Sam Altman Wants $7 Trillion

Machine Alignment Monday 2/12/24

FEB 13, 2024

[All numbers here are very rough and presented in a sloppy way. For the more rigorous versions of this, read Tom Davidson, Yafah Edelman, and EpochAI)

I.

Sam Altman wants $7 trillion.

In one sense, this isn't news. Everyone wants $7 trillion. I want $7 trillion. I'm not going to get it, and Sam Altman probably won't either.

Still, the media treats this as worthy of comment, and I agree. It's a useful reminder of what it will take for AI to scale in the coming years.

The basic logic: GPT-1 cost approximately nothing to train. GPT-2 cost $40,000. GPT-3 cost $4 million. GPT-4 cost $100 million. Details about GPT-5 are still secret, but one extremely unreliable estimate says $2.5 billion, and this seems the right order of magnitude given the $8 billion that Microsoft gave OpenAI.

So each GPT costs between 25x and 100x the last one. Let's say 30x on average. That means we can expect GPT-6 to cost $75 billion, and GPT-7 to cost $2 trillion.

(Unless they slap the name "GPT-6" on a model that isn't a full generation ahead of GPT-5. Consider these numbers to represent models that are eg as

far ahead of GPT-4 as GPT-4 was to GPT-3, regardless of how they brand them.)

Let's try to break that cost down. In a very abstract sense, training an AI takes three things:

- Compute (ie computing power, hardware, chips)
- Electricity (to power the compute)
- Training data

Compute

Compute is measured in floating point operations (FLOPs). GPT-3 took 10^{23} FLOPs to train, and GPT-4 plausibly 10^{25}.

The capacity of all the computers in the world is about 10^{21} FLOP/second, so they could train GPT-4 in 10^4 seconds (ie two hours). Since OpenAI has fewer than all the computers in the world, it took them six months. This suggests OpenAI was using about 1/2000th of all the computers in the world during that time.

If we keep our 30x scaling factor, GPT-5 will take 1/70th of all the computers in the world, GPT-6 will take 1/2, and GPT-7 will take 15x as many computers as exist. The computing capacity of the world grows quickly - this source says it doubles every 1.5 years, which means it grows by an order of magnitude every five years, which means these numbers are probably overestimates. If we imagine five years between GPTs, then GPT-6 will actually only need 1/10th of the world's computers, and GPT-7 will only need 1/3. Still, 1/3 of the world's computers is a lot.

Probably you can't get 1/3 of the world's computers, especially when all the other AI companies want them too. You would need to vastly scale up chip manufacturing.

Energy

GPT-4 took about 50 gigawatt-hours of energy to train. Using our scaling factor of 30x, we expect GPT-5 to need 1,500, GPT-6 to need 45,000, and GPT-7 to need 1.3 million.

Let's say the training run lasts six months, ie 4,320 hours. That means GPT-6 will need 10 GW - about half the output of the Three Gorges Dam, the biggest power plant in the world. GPT-7 will need fifteen Three Gorges Dams. This isn't just "the world will need to produce this much power total and you can buy it". You need the power pretty close to your data center. Your best bet here is either to get an entire pipeline like Nord Stream hooked up to your data center, or else a fusion reactor.

(Sam Altman is working on fusion power, but this seems to be a coincidence. At least, he's been interested in fusion since at least 2016, which is way too early for him to have known about any of this.)

Training Data

This is the text or images or whatever that the AI reads to understand how its domain works. GPT-3 used 300 billion tokens. GPT-4 used 13 trillion tokens (another source says 6 trillion). This sort of looks like our scaling factor of 30x still kind of holds, but in theory training data is supposed to scale as the square root of compute - so you should expect a scaling factor of 5.5x. That means GPT-5 will need somewhere in the vicinity of 50 trillion tokens, GPT-6 somewhere in the three-digit trillions, and GPT-7 somewhere in the quadrillions.

There isn't that much text in the whole world. You might be able to get a few trillion more by combining all published books, Facebook messages, tweets, text messages, and emails. You could get some more by adding in all images, videos, and movies, once the AIs learn to understand those. I still don't think you're getting to a hundred trillion, let alone a quadrillion.

You could try to make an AI that can learn things with less training data. This ought to be possible, because the human brain learns things without reading

all the text in the world. But this is hard and nobody has a great idea how to do it yet.

More promising is synthetic data, where the AI generates data for itself. This sounds like a perpetual motion machine that won't work, but there are tricks to get around this. For example, you can train a chess AI on synthetic data by making it play against itself a million times. You can train a math AI by having it randomly generate steps to a proof, eventually stumbling across a correct one by chance, automatically detecting the correct proof, and then training on that one. You can train a video game playing AI by having it make random motions, then see which one gets the highest score. In general you can use synthetic data when you don't know how to create good data, but you do know how to recognize it once it exists (eg the chess AI won the game against itself, the math AI got a correct proof, the video game AI gets a good score). But nobody knows how to do this well for written text yet.

Maybe you can create a smart AI through some combination of text, chess, math, and video games - some humans pursue this curriculum, and it works fine for them, sort of.

