Hey there, humanoids. This is David Chewmaker here with a very exciting announcement.
Your favorite wrestling podcast feed, The Ring of Wrestling Show, is now going daily.
And you can hang out with me and Kaz on Mondays and Thursdays for the Masked Man show.
And you can join me, Peter Rosenberg, alongside stat guy Greg and Dip, every Tuesday with Chief Pete. And on Fridays, I'll welcome a friend or special guest from the world of wrestling. And on Wednesdays, we have a very special new show called Wednesday Worldwide that you're going to want to check out.
Pay-per-view reaction, one-of-a-kind interviews, fantasy booking, talking about bagels.
That's what we do here on The Ring of Wrestling Show.
Follow the show now on Spotify and do us a favor.
Give us five stars.
And do us another favor and stay mage.
I think listeners of this show know that I am fascinated by this new suite of generative AI, these tools that allow us to create images or articles from word prompts.
And I am fascinated by it.
I'm spooked by it.
I'm confused by it.
I'm excited by AI.
And sometimes I'm purely scared of it.
I think that when we look at tools like chat GPT, we are seeing like the early embryonic signs of something that is really quite strange and extraordinary.
But one of the reasons that I've been holding off on a longer conversation about this technology is that, and I don't really know how to put this, like, I think I have trouble seeing the landscape of artificial intelligence.
There are some topics.
For example, we did an episode on obesity medication a week ago.
There are some topics where the landscape of questions seems relatively clear to me from the onset.
I go, okay, they're these new weight loss pills that have become really popular really quickly.
Okay, where did they come from?
What do they do biologically?
How well do they work?
What's the upshot for the obese?
What's the upshot for people who just want to lose weight?
What's the downside for culture?
How will they change healthcare?
How will they change society?
The question's just like materialized somewhat automatically.
But with AI, I feel like I don't even know what I'm looking at sometimes.
I don't even know what to compare it to.
I saw someone say that chat GPT is like a calculator, but for creativity.
A tool to amplify the speed of writing or idea generation.
And maybe that is perfect, or maybe it's totally wrong.

Like, I feel like I'm looking at a modern art piece, not understanding it.
So I have for weeks been looking for a conversation to help ground me, to help me understand what I consider two of the big unknowns in this space.
Number one, what's the best way to understand what this class of technology, let's call it generative AI, could mean for the world.
And number two, maybe even more concretely, what's the best way to understand how the biggest tech companies, Apple, Google, Microsoft, are going to use this tech to change the world.
Earlier this year, the excellent tech writer Ben Thompson, no relation to me, of the Stratechery newsletter and podcast, wrote this, quote,
the story of 2022 was the emergence of AI.
First with image generation models, including Dali, mid journey, and the open source stable diffusion, and then chat GPT, the first text generation model to break through in a major way.
It seems clear to me that this is a new epoch in technology.
End quote.
When I read these words, a few things come to mind.
One is that, yes, this is someone who I think sees the promise of AI just the way I see it, and that is thrilling to me.
But also it's like, what does that even mean?
What does it even mean that we're entering a new epoch?
E-P-O-C-H, is that how you say it?
A new era of technology.
What does that new era even look like?
So I asked him to come on the show and to describe the landscape of AI as he sees it.
And the following is our conversation.
I really hope you enjoy it.
I think you're in for a bit of a ride.
I'm Derek Thompson.
This is plain English.
Ben Thompson, welcome back to the show.
Thank you. It's good to be back.
I was last on when the news broke about Elon Musk potentially buying Twitter.
That's right.
A lot has happened since then, I think it's fair to say.
Yeah. Elon has had a decade in the last seven months since you were last on.
In a recent article for the Stratechry Newsletter, you said that AI was the story of the year in 2022.
Why?
Well, it's something that's been burbling under the surface.
Like a lot of things in tech, it takes many sort of years for something to come to the surface.
I think it was around 2017 or 18.
There's a group of Google researchers that published this paper.

Now that Dave's escaping me, it's something very clever.
Like all you need is attention or something like that.
That detail, this new sort of capability called the transformer that completely transformed the way that data was sort of processed in sort of for machine learning models
and for AI.
And that, I mean, it's difficult to overstate what a big shift that was.
So that's actually the story, but that's a good example of something that happened five years ago.
Didn't emerge until 2022.
What happened in 2022, I think was a series of things.
So the first thing was from the public perception of AI.
The first was the emergence of Dolly to, again, a good example of why it takes a while for this stuff to emerge.
It wasn't even the first version.
It was the second version where you could make these incredible images and, and this came from open AI.
And on one hand, it was amazing capability.
On the other hand, it sort of fed into the sort of conventional wisdom around AI, which is that it was going to be centralized.
It was going to be these entities that had a ton of power, compute power and data processing ability that would sort of dominate.
Then over the summer came mid journey, which was a startup, an unfunded startup itself funded.
And they had arguably even better or different, very compelling sort of image generation run via discord.
And then the big bombshell in the image generation space was stable diffusion, which was open source.
And it could run on a single GPU on your own computer.
And that was a really big deal because it opened up the possibility that this AI capability is maybe going to be more distributed than we thought,
which changes all sorts of things like what the competitive landscape might look like, you know, what, you know, where is this going to be a centralized thing is going to be much more of a commodity.
So that's the image generation space.
And the thing with images is images are, you know, an image is worth a thousand words, which we'll get to in a moment.
That's actually a very important insight.
Images are so evocative and you can sort of see that and realize that this was a big deal.
The biggest release of all though was, you know, chat GPT, which I'm, you know, all your listeners I'm sure are familiar with.
And again, chat GPT is underneath the surface built on GPT three.
And it's an evolved version of it.
They call it GPT 3.5, but GPT three came out in 2020.
And at that time, if you were sort of plugged in, it was pretty mind blowing what it could generate.
But chat GPT productized that it did.

