is there, in effect, a finite proof?
Well, Godel's theorem and, kind of, sort of very foundational facts about computation, sort of, revealed by things like computational irreducibility.
kind of tell you, actually, it could be arbitrarily hard to prove this particular thing. There might be… you might be trying to say, let me prove it in an infinite number of cases, and it might be that there just isn't a way to do that with less than an infinite amount of effort.

One has to hope, then, that the program one's trying to prove correct is actually, sort of has the property that it is decidable, it has… it has… it does not have this problem that you might have to do an infinite amount of computational work to figure out whether it's true or not.
And the problem is that there's a certain level of programming where things do tend to be decidable, and… but when you do a sort of arbitrary programming.
In a sense, by definition, there's no hope, because if you're doing things that can be computationally sophisticated, by definition, they can kind of escape the box of what's decidable.
So, for example, one place where people look at this a bunch is in type theory, where you say, oh, I've got a program.
And it's like I can have this thing as an integer, or it's an array of real numbers, or it's a thing that returns a pointer to a function that does this, and so on. These we can think of as types of functions and types of data. And there are certain situations… simple enough types
Form a decidable theory, and you can figure out, you can realistically find a proof of limited size that, yes, this function will only return an array of integers or something.
That works in very simple cases.
it doesn't work in many practical cases. I know, you know, in Wolfram language, we have a compiler, and one of the highest tech parts of the compiler is the type inferencer, which is essentially trying to prove theorems about, in this particular set of functions, if you do, I don't know, the,
if you take, you know, the transpose or the reverse of a list, do you always get a thing that has the same shape as a this, and so on. It's lots of little theorems you have to prove.
Well, when it comes to types that don't compute very much, it's just like, this is an array of integers, this is a this, this is a that. But if the type says, oh, it's got to be a prime number or something, where you're going to check that it's a prime, or it's got to be something that satisfies this criterion, so-called dependent types and type theory, all bets are off.
then you quickly descend into this kind of morass of undecidability, and you can't make, kind of, you can't expect to make sort of finite proofs. So then the question is, in practice.
What?
You know, what can you do to verify practical systems? And the answer is, it's pretty difficult.
Like, with cryptocurrencies, no realistic cryptocurrency has ever had serious verification done for its core operations. I mean, there's some sort of sideshows you can verify, but not kind of the core operation of the thing.
There are certainly things that can be done, and have been done in microprocessor design and so on, and in protocol design, where you can sort of bite off little pieces of what's being done, and say, yes, we can prove that.
My impression is that that is a useful practical technique, but it doesn't go that deep.
And, the, the question is, you know, can you go deeper? And the problem is, there's this looming problem of undecidability. And the problem then is, well, in a practical computer system, how much of what goes on
is kind of computationally sophisticated, and how much of it is straightforward enough to be able to be decidable and be validated in some… in some way like that.
it sort of relates also to the question of, sort of, what causes bugs in things. And, you know, in recent times, there's been lots of excitement about LLMs finding, sort of, obscure bugs and vulnerabilities, computer security vulnerabilities, and code, and so on. I mean, I think

My impression, and it's a bit anecdotal, is that part of what's happening there is humans have written this code, and humans make certain kinds of mistakes. There are certain patterns of mistakes that humans make, and the AIs are really good at picking up
On those patterns of human mistakes.
more so than the idea of, oh, let's trace through this very complicated logic and find out this weird thing happens in this corner. I think that that's, so, I mean, I… I… you know, I… I tend to think that the problem of verifying things is a fundamentally computationally difficult problem, and
the…
and insofar as… and it's really a somewhat different problem than the problem of doing mathematical proofs. Certainly, though, it's very different from the problem of doing mathematical proofs the way that pure mathematicians typically do it, because pure mathematicians are thinking in terms of intellectual structures, not in terms of the grains of sand lining up, so to speak. And I think that the,
Even when AIs manage to do proofs.
A lot of what they're really doing well is picking up this sort of
fact from one paper over here, and gluing it into a fact from some other paper over here, and they're kind of building a successive, you know, sequence of lily pads that they can kind of jump one to the next to kind of get to the proof, rather than they're going deep into the sand, so to speak, to fit together all these pieces. That's the kind of thing that automated theorem proving does, the kind of thing that I did for that Boolean algebra proof.
the kind of thing that is very different from what human mathematicians find interesting. I tend to think the problem of doing kind of… the problem of doing serious proofs in mathematics a bit different from the problem of verifying things in, for example, software and so on, and I am suspicious
That there's quite a cliff Of things you can verify, and things where you don't have a hope.
But we'll see. I think that's,
familiar with a bunch of the companies in this space.
It's, that mostly have slightly strange stories behind them.
But, not necessarily bad, but strange.
So, I think,
It's actually… you guys are motivating me, because I have a piece of almost finished work about the theory of bugs, and you guys are motivating me to get that actually finished, which maybe I'll try to do in the next week or so.
Alright, with that, I had better return to my day job, and thank you for all sorts of interesting questions, and
Look forward to chatting with you all another time.
Bye for now.