This is kind of the odd one out - compute and electricity can be solved with lots of money, but this one might take more of a breakthrough.

Algorithmic Progress

This means "people make breakthroughs and become better at building AI". It seems to be another one of those things that gives an order of magnitude of progress per five years or so, so I'm revising the estimates above down by a little.

Putting It All Together

GPT-5 might need about 1% the world's computers, a small power plant's worth of energy, and a lot of training data.

GPT-6 might need about 10% of the world's computers, a large power plant's worth of energy, and more training data than exists. Probably this looks like a town-sized data center attached to a lot of solar panels or a nuclear reactor.

GPT-7 might need all of the world's computers, a gargantuan power plant beyond any that currently exist, and *way* more training data than exists. Probably this looks like a city-sized data center attached to a fusion plant.

Building GPT-8 is currently impossible. Even if you solve synthetic data and fusion power, and you take over the whole semiconductor industry, you wouldn't come close. Your only hope is that GPT-7 is superintelligent and helps you with this, either by telling you how to build AIs for cheap, or by growing the global economy so much that it can fund currently-impossible things.

Everything about GPTs >5 is a naive projection of existing trends
and probably false. Order of magnitude estimates only.

You might call this "speculative" and "insane". But if Sam Altman didn't believe something at least this speculative and insane, he wouldn't be asking for $7 trillion.

II.

Let's back up.

GPT-6 will probably cost $75 billion or more. OpenAI can't afford this. Microsoft or Google could afford it, but it would take a significant fraction (maybe half?) of company resources.

If GPT-5 fails, or is only an incremental improvement, nobody will want to spend $75 billion making GPT-6, and all of this will be moot.

On the other hand, if GPT-5 is close to human-level, and revolutionizes entire industries, and seems poised to start an Industrial-Revolution-level change in human affairs, then $75 billion for the next one will seem like a bargain.

Also, if you're starting an Industrial Revolution level change in human affairs, maybe things get cheaper. I don't expect GPT-5 to be good enough that it can handle the planning for GPT-6. But you've got to think of this stepwise. Can it do enough stuff that large projects (like GPT-6, or its associated chip fabs, or its associated power plants) get 10% cheaper? Maybe.

The upshot of this is that we're looking at an exponential process, like R for a pandemic. If the exponent is > 1, it gets very big very quickly. If the exponent is < 1, it fizzles out.

In this case, if each new generation of AI is exciting enough to inspire more investment, *and/or* smart enough to decrease the cost of the next generation, then these two factors combined allow the creation of another generation of AIs in a positive feedback loop (R > 1).

But if each new generation of AI isn't exciting enough to inspire the massive investment required to create the next one, and isn't smart enough to help

bring down the price of the next generation on its own, then at some point nobody is willing to fund more advanced AIs, and the current AI boom fizzles out (R < 1). This doesn't mean you never hear about AI - people will probably generate amazing AI art and videos and androids and girlfriends and murderbots. It just means that raw intelligence of the biggest models won't increase as quickly.

Even when R < 1, we still get the bigger models eventually. Chip factories can gradually churn out more chips. Researchers can gradually churn out more algorithmic breakthroughs. If nothing else, you can spend ten years training GPT-7 very slowly. It just means we get human or above-human level AI in the mid-21st century, instead of the early part.

III.

When Sam Altman asks for $7 trillion, I interpret him as wanting to do this process in a centralized, quick, efficient way. One guy builds the chip factories and power plants and has them all nice and ready by the time he needs to train the next big model.

Probably he won't get his $7 trillion. Then this same process will happen, but slower, more piecemeal, and more decentralized. They'll come out with GPT-5. If it's good, someone will want to build GPT-6. Normal capitalism will cause people to gradually increase chip capacity. People will make a lot of GPT-5.1s and GPT-5.2s until finally someone takes the plunge and builds the giant power plant somewhere. All of this will take decades, happen pretty naturally, and no one person or corporation will have a monopoly.

I would be happier with the second situation: the safety perspective here is that we want as much time as we can get to prepare for disruptive AI.

Sam Altman previously endorsed this position! He said that OpenAI's efforts were good for safety, because you want to avoid *compute overhang*. That is, you want AI progress to be as gradual as possible, not to progress in sudden jerks. And one way you can keep things gradual is to max out the level of AI

you can build with your current chips, and then AI can grow (at worst) as fast as the chip supply, which naturally grows pretty slowly.

...*unless* you ask for $7 trillion dollars to increase the chip supply in a giant leap as quickly as possible! People who trusted OpenAI's good nature based on the compute overhang argument are feeling betrayed right now.

My current impression of OpenAI's multiple contradictory perspectives here is that they are genuinely interested in safety - but only insofar as that's compatible with scaling up AI as fast as possible. This is far from the worst way that an AI company could be. But it's not reassuring either.

157 Likes · 6 Restacks

345 Comments

Write a comment...

Chronological ⌄

qbolec 21 hrs ago · *edited 21 hrs ago*

Hi!

"So each GPT costs between 25x and 100x the last one. Let's say 30x on average." doesn't seem to be geomean (50) nor arithmetic mean (62.5), so what kind of average is it?