It used this method called, you know, reinforcement learning with human feedback where basically humans were in the loop, really guiding it towards the, not just controlling what it did or did not say, which is obviously a controversial topic, but also how it said, right, really shaping it.
So it would give you high school quality essay answers, you know, a lot of GPT answers, you have a topic sentence, you have supporting point one, supporting point two, supporting point three to concluding sentence, right?
And people sort of, it's easy to sort of mock that, but you realize there's a reason we teach people to do that for a reason, right?
And it's an effective way to communicate against the point across.
And it was in this super easy to access chat interface that really woke people up.
And I think the reason why 2022 is the year of AI is because it was a wake up publicly about what was coming, even if that stuff that was coming had actually been sort of burbling for a while.
And there's lots of interesting takeaways there about, you know, where this can be done, the importance of products.
You know, the thing about, about the language models is they are more complicated.
You can't really run them locally to the, to the extent or quality you can GPT three.
I think like chat GPT, like one question runs across like 16 GPUs as opposed to one.
And part of that is the thousand word sort of thing. Language is more complex.
It's harder to sort of get it right. And so that is still fairly centralized.
But anyhow, I'm diving into like 47 details in one answer.
It was the emergence of this in the popular consciousness and the real realization that this is a lot closer than people realize that made it such a momentous year.
You wrote a great menu there and we're going to dive into some of those appetizers and entrees in just a second.
I want to begin by saying, you know, I have friends who aren't as interested or aren't as deep in a tech as I am.
And one of their questions to me about AI is often, how is this not just crypto again?
How is this not just the metaverse? And when they say that, what they're saying is, how can you ensure?
How can you promise me this isn't something that people are really excited about for like three months in early 2023.
But like nine months from now, it's not going to be a thing at all.
What do you say in response to that? It's just a lot of hype case.
Well, I think the most damning thing for crypto is the fact that it has been around for 15, 16 years.
And there has yet to be a single demo or use case that's as compelling as any of the ones that came out last year for AI.
And so there is an extent where machine learning has obviously been a thing for a long time.
Transformers, as I noted, have been around for five or six years.
But it is meaningful that the use cases that we have already seen, the demos we've already seen, like the quality of a demo and the degree to which it grabs people's attention is, I think, a meaningful signal.
And the reality is, is that chat GPT is more compelling than basically anything crypto has developed in 16 or 17 years.

And a lot of the crypto sort of use cases are compelling to the extent they are in the back end and hidden from users.
And, you know, and even then there's not really many good demos.
Well, just to jump in right there, this might be too glib or a too glib gloss on the point that you're making.
But one thing that I said to a friend is crypto was a big pool of money looking for a compelling use case.
And chat GPT is a compelling use case looking for a big pool of money.
There's no commercialization yet for chat GPT.
We don't know what kind of business is going to be built around it.
But to your point, the demo is already more compelling to more people in a use case beyond gambling on assets and hoping that they go up.
Then crypto could have produced or has produced, I think, in the last 15 years.
This is not a prediction that crypto will never produce anything in the next 15 years.
But in the use case race, it seems to me that AI is already ahead.
Yeah. And I think you make a good observation, which is that the money aspect has a warping effect.
And I think it's been bad for crypto.
At a fundamental level, what makes crypto compelling from a theoretical perspective
is that in digital, everything is infinitely copyable and infinitely reproducible.
And a lot of the economic value of digital flows from that point.
And you have zero marginal cost production and duplication of content.
And that has all sorts of implications.
So it's kind of like the core insight of Shoshakari, like what's the implication of zero marginal cost content or zero marginal cost information
and applying that to company after company and industry after industry.
And what makes crypto potentially compelling is crypto enforces scarcity in the digital space.
So you can have a lot of the benefits of digital of easy to transfer stuff, easy to move stuff around.
But you can guarantee there's only one of something.
And that is theoretically compelling.
It's sort of like what could be done with this.
But by extension, that's a very small sort of add on to the vast majority of digital, which is and will always be infinitely duplicable.
And it's like, where's some unique cases?
Now, if you want something to be scarce, an obvious sort of manifestation of that is money.
Like money is valuable because it's scarce.
But the problem is you had all this investment, all this attention that was basically naked self-interest.
And like, how can I get rich as soon as possible?
And there wasn't nearly enough or much investment in what, how can I add on this little sprinkle of scarcity to a generally digital product,
which I still think is something that is interesting.
Like, like, can you have, can you have like entitlements that you can carry across from service to service, right?

Where, and you can verify who you are in a way that's independent from any company that could go out of business or change it,
or you can have different things.
Like, there's theoretical use cases.
But again, we're in, I keep saying theory for a reason, because we haven't seen the actual manifestation of these.
Whereas to your point, AI is the exact opposite.
It's the, you know, if you want to get to the scarcity abundance sort of thing,
the long-term output of AI is massive abundance of all sorts of information because it's generated by a computer.
It's not generated by a human.
So it's kind of in the exact opposite direction from crypto in that regard.
I'm going to steal that slash also reference you.
But the idea, yes, that AI is the ultimate expression of abundance,
and crypto is the ultimate expression of scarcity is a really nice way to distinguish these two technological movements.
I said this other context, and I'm sorry for repeating myself for this audience,
but when I look at-
Never apologize.
People don't care what you say as much as you do.
That's really funny.
Sorry, but I want to apologize.
I rescind the apology.
When I look at the genitive AI tools that are producing synthetic content based on prompts,
and that is like in a very abstract way, the most boring way I can summarize what's actually happening here with these AI tools.
When I look at this technology, I think we don't even know what this is going to become.
This is like seeing a tadpole in a uterus and trying to predict if it's going to be a frog or a human or a woolly mammoth.
You don't know.
It could be any of those things.
The biggest breakthroughs might not be on our radar at all,
and Nat Friedman, the writer in VC, made this point very concrete when he was talking to you about a program called,
I think, Refusion, which generates music from text prompts using visual sonograms.
You can be like, you know, write me a 90s alt rock jam and C minor, and it will do that.
It won't be very good, but it will do that.
And this kind of-
It's better than you think.
Yeah, I just used it.
Some of them are incredible. I think Emotional Synth was actually quite good.
Yeah, it makes sense.
Synth is like the entry of computers to music, so that makes sense that that's where it would do

