I'm Ezra Klein, this is The Ezra Klein Show.
So this is a bit of something different, a bit of an experiment.
I got two parts to my work here at The Times, I do the podcast and I do my columns.
And the podcast is really where I explore what other people think, the columns is where I work out what I think.
But sometimes I want to cross the streams a little bit more than I do because there's a tendency for people to think that whatever podcast I did last on the topic is what my actual view is.
We're going to be covering AI a lot, you might have seen GPT-4 just came out, that's a big deal, what it can do is a really big deal, we're going to have a big conversation around that on Tuesday.
But I just wrote a column trying to work through my own thinking on AI and why it's come to dominate so much of my thinking.
And I thought I might read it here to give people a little bit of context for the coverage we're doing and where I'm coming from as I do it.
So in 2018, Sundar Prasai, the chief executive of Google, and I would say he's not one of the tech executives known for constant overstatement, he said and I quote, AI is probably the most important thing humanity has ever worked on.
I think of it as something more profound than electricity or fire.
That's a hell of a quote, and I was back in 2018 when these systems were a lot weaker than they are now.
Try to live, I've been trying to live for a few minutes, in the possibility that what he's saying is true.
People talk about human cognitive biases all the time, but there's no more profound human bias than the expectation that tomorrow is going to be like today.
It's so powerful because it is almost always correct, tomorrow probably will be like today and next year probably will be like this year.
But try to cast your gaze 10 or 20 years out.
One thing that's been unnerving me, and I think this is partially me as a parent thinking about the world my kids will grow up into, is that I don't think I can in a way that doesn't feel that common in human history.
When I think 10 or 20 years out, I really don't know what kind of world to envision.
Real intelligence is a loose term, and I mean it here loosely.
I'm describing not the soul of intelligence or understanding, but the texture of a world populated by programs that feel to us as though they are intelligent and that shape or govern much of our lives.
I really think that's the important thing, not whether it is intelligent, but whether to us it seems so.
Such systems, if you're using some of these, they're already here, and this is the part that I really want to emphasize and underline.
What is coming, and it's coming fast, it is going to make them look like toys.
What is absolutely hardest to appreciate in AI is how fast the improvement curve is.
There's a comment by Paul Cristiano, who was a key member of OpenAI, and he left to found the Alignment Research Center that I think about sometimes.

The broader intellectual world seems to wildly overestimate how long it will take AI systems to go from large impact on the world to unrecognizably transformed world.
This is more likely to be years and decades, and there's a real chance that it is months.
What he's saying there is not that it is all going to happen in 10 months.
What he's saying there is that when these systems get powerful enough, they can already write code, they will begin improving themselves, and so when you get to a system so powerful that it's the kind of thing we would want to turn over a lot of power to, that system can go to unimaginably powerful, potentially very quickly.
People call this a take-off, or the intelligence explosion, and maybe it's wrong.
Maybe it doesn't happen.
Perhaps developers hit a wall they don't expect.
But that's not a plan.
What if they don't?
I've been trying to think through what this feels like to me, and spending a lot of time reporting with these people, and understanding the sort of exponential nature of the curves of the improvement.
I keep finding myself thinking back to the early days of COVID.
There were these weeks when it was clear that lockdowns were coming, that the world was tilting into crisis, and yet everything felt pretty normal.
And you sounded like a loon, I sounded personally like a loon, telling your family they needed to stock up on toilet paper.
This is a difficulty living in exponential time.
It creates this impossible task of speeding policy and social change to match the rate with COVID, a viral replication.
And I think that some of the political and social damage we still carry from the pandemic, it reflects that impossible acceleration we had to make.
There is this natural pace to human deliberation, and a lot breaks in our society when we are denied the luxury of time.
But I think that's the kind of moment we're entering now.
We do not have the luxury of moving this slowly in response to these technologies, at least not if the technology is going to keep moving this fast.
So I moved to the Bay Area in 2018, and one of my personal projects while I've been here has been to spend a lot of regular time with people working on AI.
I always thought this was the big technology story of this era.
And I have, and I don't know that I can properly convey to you just how weird that culture is, just how weird the culture that people building these powerful systems are.
And when I say weird, I don't mean it dismissively.
I mean it descriptively.
This is a community of people living with a completely altered sense of time and consequence.
They are, or they at least believe themselves to be, creating a power that they do not understand at a pace they often cannot believe.
This is just wild to me.
There was a 2022 survey of AI experts, and they were asked, quote, what probability do you put on human inability to control future advanced AI systems causing human extinction