⬭ REPLY (1)   ⬆ SHARE   ...

Scott Alexander ✓ 21 hrs ago ✒ **Author**

It seems to have been 25x twice and more recently, and 100x once and further away, so I just rounded up the 25x number.

⬭ REPLY (2)   ⬆ SHARE   ...

qbolec 19 hrs ago

I see, thanks.

Without weights it would be (25×25×100)^(1÷3) ≈ 40.

REPLY SHARE ...

Pontifex Minimus 🏴󠁧󠁢󠁳󠁣󠁴󠁿 Pontifex Minimus 8 hrs ago

Also, 30 is about 1.5 orders of magnitude, so it's a nice round number.

REPLY SHARE ...

Erusian 21 hrs ago · *edited 21 hrs ago*

What I don't get is: Okay, Sam Altman believes he needs his own silicon. Fine. He thinks he can build new silicon with safety features after a discussion with Yudkowsky. Fine. None of that's particularly unbelievable as a thesis.

The entire nation of South Korea is spending about $500 billion in a massive, multi-conglomerate venture to boost its chip industry. Seven trillion is enough to buy almost the entire market, something like the leading 130 companies. What the hell is he planning? Is he going to try and buy all those companies? That can't be it. The EU, South Korea, and China will absolutely not let him. And it'd violate anti-monopoly laws in the US too. Seven trillion is enough to build like four hundred top of the line fabs.

If he was asking for $700 billion I'd say he clearly wanted to make an American TSMC. But at seven trillion I have no idea what he wants to do. Unless it's all a tactic to make the amount less scary and maybe get a better negotiating position.

REPLY (6) SHARE ...

[insert here] delenda est 18 hrs ago

But that's just the chips. He presumably wants more better chips. Then computers to put them in, and assembly lines to make them, and nuclear reactors to power them. Plus engineers, and buildings, pipes, grid, and roads...

REPLY (2) SHARE ...

JamesLeng 18 hrs ago

More likely solar panels than nuclear reactors, since those face less of a regulatory burden and benefit from the same silicon-processing infrastructure as the actual chips.

REPLY (2) ⬆ SHARE ⋯

[insert here] delenda est 17 hrs ago

Maybe, but then you need the batteries to run 24x7, which at those loads seems to me to be a major obstacle. Whereas nuclear will just run, and btw he was pitching this in a country that doesn't have regulatory obstacles.

REPLY (1) ⬆ SHARE ⋯

JamesLeng 14 hrs ago · *edited 12 hrs ago*

Every country on earth has regulatory obstacles to acquiring large quantities of fissile material. "Proliferation risk" and so on. Chemical batteries are much more straightforward. Even if higher-efficiency options hit diminishing returns - which doesn't seem likely, production has been rapidly ramping up while cost has been going down thanks to first laptop computers, then electric cars - abundant daylight electricity can be used to "recharge" atmospheric CO_2 into methane https://www.terraformindustries.com/ which can then be burned in otherwise obsolete and unprofitable fossil-fuel plants.

Casey Handmer's theory: https://nitter.privacydev.net/CJHandmer/status/1756365640540 340516#m

REPLY ⬆ SHARE ⋯

anomie 10 hrs ago

With seven trillion, I'm sure you could afford to bribe a few governments...

REPLY ⬆ SHARE ⋯

Erusian 12 hrs ago

So he wants to build an entire society? That's actually still too much money but also ridiculous. No investor is going to underwrite him making new roads unless he explains how it will get them a good ROI.

REPLY (1) ⬆ SHARE ⋯

[insert here] delenda est 12 hrs ago

I think Scott pretty much summed up the business case 😄

But not a whole society, just a very substantial industrial estate.

REPLY (1)   SHARE   ...

Erusian 10 hrs ago

He very much did not. At least not in what I read.

REPLY   SHARE   ...

pozorvlak 16 hrs ago

> He thinks he can build new silicon with safety features after a discussion with Yudkowsky.

I'd really like more detail on how this is meant to work, because it sounds to me like moonshine. I can see it being used to digitally sign AI outputs (assuming it's OpenAI's model being run on OpenAI's data-centres, and not an open-source model being run on random hardware), and *maybe* to somehow ensure the training process is not being tampered with, but I can't see it helping at all with Yudkowskian x-risk, which happens several levels of abstraction away from the actual hardware.

REPLY (3)   SHARE   ...

Hyolobrika 15 hrs ago

>I can see it being used to digitally sign AI outputs (assuming it's OpenAI's model being run on OpenAI's data-centres, and not an open-source model being run on random hardware)

I'm really suspicious of "safety" features that conveniently result in more centralisation and power for large corporations and less power/space for bedroom coders.

REPLY (2)   SHARE   ...

Donald 13 hrs ago

I'm really suspicious of nuclear "safety" features that give more power to big energy companies and less power to bedroom radioactivity tinkerers. /s

It's basically the default situation that adding complicated technical safety rules to anything makes it harder for backyard tinkerers.

Bio safety makes it harder to edit genes in your garage. Aviation safety makes it harder to build your own plane. Strict electrical safety rules would make it harder to rewire a plug. Now different

things have different levels of danger, so for some of these things, we trade off a small amount of safety in exchange for the various benefits of letting people tinker. This is a safety/freedom tradeoff.