better.
Yeah, I tried to give it something acoustic, and it was not very good at all.
But this comment that you can generate music from text that uses visual sonograms inspired you to write this,
and I want you to respond to it.
Quote, right now, text is the universal interface, because text has been the foundation of information transfer since the invention of writing.
Humans, though, are visual creatures, and the availability of AI for both the creation and interpretation of images
could fundamentally transform what it means to convey information in ways that are impossible to predict.
End quote.
Ben, what did you mean by that?
Well, I think the conventional wisdom, and rightly so, is that these large language models, like GPT, are more important economically than the image models, right?
I mean, images are great. Art is important. All those sorts of things.
But at the end of the day, everything about our economy and the way we work runs on text.
And it's not just that it goes back to the written word, but also the foundation of computers are text, right?
You know, like, you code in text.
And, you know, text was first on the Internet.
You know, the reason, especially in the early years of Shachakari, I would write a lot about what happened in newspapers
and what was happening to that space, because they were the first online, because they were text.
And you go back to the early days of the Internet, and there was very well-bandwidth and hard to transfer stuff.
Like, I remember, you know, being in college, you know, and we were transferring music files via FTP,
and it took forever because they were so large, right?
And the Napster comes along. It's, oh my God, I could get all this music. It still took forever.
And, you know, and then movies. No, no way, right?
And so text was very, very early. And so text is easily transferable.
It's easily storable. It's easily, and a lot of the time we spend in schools for kids
is really about getting them to be able to understand, interpret, and generate text
because that's sort of the core of communication.
Now, because of that, anything text-related is going to be more important, right?
And so given that you start from the premise of analyzing what's going to be the impact on the ecosystem,
you start with the assumption that text is most important.
Well, then you get to things like, well, text actually takes a lot more resources than image generation does.
And it's like, well, okay, that's going to be more centralized players. It's going to be more important, X, Y, Z, et cetera.

But that is a sort of sustaining innovation.
It's like, we're just doing text, but we're doing it better.
And the assumption there is that the existing players will benefit from a sustaining innovation.
But it's interesting to back up and say, what could be a truly paradigmatic shift here?
I'm not sure if I said that word right, but I think you don't.
I think it's good. So now you have paradigmatic, yeah.
So paradigmatic, is that how you say it?
Maybe.
I suffer from the writer affliction of, I know a lot more words than I know how to say.
But where, what happened, like, things that are meaningful, there's usually a V1 that is sort of does what came before in a different way
and sometimes in a better way, sometimes in a worse way.
Then there's a V2 that transforms the way things actually work, okay?
So you go back to advertising online.
In newspapers, you would have text, and next to that text, you would have an ad.
Because that's, we were limited by the capability of a printing press, which had to, like, put that page down one at a time.
And so you get to the web, and how did advertising work?
You'd have text, and next to it, you would have an ad.
And those ads didn't work as well.
There was a lot of talk about print dollars and digital dimes.
And when advertising took off in a meaningful way, you know, Google is obviously one of them was Facebook.
What did Facebook do?
They invented the feed.
And the feed is something that's fundamentally the printing press is incapable of.
You can't generate, on demand, on an ongoing basis, an individualized publication for every single reader.
That's what Facebook does.
Facebook is an unbelievable technological marvel, if you think about it, where billions of people are going onto the site
and every single person is seeing something different.
And it's completely customized to you.
And it's infinite.
You just keep scrolling.
And you get a picture of your nephew, and you get a story from the New York Times.
And all this stuff has interesting bits about the commodification of sort of content.
But one of those bits of content could be an ad.
And that ad, you think about it, now instead of an ad off to the side, you're scrolling your phone.
And even if it's just for a split second, the entire screen of your phone is taken up by an ad.
And it's very compelling from a sort of business perspective.
And it makes a whole bunch of money.
But the point is the effectiveness had to come from transforming what you were doing in a way that

was only possible with the medium that was sort of in place.
So you need this transformational change.
Now you back up, that's relatively small stakes compared to what we're talking about with AI.
Again, the conventional wisdom and the best bet is that AI does all our text-rooted jobs and everything sort of better than we do.
It is very interesting to think about the fact that text is to some extent unnatural for humans.
Humans, as I said in that blurb, are visual creatures.
We communicate with images.
Before we had text, we had hieroglyphs, and we had drawings on the walls of cavemen or whatever it might be.
That's how we process information.
Again, I go back to that pictures worth a thousand words.
That has all sort of manifestations, if you think about it, and what makes the image generation space really interesting.
A lot of people are pretty dismissive of it.
It's like, okay, that's fine.
I'll kill stock photos or whatever.
But if you think about it, if anyone can generate images and sort of generate them on the fly, and you could imagine this is like the long-term metaverse play where you're in a metaverse and it's generated on demand on the fly like a Facebook feed.
It's customized to you and something completely unique.
That's pretty compelling.
That's pretty interesting.
And what if we can start to communicate and our means of interaction and information transfer back away from text back to something that's more sort of human oriented?
Guess what?
Every single entity that's predicated on text is in trouble then, right?
Because they're not going to completely shift how they work.
That's going to be the opportunity for completely new companies, completely new sort of things.
But we're a long ways away from that.
There was so much there, and it might take me a few days to just chew over all the implications.
But let me ask one little follow-up question before we move on to OpenAI and some of the big tech companies.
I just read this essay about the invention of the alphabet and how the invention of the alphabet changed the course of human civilization.
And the alphabet allowed individuals to communicate with each other via text.
It accelerated textual literacy.
But one thing we don't have similar literacy for is the ability to produce incredibly complex images for each other.
I have no idea how to create an incredibly complex image for a friend, a wife.
I'm not an illustrator.
That's actually a very rare skill to be able to do that.
You have to work for a video game company.