or similarly permanent and severe disempowerment of the human species?
So that's pretty extreme.
The median reply of these researchers, the people creating these technologies was 10%.
The median reply was 10%.
On the one hand, I find that hard to fathom.
And then on the other hand, I have sat for hours talking with people who put that probability much higher.
And it gets to how weird this culture is.
I mean, would you work on a technology that you thought had a 10% chance of wiping out humanity?
Would that be something you'd spend your time on?
We tend to reach for science fiction stories when thinking about AI.
And I've come to believe that the correct metaphors are in fantasy novels and occult texts.
My colleague rushed out that had a good column on this.
We talked about it as an act of summoning.
The coders casting what are basically literally spells, right?
They're strings of letters and numbers that if uttered or executed in the right order, create some kind of entity.
They have no idea what will stumble through the portal.
And what's oddest in my conversations with them is that they speak of this completely freely.
They're not naive in the sense that they believe their call can be heard only by angels.
They believe they might summon demons and they're calling anyway.
And I often ask them the same question.
If you think calamity is so possible, why do this at all?
And different people have different things to say, but after a few pushes, I find they often answer from something that sounds to me like the AI's perspective, many, not everyone I talk to, but enough that I feel comfortable making this characterization.
They'll tell me they have a responsibility, they feel they have a responsibility to usher this new form of intelligence into the world.
Now, there is a gulf here and I suspect I've just fallen into it for a lot of you.
The very tempting thought at this moment is these people are nuts.
And that has often been my response.
Perhaps being too close to this technology leads to a loss of perspective.
Loss of people writing this code are not the people you should trust.
Look, this was true among cryptocurrency folks in recent years.
The claims they made about how blockchains would revolutionize everything from money to governance to trust to dating, I would say they never made much sense, but they were believed most fervently by those closest to the code.
Maybe they knew something we didn't, or maybe they just lost all perspective.
I think in the crypto case, it was a ladder.
As you might say then, is AI just taking crypto's place?
Is it just now a money suck for investors and a time suck for idealists and a magnet for hype men and a hotbed for scams?

And I don't think so, or at least I don't think it is safe to assume so.
Crypto was always a story about an unlikely future searching for traction in the present.
With AI to imagine the future, you need only look closely at the present.
Could these systems usher in a new era of scientific progress as a booster's hope?
In some cases they already have, in 2021, there was a system built by DeepMind that managed to predict the 3D structure of tens of thousands of proteins.
This was a breakthrough so remarkable that editors of the journal Science named it the breakthrough of the year.
Will AI populate our world with non-human companions and personalities and become our friends and our enemies and our assistants and our gurus and perhaps even our lovers?
Well, look at this piece that was just published in New York Magazine, quote, within two months of downloading Replica, which is a chatbot, basically, a chatbot companion.
Denise Valenciano, a 30-year-old woman in San Diego, left her boyfriend and is now happily retired from human relationships.
I downloaded Replica, by the way.
I didn't find it that interesting, but again, it's going to be so good so soon.
Can AI put millions of people out of work?
I mean, of course, automation already has.
It's done that again and again.
Could it help terrorists or antagonistic states develop lethal weapons and crippling cyber attacks?
These systems will already offer guidance on building biological weapons or nuclear weapons if you ask them cleverly enough.
Could it end up controlling critical social processes or public infrastructure in ways we don't understand and ways we may not like?
Well, it's already being used for predictive policing and judicial sentencing, so yes.
And there are two points to make about all this.
One is that this is happening now with very weak systems compared to what we're going to have in five years.
If we are giving them all this power and connecting to them so emotionally now, think of what is coming.
Try to think of what is coming.
But the second point, and the bigger one to me, is I don't think listing the obvious implications here does all that much to prepare us.
We can plan for what we can predict, though I think it's telling that we have not planned for most of this.
But I think what's coming is going to be much more unpredictable.
It's going to be much weirder.
And I use that term here in a very specific way.
In his book, High Weirdness, Eric Davis, the Historian of California and Counterculture, he describes weird things as, quote, anomalous.
They deviate from the norms of informed expectation and challenge established explanations sometimes
quite radically.