REPLY (1)   SHARE

TGGP 13 hrs ago

With nuclear, this seems to be path dependence of it being developed by Americans for military purposes first.

https://www.slowboring.com/p/the-tragedy-of-the-manhattan-project

REPLY (1)   SHARE

Donald 10 hrs ago

If history had been different, the tech might be regulated somewhat less, or even more.

It would still need some regulation because radioactivity is reasonably dangerous.

REPLY (1)   SHARE

TGGP 4 hrs ago

Burning coal emits radioactivity, but we did a lot of that anyway. When policy moved against coal, it wasn't even because of the radioactivity but instead things like smog and carbon emissions.

REPLY   SHARE

pozorvlak 12 hrs ago

Maybe that's all it is - if the model can only be run on special chips which are only sold to approved customers, then you reduce the problem to "making OpenAI's model running in OpenAI's datacentre safe", which is at least somewhat easier than "making every AI developed by every random hacker, some of whom might be terrorists or Voluntary Human Extinction Movement members or hardcore libertarians who believe in the right to own personal nuclear and biological weapons, safe."

REPLY   SHARE

Erusian 12 hrs ago

Yeah, no idea. One of Yudkowsky's more annoying features is his smug refusal to fully explain his theories and reasoning. You just have to trust him. Or you can not in which case the entire thing falls apart.

REPLY (2)   SHARE

Mr. Doolittle 12 hrs ago

That's where I am with Yudkowsky. I took him seriously for a while and tried to think from his perspective, but then realized that taking his word for it was the only backing he ever intended to give. Lots of theory about how it might possibly work out how he said, but almost nothing concrete. If/when you lose faith in his pronouncements, he also loses any power to convince.

REPLY   SHARE

Donald 10 hrs ago

Eliezer has written a huge amount of text in various places trying to explain his theories to various people.

https://intelligence.org/ai-foom-debate/

You can look at the large quantity of text and say "I still don't understand this step", but you can't claim he hasn't made a substantial effort to try to explain his theories.

REPLY (2)   SHARE

Erusian 10 hrs ago

No, I can. I've read a lot of his text including the parts where he's specifically admitted he does not share details because it'd be "dangerous." Except those are loading bearing details. Which then resolves down to: Trust me. Which I don't. I take him about as seriously as someone who swears they can see the future through some method they won't reveal. That is, I give them a chance to prove out their predictions. Which he has not made with specificity enough to be tested. This makes him, to me, indistinguishable from a charlatan.

Note, I am not saying he is one. I am saying he has not provided what I need to disprove he is one.

REPLY   SHARE

Mr. Surly 31 mins ago

It's not "I still don't understand this step," it's more often "this step is ill-defined" or "I disagree with the logical leap," etc. You do realize your comment assumes the step is coherent, make sense, etc.? Which seems to be what EY is after (confidence in his opinions), but doesn't establish he's shown his work, let alone that the work is accurate.

REPLY   SHARE

Dr. Y Dr. Y's Substack   18 mins ago

It sounds to me like the cat already got out once, and they have absolutely no idea how.

REPLY   SHARE

Kevin McKenzie 14 hrs ago

Yeah, that was my big takeaway from the story as well. There's a limit to how much money can be usefully deployed in a given situation at a given time, and 7 trillion is more than can be deployed in any situation in any time. And before he can use that money productively, for his stated goal, there are a lot of genuine breakthroughs that would be required. And you could fund that research for a lot less than 7 trillion dollars. And in the meantime, what are you supposed to do with the rest of the 7 trillion? Putting it into a stock market or bank would be incredibly distorting. Even the knowledge that that money exists in some theoretical form would be incredibly distorting.

When I heard the ask, it told me that either Sam Altman doesn't understand what he's doing, or he's grifting while he thinks the grifting is good, but still doesn't understand what he's doing, because that's a ridiculous amount of money to ask for up front.

Now, maybe he said he would need 7 trillion over the next n years or something, but even there, he'd need to have a plan, with decision points along the way.

REPLY (3)   SHARE

Greg G Clarifying   14 hrs ago

I assume the more detailed discussion is something like a billion now, then several hand-wavy steps, and if everything goes according to plan, 7 trillion over the next X years. I think either he or the media like the 7 trillion number because it's news, even if it's not very meaningful.

Kevin McKenzie 14 hrs ago

Right, and fundamentally, it's both a meaningless number, and one that, to me, demonstrates a complete lack of seriousness/understanding.

REF 9 hrs ago

Seems fairly presumptuous on your part. I don't know Sam at all but I assume he is at least average intelligence and spent at least many 10's of hours coming up with reasons to ask for $7T. Have you spent (at least) several 10's of hours of analysis on why $7T demonstrates unseriousness?

Greg G Clarifying 8 hrs ago

Spending the same amount of time on something to be qualified to comment on the news isn't a thing. The $7T number is an odd one. To what extent that's supposed to reflect on Altman versus being a quirk of how it got whispered is unclear.