You have to take a long time to become an illustrator.
If ordinary people become literate in the ability to create using AI and stable diffusion, these on-demand images,
you're saying we don't necessarily know how that's going to change communication, social media, entertainment.
It might open up entire new vistas in the way that we talk to each other, the way that we text each other,
the kind of media that we consume on our downtime when we're living in this kind of cornucopia of generative AI 3D image creation
that is at the fingertips of everybody.
Is that kind of what you're saying?
That's exactly it.
And I'm not making any grand predictions when or if this will happen or what it will mean,
because to your point, it's kind of impossible to know, right?
No, you're just saying it's weird, right?
It's larval and it's weird.
It's the tadpole and the belly.
You don't know what it's going to be.
And you think about it, right now, if you want to generate a great image,
there's a whole category now of prompt engineering, right?
How do you put in the exact right words to get what you want?
But I suspect it's a somewhat temporary phenomenon and not just that,
but you're going to be able to do it by voice.
So you could generate these images by speaking them into existence because that's actually a fairly trivial problem
that a lot of companies have mostly already solved.
And so you can communicate and speak an image into existence
without even needing to interact with text at all.
It just goes from your voice to images on a screen or in a metaverse or whatever it might be.
Right.
There was a tweet that I saw that said,
the most important programming language of the future is going to be English.
Rather than understand Python or C++, you just tell the computer.
This is generally the kind of software that I want to exist.
And then you can create and edit and optimize and do all this stuff just with the prompting of your voice.
That's different.
That's on the text side and not simply the image creation side.
But you're going to be able to speak images into existence, right?
You're not going to need to type.
All the pieces to do that are there right now.
It's just a matter of putting them together in a way that works well.
And the other thing that's important with all this stuff, this is always the case in tech.

And it makes predictions both interesting and difficult to a certain extent is that whatever products you see today are the state of the art now.
They're not, they're going to be progressing rapidly.
And they progress on multiple angles.
They progress on a sort of the underlying tech sort of perspective.
They progress on the user interface perspective and how you interact with it.
And they progress from the underlying compute perspective as compute continues to get better and faster.
And we have more access to powerful computers.
And so that's why you have to look for what's the trend line, right?
And so if you can, so the trend line is absolutely, I say something in an image appears like all those things are technically possible today.
It's not good enough to be a product, but that is not at all a limitation because those problems will be solved.
You mentioned chat, EPT.
You mentioned Dali too. We should talk about the company that released both of these programs, that being open AI.
What's the most important thing that people should know about open AI and what they and Sam Altman, their CEO is trying to accomplish?
I mean, open AI is an interesting creature, I think to say the least.
Open AI, another one of whose co-founders is Elon Musk, by the way, who did leave the company a couple years after it started.
But it's kind of funny because they're sort of founding thesis.
And I'm simplifying somewhat, but it was basically artificial general intelligence is a big risk to humanity.
So we need to invent it so that we can make sure it's not used poorly, which is, you know, we happen to make a whole ton of money along the way.
That's great. I mean, we need to be in charge because we'll make sure it's used correctly, which is a little sort of, probably feels like a bit of a self-serving sort of explanation.
But there is some, you know, it was founded as a nonprofit because, you know, like actually this is, we believe this.
And so we're going to actually put it in the formation of our company.
And then what happened was they realized that to accomplish this, they're going to need a whole bunch of money.
Because this compute is still very, very expensive.
And so they changed from being a nonprofit to a capped profit company, which is investors would get up to 100x their return.
And then past that point, the money would revert.
And this is sort of like dribbled downstream.
They actually started out running their compute on Google's cloud.
Then Microsoft invested and, you know, as part of that investment moved their compute to Azure.
And basically there's a new deal this week.
And basically the long and short of it is OpenAI, I think the way to understand their goals and