And weird in that sense, that's the world we're building here.
I cannot emphasize this point enough.
We do not understand these systems.
We do not understand them.
Not even the people building them.
And it's not clear we even can.
And when I say that, I don't mean that we can't offer a high-level account of the basic functions.
Let me try.
These are typically probabilistic algorithms trained on digital information that make predictions about the next word in a sequence or an image in a sequence or some other relationship between abstractions that it can statistically model.
But when you zoom into the specifics of what it is doing, that picture, it dissolves into computational static.
Megano Gieblin has this brilliant book, God, Human, Animal Machine, which I highly recommend.
And she writes in it, if you were to print out everything the networks do between input and output, it would amount to billions of arithmetic operations, an explanation that would be impossible to understand.
That's the point.
That is the weirdest thing about what we are building.
The thinking, for lack of a better word, is utterly inhuman.
What we have trained it to present is deeply human.
And the more inhuman these systems get, the more billions of connections they draw in layers and parameters and nodes and computing power they acquire, the more human they come to seem to us.
The stakes here are material about jobs we have or don't have and income and capital and they're social about what kinds of personalities we spend time with and how we relate to each other.
And they're metaphysical, too.
Gieblin observes, I quote, as AI continues to blow past us in benchmark after benchmark of higher cognition, we quell our anxiety by insisting that what distinguishes true consciousness is emotions, perception, the ability to experience and feel.
The qualities, in other words, that we share with animals.
This is an inversion of centuries of thought in which humanity justified its own dominance by emphasizing our cognitive uniqueness.
We may soon, arguably we already are, find ourselves taking metaphysical shelter in the subjective experience of consciousness, the qualities we share with animals, but not so far with AI.
Ogieblin writes, if there were gods, they would surely be laughing their heads off at the inconsistency of our logic.
If we had eons to adjust, perhaps we could do so cleanly, but we don't.
The major tech companies are in a race for AI dominance.
The US and China are in a race for AI dominance.
Money is gushing towards companies with AI expertise.
They're going faster.

To suggest they go slower or even stop entirely, it's come to seem somehow childish.
If one company slows down, well, look, another is going to speed up.
If one country hits pause, the other is going to push harder.
Fatalism becomes the handmaiden of inevitability, and inevitability becomes the justification for acceleration.
I think Katja Grace, who's an AI safety researcher, summed up the illogic of this position really well.
Slowing down, she wrote, would involve coordinating numerous people.
We may be arrogant enough to think that we might build a god machine that can take over the world and remake it as a paradise, but we aren't delusional.
I think one of two things must happen, or should happen.
Security needs to accelerate our adaptation to these technologies, our governance to them or of them, or a collective, enforceable decision has to be made to slow them down.
And even doing both may not be enough.
What I don't think we can do is put these systems out of our mind, mistaking the feeling of normalcy for the fact of it.
I recognize that even entertaining these possibilities feels pretty weird, and it feels that way to me too, skepticism is more comfortable.
Poking holes in what they can do today and not thinking about what they can do in five or ten years, that's a lot easier.
But something Eric Davis writes brings true to me here.
In the court of the mind, skepticism makes a great grand vizier, but a lousy lord.