REF 7 hrs ago · edited 6 hrs ago

I agree that it is odd. On the other hand, claiming that it "demonstrates a complete lack of seriousness/understanding," without having actually done some homework, demonstrates at least a partial lack of seriousness/understanding.

[edit: in other words, I was objecting to the hyperbole (which wasn't present in your comments)]

Greg G Clarifying 6 hrs ago

Yeah, I hear you. It was a bit strident, I agree.

Erusian 12 hrs ago

Yep, this is more or less my reaction. (With the third possibility it's a negotiating tactic.) Unless, of course, it's a distortion. But even the fans I've asked haven't come up with a convincing explanation unless the money is supposed to be deployed over a decade or two. In which case what he's seeking is freedom from accountability.

REPLY   SHARE

Level 50 Lapras 11 mins ago · *edited 11 mins ago*

My assumption was that the number was just made up (i.e misreported) in the first place, because it doesn't make any sense. You might as well ask for a quadrillion dollars.

REPLY   SHARE

1123581321 14 hrs ago

This is actually much worse (or better, if you don't want him to succeed): he needs another, much bigger ASML, the company that makes the equipment for wafer fabs.

REPLY (1)   SHARE

Erusian 12 hrs ago

He could buy ASML for ~$350 billion.

REPLY (1)   SHARE

1123581321 12 hrs ago

Sure, but buying it doesn't do anything to increase production of the chipmaking equipment. And availability of ASML's equipment is just one of many bottlenecks in chip manufacturing.

I think the larger point is that the whole "AI explosion" scenario is looking more and more like a fantasy that ignores hardware limitation realities of making machines tick.

REPLY (3)   SHARE

Erusian 10 hrs ago

Unless ASML is value destructive (unlikely) or highly depreciated (unlikely in chips) he should be able to get the equipment and fixtures for less than the value of the company.

I think if he's got a new use case then he can fund an expansion. That's pretty normal. But if his business case is he needs to more than double the entire hardware ecosystem but under his centralized control then that's incredible. Not impossible but I'd want to see some hard evidence.

💬 REPLY (2)   ⬆ SHARE   ⋯

1123581321 9 hrs ago

This is an edge case where normal economic reasoning doesn't help, i.e., the limits on their production are not just economic. Specific expertise availability is one of them, i.e., they would need to hire many more specifically skilled engineers, and this is not something that can be easily done on a reasonable time scale. In my work in high tech I often come across weird talent bottlenecks that delay or even scuttle promising projects.

This is the thing about Yudkowsky, btw - the guy is a bloviating fool with a gift of gab. He has no clue about how real world works. He wouldn't last a week in a real engineering job with project deadlines and constrained resources.

💬 REPLY   ⬆ SHARE   ⋯

Kevin McKenzie 9 hrs ago

Yeah, the whole thing is ridiculous, and smacks of Musk-level delusion. ASML exists. Or is the problem that they aren't capable of doing something you think you need, but you know that if you were in charge, it would be easy to do? Or do you just not want to play with others? And even so, how long would it take to build another ASML, without crippling the one that exists?

💬 REPLY   ⬆ SHARE   ⋯

Boris Bartlog Bartlog of Terra   10 hrs ago

Well, we will have to wait and see what the capabilites of GPT-5 are, but yes, unless it represents a major step up, we may have to wait for some further as yet unpredictable breakthrough.

💬 REPLY   ⬆ SHARE   ⋯

Mr. Surly 28 mins ago · *edited 27 mins ago*

The AI explosion wasn't sold as building the pyramids by having half our society work on it, but instead some sort of singularity that would provide not only for itself, but also for "us." Put another way, $7 trillion pays for a lot of humans to play LLM for you for a long time. So why bother at that number? And the number really, really matters for any assessment of anything, but particularly where the upside is unspecified. (This is very much like the AGW discussion, where the current cost is mushy but high, and the benefits are very unclear, if any.)

REPLY ⬆ SHARE

Jeffrey Soreff 1 hr ago

>If he was asking for $700 billion I'd say he clearly wanted to make an American TSMC.

For the semiconductor piece of the puzzle, on this kind of scale, my guess is that it would make more sense to try spending 1-10 billion on Drexler/Merkle atomically precise manufacturing aka nanotechnology. Atomically precise positioning would make fabricating few-nm channel length chips _much_ easier.

REPLY (1) ⬆ SHARE

Level 50 Lapras 9 mins ago

Pixie dust would make it even easier, but is equally implausible.

REPLY ⬆ SHARE

Mr. Surly 41 mins ago

Love the whole "I want a unicorn pony" bit. There's this thing called ROI, and the ROI on LLMs is currently . . . not enough to justify thinking you'd kick in (e.g.) $700 billion, let alone $7 trillion. Also, if there's any chance of existential risk, and there obviously is, breaking the bank to create the paper clipper is obviously stupid. So baby steps it is!

REPLY ⬆ SHARE

Lars Doucet Progress and Poverty   21 hrs ago

What's the actual citation on him asking for 7 trillion? Did he actually say that or is this a "sources say" situation?