approach.
And again, this is me looking in from the outside.
So I might have, you know, not be quite right, but is they do believe it is possible to have an artificial general intelligence
that is sort of self-directing and can solve world problems and can make great scientific breakthroughs.
And, you know, like all the sorts of things that, you know, if we have a truly sort of intelligent computing entity could potentially do all these good things
and of course could do lots of bad things, right?
If that happens, the value is going to be astronomical, like trillions and trillions of dollars.
And to get there is going to be very, very expensive.
And so the deal they've basically cut with Microsoft is, and this is simplifying dramatically,
but Microsoft gives them the compute they need.
And Microsoft basically, one gets all the profits up to like a hundred billion dollars or something like that.
The number is not quite right, but by and large.
And Microsoft gets to incorporate all these sort of capabilities along the way into their products.
And so you see that Microsoft is going to be building GPT into like Word, right?
So you can just generate your essay right there in Word using this sort of capability.
They're incorporating into Bing so you can like, like get a different sort of search experience with artificial intelligence, which we can get a little bit.
But, and then also you can access the OpenAI capabilities via API on Azure.
And so if you run your stuff on Azure, you have access to the sort of thing.
And their sort of bet is, look, we're not actually interested in a quote unquote small result,
where a small result would as a standalone company be worth, you know, would change the big five to the big six.
We are going for the ultimate goal.
And so we're going to sacrifice all the intermediate goals and basically give them to Microsoft.
Because if we can get to the finish line, we're going to be worth more than anyone else combined.
And by the way, once it gets that big, the money then goes back to the nonprofit.
Like that structure is sort of still in there.
It's definitely kind of weird, but that is sort of OpenAI.
They're shooting for this ultimate goal.
And they found a partner in Microsoft that will fund that and they'll give away all the intermediate benefits to Microsoft along the way.
How catastrophic do you think this deal is for Alphabet?
Do you think it exposes their inability to ship certain AI products?
Or do you think in the medium term, you might see it as being a galvanizer that gets Google to use their, you know, as far as I can tell,
frankly, you know, brilliant AI research table to accelerate the development of their own wondrous products.
Yeah. Well, to be, I mean, Google is generally been thought to be ahead in this space by and large.
And Google will tell you that.

It's like, oh, our image generation is better than those guys.
Our text generation is better than them.
It's like, well, it'd be nice if we could see that Google.
And the reality is they do and have been shipping AI stuff, but it's been in their products, particularly search.
And the reason we know this is not just because Google said that, but you can actually trace Google's costs of good sold, basically,
like how much they're spending on compute relative to the revenue has been increasing for, in particular, the last five to six years.
And that's not the payments to make to Apple. That's somewhere else on their sort of income statement.
This is the actual cost of the computers they're running to provide their services is going up.
And it's like, why would that be happening?
Generally, you would think costs would be going down and it's going up faster than revenue.
And the reason, the likely reason and the one that makes sense with both what they're saying and what's happening in tech is they are like,
AI compute is more expensive.
You have to like, you have to run these GPUs.
You have to do a lot more compute.
And so they are spending more relative to what they're delivering.
And they say that's for AI and that makes total sense.
And so the assumption should be that they actually are doing a lot in this space.
It is showing up in their products, is just showing up in their products sort of as they exist.
And a lesson I've learned over the years as someone who was hilariously wrong about Google, early in the mid sort of mobile area.
Remind me, how are you hilarious?
Wrong about Google?
Oh, I thought that, you know, they sort of peaked in sort of importance and relevance and their profit is like TEDx since then.
It was an incredibly dumb take.
But what I got wrong is, you know, just the power and importance of distribution and ownership.
And, you know, and that influenced my later writing, right?
And sort of talking about things like aeration theory.
And if you control demand, it gives you all the leverage in the marketplace and things like that.
And Google, you know, with Android and growing, you know, search on mobile was a huge tailwind to them.
And, you know, the way I think about it is I was totally wrong.
But the best way to be right in the long run is to admit when you got wrong and try to learn from it.
So I hopefully I accomplished that sort of in this case.
But it does make me gun shy here because so the front, the problem facing Google with things like what like chatGPT is.
And this is the where companies get in trouble.
It's not a technical issue.

We it's totally, I think we should assume that Google stuff is better.
They have more resources, they've been working on it longer.
I mean, they invented the technology for, for, you know, for goodness sake.
The problem is a business model problem.
And that's where companies do get in trouble.
The way Google makes the vast majority of their money is when you do a search,
they're running an auction for all those ads on the page.
And who wins an auction?
Well, usually, you know, in a traditional auction, it's whoever pays the most.
And that was how web advertising used to work, right?
I pay for these number of pressions, XYZ.
What made Google advertising so brilliant was they didn't pay just for showing ads.
You could show all the ads you wanted to for free.
You paid Google when someone clicked on that ad.
And so that would make you more willing to pay because you're only paying for success.
You're actually getting a customer or a potential customer, I should say.
And so in the case of a Google auction, the winner is decided by the user.
The user is displayed a bunch of search results and a bunch of ads,
and they pick the winner by clicking on it.
And when that, and when they click, then the, that's when the advertiser sort of sort of pays Google.
Now, there's lots of games that are played at this, right?
Google's ads have become increasingly indistinguishable from results.
There's so many ads, particularly on mobile, that companies who are the first results feel the need to buy an ad so that they're still clicked through.
Or sometimes I'll search for Levi jeans.
And the first thing that will come up is an advertisement for Levi's, which of course I'll click on.
But like, I wanted to click on Levi anyway.
I Google Levi jeans.
So there's a little bit of a cheating there, but yeah.
For sure. For sure.
A lot of the revenue is like that, to be totally honest.
This is the case for all search results.
The same thing like App Store ads, right?
Apple makes a bunch of money on App Store ads that are search results
that is what you would have found anyway, but you're going to click what's top of the list.
Search advertising, I think, is very compelling from a business perspective
and a little less compelling from a benefit to society and user perspective, to be totally honest.
With Google though, the problem with a chat interface that sort of works well is
there is no room for that auction to happen, right?
The whole idea is it's giving you an answer.
It's not presenting sort of a list of options.
And yes, theoretically, you could put the old kind of advertising where you slap an ad in there,
but the entire point of Google and the Facebook feed and things along those lines