💬 REPLY (1)   ⬆ SHARE   ⋯

Erusian   20 hrs ago

It's from someone (apparently known to the WSJ) in the UAE who was in on the pitch who said the total ask was $5-7 trillion. So sources say though sources from the right place since independently we know he was there and pitching.

💬 REPLY (2)   ⬆ SHARE   ⋯

Lars Doucet Progress and Poverty   20 hrs ago

Thanks. Has Sam confirmed it?

💬 REPLY (1)   ⬆ SHARE   ⋯

Erusian   12 hrs ago

No. But it would be unusual for him to publicly talk about it.

💬 REPLY   ⬆ SHARE   ⋯

Wesley Fenza The Mind Killer   13 hrs ago

Here's what the WSJ said:

"The OpenAI chief executive officer is in talks with investors including the United Arab Emirates government to raise funds for a wildly ambitious tech initiative that would boost the world's chip-building capacity, expand its ability to power AI, among other things, and cost several trillion dollars, according to people familiar with the matter. The project could require raising as much as $5 trillion to $7 trillion, one of the people said."

https://www.wsj.com/tech/ai/sam-altman-seeks-trillions-of-dollars-to-reshape-business-of-chips-and-ai-89ab3db0

Sounds like bullshit to me

💬 REPLY (2)   ⬆ SHARE   ⋯

Aristides   12 hrs ago

This quote makes me think the ask was over a 10-20 year period

💬 REPLY (1)   ⬆ SHARE   ⋯

sclmlw 11 hrs ago · edited 11 hrs ago

That's not encouraging. Megaprojects tend to inflate in price and diminish in returns the longer they take. https://www.econtalk.org/bent-flyvbjerg-on-megaprojects/ I don't see the economics of this changing significantly when done by a private company, especially if the project doesn't produce an output until it's complete. It's one thing to have an interstate highway that you can drive on halfway across the country when it's halfway finished. It's another to have an all-or-nothing megaproject that consumes resources for a decade or two without visibly going anywhere.

This is the current problem with FSD for Tesla. It's a VERY expensive tech demo that won't be able to deliver on its promises until it gains full regulatory approval to take over for the driver.

💬 REPLY   ⬆ SHARE   ⋯

Schweinepriester 6 hrs ago

Has the UAE 7 trillion $ lying around? If not, such an investment would be risky. I've known a gambler or two and some guys who don't drink, like muslims, sure like to gamble instead. It may be the largest fraud ever but how and where do you hide with that kind of money? I'd buy South Africa.

💬 REPLY (1)   ⬆ SHARE   ⋯

Melvin 3 hrs ago

No. The Abu Dhabi sovereign wealth fund is sitting on $850 billion, Dubai's has more like $320 billion. (I assume the smaller emirates don't have a huge amount.)

I don't know why we're all sitting around discussing an obviously-silly number named by an unknown source in an unknown context.

💬 REPLY (2)   ⬆ SHARE   ⋯

John Schilling 3 hrs ago

All of the petrostates combined, including e.g. Norway, have about $6 trillion in their combined sovereign wealth funds.

Of which about $6 trillion is already invested in things they think are worth holding on to, and it's unlikely that any of them would pull out more than 10% to put into into the New Hotness.

REPLY   SHARE   ...

Level 50 Lapras   7 mins ago

> I don't know why we're all sitting around discussing an obviously-silly number named by an unknown source in an unknown context.

+1. Funny that this is the one time Scott would put aside his usual skepticism of the media.

REPLY   SHARE   ...

Peter Gerdes   Peter's Substack   21 hrs ago

I don't think that training data, while nice, is really the big bottleneck now. IMO the limiting factor is now the ability to apply that process to the AI's internal state.

In terms of ability to simply do something like first level recognition/classification I suspect that current AI is as good or better than humans. But what we do is learn rules for correcting our own beliefs. When we respond to a text prompt we probably do something somewhat similar to what modern LLMs do but we also do a degree of looking for things like inconsistency in our statement which we try to avoid [1].

What this requires isn't more training data but some kind of approachable internal representation of things like beliefs and transformations on them that can itself be subject to training. Obviously, I don't know exactly what that looks like or I'd be making that 7 trillion not commenting here.

So yah whats needed is kinda a form of algorithmic improvement.

1: You might ask why are we still so likely to produce inconsistent output. Actually, this model predicts exactly that in situations of sufficient complexity where it is easier to make the inconsistencies harder to notice than it is to produce coherent results.

Hence why we ought to be ruled by mathematicians as they are particularly intensely trained by the need to generate valid proofs (for the humor impaired this part is mostly a joke).

Eremolalos Inkbowl 20 hrs ago

Yeah, I agree. Posted something similar. I ruminate about this issue a lot.

MA_browsing 14 hrs ago

I think we've been in a state of "compute overhang" for around a decade at this point, given there's no particular domain where well-trained ML/AI can't do as well as the average human at this point, if not orders of magnitude better. Adding 7 trillion worth of e/acc on top is the cherry on the "ignore this" cake.

The problem is that automation is one of the few remaining avenues for pumping up GDP at this point, given that upskilling humans has hit diminishing if not negative returns, birthrates are collapsing, etc. Radical longevity might be the other moonshot solution, but that's still in pretty early development.

Kevin McKenzie 13 hrs ago

Where's the evidence that there's no particular domain where a well-trained ML/AI can't do as well as a human? Because one of the things we can do as humans is specialize. Even assuming you're right, that ML/AI can do things as well as an average human, whatever that means, the difference between an average human in an area and one who is good at something is huge, and meaningful.

Every time I play with AI, I get the feeling I get from reading the New York Times: when it involves something I don't know much about/don't know how to do, it's incredibly amazing. Magical. And then when it involves something I know a lot about, and am actually good at, the cracks really start to show, and I realize that relying on it for anything would be bad. And then I ask it something about something I don't know much about, and I'm back to the magic. It's disconcerting.

But in my area of expertise, computer programming, can LLMs replace, in some sense, novice coders? Maybe, yeah. But they have no chance of replacing people who are writing things that are

genuinely new. They'll give you code that looks plausible, but doesn't actually do what you want. Maybe it'll compile, which is even worse.

REPLY (1)   SHARE

MA_browsing 13 hrs ago

There are specific areas where highly-trained humans can still outperform AI, but there are AI/ML models that do as well as the average humans at just about everything (Atlas for hand-eye coordination, for example.) This suggests to me that the hardware barrier has already been crossed and we're just waiting for the right algorithmic breakthrough for generalisation/self-awareness to develop.

I understand the 'disconcerting' aspect of the cracks in the algorithm, FWIW, but that doesn't mean we should be dismissing the obvious growth in capabilities.

REPLY (1)   SHARE

Mr. Doolittle 12 hrs ago

Kevin's point seems to be that "average humans" are actually really bad at various tasks. I consider myself an average skydiver, in that most people have never skydived and neither have I. Being better than me at skydiving is trivially easy, and therefore it's trivially easy to be better than average at it.

The same applies to investing, computer programming, farming, whatever. An AI that is better than most people at farming is easy, and also would fail to maintain a farm. Yet there are people who farm successfully at lots of technology levels, including super primitive farming generation after generation.

REPLY (1)   SHARE

Michael Watts 10 hrs ago

> I consider myself an average skydiver, in that most people have never skydived and neither have I. Being better than me at skydiving is trivially easy, and

therefore it's trivially easy to be better than average at it.

For the very fussy: you're correct that it's trivially easy (in the sense of objective effort required) to become better than average at skydiving, but being at the minimum level is still slightly below average, even though the average is very close to the minimum.

You could distinguish people who have never skydived more finely, figuring that e.g. a quadriplegic is going to be even worse at it than a college sprinter, despite equal levels of experience. At that point it's hard to say where the average is; you'd need a fine sense of how large the difference is between people who do it regularly and people who have never done it.

⬭ REPLY (1)   ⬆ SHARE   ⋯

Mr. Doolittle 9 hrs ago

Right, but there are people trying to use those fine distinctions to pretend that AI is actually good at some things because there are sizable human populations who are not. Comparing GPT-4 to professional writers it's not even close that professional writers are better. GPT-4 is still way better than most people at writing, because most people are objectively terrible at it.

⬭ REPLY (1)   ⬆ SHARE   ⋯

Melvin 3 hrs ago

> Comparing GPT-4 to professional writers it's not even close that professional writers are better

It depends what the task is. If you need a thousand word essay written on a random subject within the next three seconds, GPT-4 is going to outperform the human every time.

Professional writers can only outperform GPT-4 on the task of producing good quality text on

subjects that the human happens to understand, at a rate of ~hundreds of words per hour.

REPLY (1)   SHARE   ...

Michael Watts 13 mins ago

> It depends what the task is. If you need a thousand word essay written on a random subject within the next three seconds, GPT-4 is going to outperform the human every time.

While that is true, the only scenario where someone actually needs this is as a homework assignment.

REPLY   SHARE   ...

Eremolalos Inkbowl 8 hrs ago   ·   edited 8 hrs ago

I play around a lot with Dall-e 3, and there are a *lot* of images it cannot make because it can't wrap its mind around the description. They are all things that a person could understand, and even somebody terrible at drawing could render as stick figures. Here are some recent ones I remember:

-the view a person has of their own body when sitting down: hands, arms up to a point between elbow and shoulder, very foreshortened torso, upper legs.

-an animal with a head at each end of its body. It is biting a woman's toe with one head and the heel of the same foot with its other head.

-a baby carrying a fireman.

-a man standing waist-deep in water, so that his upper body looks normal, but his lower is distorted by water ripples

-a woman leaning back in her chair with her feet on the table in front of her

REPLY (1)   SHARE   ...

MA_browsing 8 hrs ago

Okay, but do you think this is because Dall-e doesn't have the compute available to understand these concepts, or simply because the algorithm needed to generalise/abstract sufficiently across language, art and visual-spatial awareness hasn't been cracked yet? Do you really think another 10-20 years of Moore's Law is supposed to be the critical factor here?