is having something that fits in the context that is unique to you and compelling
is dramatically more valuable than just throwing an ad in there.
And from an advertiser perspective, someone clicking on your ad is worth more
because now they're in your website and you can get them to sign up.
You get their email address, you start remarketing to them, right?
So that initial Google click is more valuable than just sort of a generic ad
that might prompt a transaction, but you don't know who the customer is.
And that's the big challenge for Google is they still make the majority of their money.
The most profitable business is search. The fundamental way that search monetizing works
feels incompatible with these sort of chat interfaces where it just gives you the answer.
That's one of the first things that I experienced when I was using chat GPT
and thinking about the way that it lived alongside Google
or might theoretically provide a challenge to Google.
It was the same language. I thought of you have search engines which give you links
and then there are answer engines that give you answers, paragraphs.
Sometimes even essays. And yes, the answer can be bullshit on chat GPT,
but the truth is that sometimes the search results are bullshit too, right?
Bullshit can sometimes just be an intrinsic ingredient in the digital experience.
But I've never thought of it quite like this, that if you provide an answer,
you cannot introduce links that are advertisements.
You have to find some other way to introduce advertisements around the answer,
whether it's an interstitial ad which everybody hates because they're extremely annoying, right?
The click here to see the answer.
Or you do this weird thing where maybe part of the results page is an image of an advertisement,
like a branding advertisement, and you have to click a button to make it disappear
so you can see the entire answer being provided by chat GPT or Bing.
But it requires an entire rearchitecturing of what search is online.
Well, not just that, but it's a crappy experience for everyone.
It's obviously a crappy experience for the user, but it's crappy for the advertiser.
Because what makes Google ads so compelling is the user themselves decided to look at it,
decided to follow it, as opposed to it being shoved in their face.
And so in this case, the vast majority of Google's money is made on a very small number of searches.
Things like travel, insurance, e-commerce, stuff you want to buy.
And you can tell by just the number of ads on the page.
If you're getting five ads, that's probably a very profitable search term for Google, right?
There's other stuff you search for, like a Shakespeare quote, there's no ads.
And the vast majority of Google's searches are of that type.
And so my best guess as to how Google will respond is they will introduce chat interfaces
for the vast majority of queries that don't make money,
because those queries are important because they keep people using Google
because it's valuable and gives them answers they want, while retaining the sort of different format
for stuff that actually does make them money.
And given their position and the fact they have so much distribution, that will probably work.

And because I think it's safe to assume they have the same capabilities
and probably better capabilities than OpenAI or Bing or whatever.
But the speed with which they decide to do this or respond is going to be sort of very interesting to watch.
And the bullish case for Google is, look, we know they can do this technically.
They've figured this out before how to adjust search to a new sort of paradigm.
They'll probably do it again, and their existing advantage in users' habits and where they're existing will be enough to carry them through.
But at least there is the outline of where you could see some serious disruption for them.
That's a very clear frame. I think I'm going to remember that, that Google's version of chatGPT might be rolled out for, let's call them, the cheap queries.
But for the expensive queries, I want to buy a car. I want to buy jeans. I want insurance.
I want to go shopping. I want to fly to Singapore.
Okay, for those, you're not getting anything like chatGPT.
You're getting the OG Google experience.
I don't think I want a chatGPT answer, right?
I don't even know what chatGPT's answer to the question would be.
Hey, I found you a flight on Singapore Air?
Right, I don't trust anyone to book my flights. I'm not going to chat and trust the computer.
I have very specific preferences.
I have a question that might be stupid and maybe even cheap that brings us to Amazon and Apple.
But that's that I have an iPhone. I have Alexa.
I don't think Siri or Alexa are very good.
At least they aren't very good at interacting with me and giving me what I want.
And one of these ironies of living in this early...
It could be a you problem, Derek.
It could be a me problem. That's why I brought someone else on to talk about it.
This is not putting the monologue.
It is kind of interesting to me that Siri doesn't seem very good,
but we're also in this golden age of AI.
And I wonder whether, A, you just totally reject the premise
and you're like, Derek, this actually is a you problem. People love Siri.
Or B, whether you have ideas for how these LLMs, these generative AIs
could be insinuated into our iPhone experience the next few years
in ways that we can't even see yet.
Yeah. No, Siri stinks. You're right. You're right on that.
And the main part of the premise of your answer I would reject
is that we're in the golden age of AI.
We are in the first inning.
And I think that's an important distinction.
Number two, in my estimation, Google's chat voice experience,
Google Assistant, is way better than Siri.
And I think it's better than Alexa as well.

And that is actually probably the most concrete manifestation of Google's AI capabilities that's out there.
You can have ongoing conversations with Google.
I'm sure you used it, but it's a much different experience than I think with Siri in particular.
It retains information better. You can go back and forth.
It understands you basically perfectly every single time.
And it's a good example of where Google is strong in this space.
And also, it's an example of Google taking risks and pushing forward because I wrote about this in the context of Google Assistant when it first came out, like, this is a business problem for Google.
That's the exact same issues I talked about.
The moment Google Assistant is inserting information based on who paid them to do it is the moment the entire value proposition sort of falls apart.
But Google has invested in it all the same.
And that's a credit to Google.
I think so, so it's, I don't know, your overall premise of the question or your overall question is a fair one.
You know, and I think the, in general, in tech, everything is additive.
It's very rare.
There's a popular narrative.
When tech first started, you had, you go back to like the piece, like you had the mainframe.
They did these things called mini, what are they called?
Mini computers.
I don't remember.
Like you had, like, long computing.
Like it was a big company that no one remembers.
This one, Boston was like the center of like the computing universe.
And then the PC comes along and queens them all out.
And that established this narrative in tech about new products coming along and wiping out the previous generation.
And that hasn't really been the case for 20 years.
Everything is additive, right?
The PC is still around, right?
Is it as dominant as it was?
No, it's been eclipsed by mobile, but we still have PCs.
We're both sitting on PCs right now conducting this because they're more capable and beneficial for use cases than a phone would.
We could do a podcast via the phone.
But why?
When we have a better, more capable machine right here.
And I think this will be the case here.

Like the fact that like search and links and research, that doesn't necessarily mean we're going away.
We might have something better.
That's on top of that.
Same thing with sort of voice interaction.
Sometimes text interactions better.
Sometimes like image communication might be better.
And that's probably a safer, better framework in general to think about this stuff is don't get locked into it being a zero sum game.
Again, to go back to the very beginning of our conversation, we're talking about massive abundance.
Scarcity is a world where there's zero sum, where there's replacement, where you either have A or you have B.
We're talking about a world where you have all the information that you want.
That could be presented in a million different ways, depending on context.
And, and a lot of mistakes and analysis are made with the assumption that stuff zero sum and the reality it probably isn't.
Well, like we do, you think about mobile.
Mobile didn't reply.
Like we don't sit at our desks with little phones doing all our work.
We're still at a desk with computer.
But now we can also be sitting on the toilet with a computer, right?
Because it's a, there's a new sort of interface that goes with it.
When it comes to the voice stuff, we can be in the kitchen with our hands all dirty, you know, cooking something and say, Alexa set a timer and it's like, well, is that really beneficial?
Well, you're now doing more computing than you did in the past.
Is it replacing your mobile phone experience?
No, is it replacing the PC experience?
No, it's adding on top of it.
And that's how this stuff is going to sort of, it's particularly in the, you know, the first short to medium, even to longterm is how it's going to manifest.
Really quick question before I ask you about some of the dark sides here.
You know, I'm not trying to get you to just summarize the movie her, but I am kind of interested in the interplay between Apple AirPods and AI.
What happens when you have something in your ear, on your face, that you can talk to and that can talk back to you, combined with what you've been discussing for the last 45 minutes, this new dawn in an ability to generate synthetic intelligence, synthetic content on the fly.
Do you have any thinking about what that combination could create?
What horizons that could open in the next few years?
I mean, I do think her is very compelling.

I've been referencing and talking about it for years.
I think, you know, it just the, this idea of sort of an assistant that's with you all the time.
And one of the powerful things about these transformer based models is they learn very quickly.
They don't have to like iterate a million times and sort of nudge them in a direction.
They have this sort of one-shot learning where you give them a result,
they immediately incorporate it and sort of understand, you know,
and keep that in mind sort of going forward.
Now it's interesting because we do have technical limitations to getting to there.
So if you go on chat GPT and you have a conversation,
it's like, wow, it's amazing how it has memory, right?
It knows stuff we talked about before.
That's a total hack.
Every single submission to chat GPT is fresh and new.
What it does is it sends your entire past history.
So when it generates the new answer, it has all the context and it generates a new,
all that sort of thing, you know, to actually have persistence.
Like this is where a big question about centralized versus local comes in.
Like there's real compute costs, there's storage costs, there's memory costs,
all of which are going to be real challenges to sort of overcome.
Like if you want to hear that knows you and it's extensive,
like you have to store that information somehow
and you have to be able to retrieve it quickly and efficiently.
And so there's actually a huge amount of technical barriers to getting to that.
But again, technical barriers will be overcome.
If we've worn one thing in technical for 40 years,
that's not a gating factor per se.
There's also weird physiological limits.
It turns out that if you have something in your ear for too long,
you will get a bad reaction to it.
You know, we're, again, we're talking about having something in your ear
basically 24 seven, like all the time, right?
And so, you know, is it going to be some sort of implant?
Is it going to be like, so like, how is that going to work?
Obviously, do I want to be speaking in public all the time?
How am I going to communicate?
You know, is it going to capture my thoughts?
Like that, now that stuff is, that's not like, we know how to do it
and it's going to get better.
That's speculative.
Like, are we going to actually figure out how this sort of stuff would ever work?
But I do think by and large, this broader idea that we all would have an assistant,
that does know us because we've interacted it for days, months, weeks, years.
I think I got my order wrong there.

But, you know, it is like that is something that is plausible.
There's things that need to be worked out to do that.
And you can imagine how beneficial that could be, right?
Where you just sort of have, you can do things, the distance between having an idea
and the execution of that, whether that's generating an image
or writing a paragraph or booking a flight or whatever it might be, is dramatically compressed.
That's certainly a vision and a place that people are going to push to go to.
You gave me one haunting thought, which is that just this week,
there was a breakthrough in, what is it, a brain interface technology
that allowed people to produce words on a screen simply by thinking them
at a rate four times faster than any previous technology to achieve this.
And it made me think, maybe the future is her as a silent film, right?
That instead of talking to Scarlett Johansson, you don't even have to say anything.
You just think it and the little bud in your ear interprets what you have thought.
And he says, oh, Derek just thought, make a reservation for Derek and his wife, Laura,
at Ellie at 8 p.m. on Friday.
And it just does that.
And I don't know, now we're getting like 20 years out in the future
and I'm not going to ask you to comment to intelligently on pure speculation.
But that could be one combination of two different streams
that people aren't necessarily bringing together here,
this sort of brain interface technology and Apple Pod AI technology.
Yeah, no, people are definitely thinking about that, I think.
But it is to your point, it's some point in the future.
There's a lot of things to be solved to sort of get there,
but certainly that is a vision for sure.
Last question for you.
There are a lot of people who are really afraid of AI
and in particular afraid of what AGI will accomplish.
I'd like you to do two things to close us out.
I'd like you to help us understand how you fear artificial intelligence.
And I'd love to know what you think AGI, artificial general intelligence,
actually means and whether you think it's something to be afraid of.
There's a reason why films like Terminator or whatever exist, right?
Like this idea that once you have an AI that, you know,
first off, there's a few questions here.
Number one, is AGI, what is AGI?
There's a big question here.
We devised things like the Turing test,
which chatGPT can obviously pass, right?
There's a criticism of chatGPT that's like a very confident poster
on an internet forum that's totally wrong.
And implicit in that analogy is we have this archetype.

It's humans, right?
Humans are bullshit all the time.
They make stuff up.
They sound very confident.
You know, the fact that GPT is basically built from the internet,
and that was a big thing that was enabled by transformers
instead of having these carefully labeled data sets,
you could just scour the internet and get all this data
and sort of glean sort of like how stuff fits together.
I mean, chatGPT is not thinking.
It's basically running probability of,
given this question in this context,
what's probably the answer that suffices to this?
You are describing one answer to my question,
which is that, you know, one dark side of AI
is that it will confidently produce bullshit.
I mean, the internet already confidently produces bullshit.
But to the extent that these large language models are simply,
you know, like speedrunning that synthesis,
they're just going to produce even more confident bullshit.
Which, by the way, is another sort of potential danger for Google.
You know, they rely on the internet having the answer.
And if there's the more bullshit there is,
the better they're going to have to do in distinguishing that.
And I think a lot of people, I certainly agree with that,
feel that Google search is not as good now as it was previously,
in part because there's so much gaming of the system, right?
That's a hard challenge for them to solve.
You're dealing with an opponent that is unknown,
unnamed and has zero marginal costability
to sort of generate stuff, right?
You have actually like the Macedonian teenagers
generating fake stories on Facebook, right?
Doesn't cost anything to do.
Like you can dupe, when you had to have a printing press,
there was a cost barrier to generating an output.
The reality is, is that we as humans are very complicated creatures.
We're filled with biases.
We want things to be true that aren't necessarily true.
And, and that is the source of so much confidence.
I mean, arguably the conflict on the internet
is not that people suddenly became full of misinformation.
It's that we suddenly became exposed to lots of people

that think about the world very differently than we do.
And that's very upsetting.
That's a lot of the consternation I think people have online.
It's not that cranks didn't exist previously.
It's that we weren't exposed to them.
And there are real downsides here where if you were,
if you had an abhorrent opinion,
you were probably the only person in your area that did
and you felt isolated.
Now you can find a community of abhorrent opinions online, right?
But all these issues are pre-AI.
Like I think a lot of the concern and consternation
with the internet generally is going to be projected onto AI
when reality, they're all human problems that emerge
when we're all in the same place at the same time, right?
And now AI is dumb.
It has no, JGB doesn't know what it's producing.
It's all based on this feedback.
Now I mentioned this reinforcement wording with human feedback
that is a reintroduction of sort of shaping what it says
on top of this sort of internet corpus, right?
And so we started saying all these datasets
have to be super highly labeled back in the early machine
running area to, wow, we can just use the internet.
This is amazing to, well, for people to actually enjoy using this,
we do need to reintroduce the editing function,
a sort of like shaping sort of function.
And so, you know, the answer there is sort of in the middle,
but it's still dumb to actually be creative,
to actually generate and to actually be sort of like sentient
as however we might define it.
It's still not clear that number one, it's going to get there.
Now, a lot of people are you look, the all of, you know,
you think about like text, for example, right?
To date, the generation of text is a deliberate act.
You have to, yes, it's cheap and easy to do
and duplicate on the internet,
but the actual generation still needs to be done.
When I'm writing an essay on Shritaqiri,
when you're writing on the Atlantic,
how many thoughts, how much processing
was never actually put down, right?
Probably 99%.

Exactly.
And even if these AIs are exposed to all the textual input
of humanity, you just said they've actually only been exposed
to 1% of the thought.
And so, what is the line and bridge to actually being able
to incorporate and understand that in a way that does become,
you know, yes, it's dumb and has no thought
in his probabilistic base, but it's just as good as a human.
Is it enough to accumulate all the right knowledge
in the world?
Maybe, I don't know.
And so, the question is number one,
are we even going to get to AGI?
Number two, if AGI comes, what do we do about it?
I actually have a very kind of, I don't know if it's cynical
is the right word.
No, it's not cynical.
It's like a resigned sort of view of it.
I think a truly sort of like bad acting AI of the sort
that's imagined and terminated by, you know,
the people that are really concerned about this.
I don't think there's anything we can do about it.
I think if it happens, we're screwed.
And so, there's a certain aspect of, you know,
think about Apple and their China operations, right?
If China attacks Taiwan and all Apple gets cut off
from China completely, they're screwed.
So there's kind of an aspect of, yeah,
they're trying to diversify on the edges,
but, you know, it's just not a functional way
to think about their business because it would be
so astronomically expensive to undo it
and unwind it immediately that they're just going to kind
of hope it doesn't happen.
And I kind of have a similar feel here.
I am a little more skeptical.
There's people in the space that are sure it's almost here
if coming.
I'm a little more skeptical for that reason.
I don't think it's particularly like, I do think there's so much
more to actual generation.
And maybe I'm biased.
This is my human, my narcissistic,

solipsistic humanity bias here absolutely might be the case.
But I also sort of feel like, well, if we get there,
we're screwed.
So I'm not sure how much it's worth to,
given the realities that this is out there,
that we have competition in the space.
China is investing just as much as we are investing.
But I'm not sure working against a potential outcome
that we probably can't do anything about is worth limiting
and stopping the massive potential benefits
that are downstream of this.
Ben Thompson, sure.
Thank you very, very much.
Good to talk to you.
Thank you for listening.
Plain English is produced by Devin Manzi.
If you like the show, please go to Apple Podcasts or Spotify.
Give us a five-star rating.
Leave a review.
And don't forget to check out our TikTok
at plain English underscore.
That's at plain English underscore on TikTok.
Thank you.