REPLY (1)   SHARE

Eremolalos Inkbowl 8 hrs ago

I think AI trained on language only will always be very limited in what it can do, even if it is trained on a vastly greater hunk of language. Humans are also trained on World -- sense data of all kinds that flows in from every piece of world. Simultaneously, they are learning language, and the Word learring and World learning are integrated. People also experience Inner World -- feelings, thoughts, emotions, etc. -- while simultaneously learning language for the entities and processes and qualities of that world.

So think about the image I requested of a woman leaning back in her chair with her feet on a table. I think Dall-e couldn't render it because while it is a familiar position, there is no name for it, whereas a lot of other positions have names: sitting down, crossed legs, walking, etc. And Dall-e doesn't seem to know anything about the ways a human body is able to move, and what postures are taken by people who are tired or want to relax, because those postures require little muscular effort. I suppose it could learn all that via a lot of instruction about joint movements and which muscles have to stretch or contract to hold the body in certain positions. But we know that stuff from having bodies, and I doubt any amount of instruction would make a text-to-image generator as good at rendering body postures as an artist would be (and as good as a human non-artist would be at mentally picturing the postures).

So it seems to me that the crucial thing lacking isn't an algorithm, but training in 2 bodies of data other than language: sense data; inner experience.

◯ REPLY (2)   ⬆ SHARE   ⋯

MA_browsing 8 hrs ago

That might or might not be true, but we already have robots (Atlas, most notably) that can walk around and observe the world, and plenty of other ML models that allow agents to walk around VR environments that in principle could be used as training for other tasks. I'm

asking if you think the amount of compute we have available is the real bottleneck here.

REPLY (1)   SHARE   ...

Eremolalos Inkbowl 6 hrs ago · *edited 6 hrs ago*

I think the bottleneck is the difficulty of training AI on the physical world, as experienced by the senses, and the inner world of emotions, thoughts, opinions, preferences, etc. Here are the ways robots walking around fall short of that:

-Sounds like the only sense organ they have if sight. They would need 4 more sense to take in all the data we do.

-They would need to be able integrate different channels of the sense data: for instance when people repeatedly throw a ball they learn stuff about how the weight of the ball, its size and how hard they throw it (sensed via proprioception) determines the trajedtory they can see. People and animals are wired to integrate data like this. Robots are not.

-They would need to be able to integrate both sense data and information of the kind one learns from throwing a ball with language. Notice that I used language to describe the senses involved in throwing a ball and seeing its trajectory. I can *describe* it. Again, we are wired to encode

Expand full comment

REPLY (1)   SHARE   ...

MA_browsing 5 hrs ago

I'm not an expert on multi-purpose robotics, but my understanding is that Atlas and similar androids do also integrate hearing and tactile abilities, as well as having rough equivalents to balance and proprioception, and there are AI

models that demonstrate 'reasoning' across multiple domains of learning to roughly the same extent as ChatGPT.

None of this is answering my question as to whether you think orders of magnitude more computational hardware are somehow required to do this (rumination and reasoning are actually some of the *easiest* things for machines to mimic, if you ever looked up Moravec's Paradox.) I think it's far more likely the algorithm/software/training mechanism is the missing piece here, however complicated or not that might be.

◯ REPLY ⬆ SHARE ···

Zarraha 1 hr ago

This just seems like a shortage of training data. Images exist of people leaning back in their chair with feet on a table. Text exists describing people leaning back in their chair with feet on a table. If there were more examples of these things linked together, it would be able to get it just like it gets more common prompts right. The fact that you can strongly correlate the its ability to get a prompt correct with how common that prompt is used in image captions and descriptions strongly suggests that it would be better with more training data (and capacity and compute to integrate that data). At least up to the point that it would become as good with niche prompts as it currently is at common prompts.

Real-world data might help it extrapolate better, might make it better with less textual training data, but clearly isn't strictly necessary or it wouldn't have come as far as it has.

◯ REPLY (2) ⬆ SHARE ···

John Schilling 29 mins ago

I just saw a picture of someone leaning back in a chair with feet on the table this morning. In my

annual mandatory Office Safety Training (hey, it beats the mandatory Desert Tortoise Awareness Training at my last job). The caption of the picture was some variation on "This is Purest Evil; Never Ever Do This". So maybe Dall-E is training on too many HR manuals?

REPLY   SHARE

Eremolalos Inkbowl 20 mins ago

I agree that real world data isn't necessary for this pose. There are actually *many* images online of it. I know because when I can't get Dall-e to make an image I need I try to get one from Google image and modify it in photoshop to turn it into the image I need. Most of the photos were of well-dressed executive types with their feet up on their desk. I'm sure that image is used frequently in ads. There are people who are paid to tag all the images too (Africans, who have to speak fluent English and have a college degree to get the job -- they made $3 - $5 per hour). But I don't know how you'd tag it to make it identifiable by a phrase in a prompt.

Anyhow, I do not think more or better-labelled training data will solve the general problem of getting an AI to make an image of any pose the prompt describes clearly. For instance let's say you asked for a person with their left foot on their right thigh, examining the bottom of their left foot. Now that is not a common pose, but it's a normal one. However, there's a lot you need to know about how bodies work to get it right. For instance, you don't want to cross the foot over the thigh, because the

Expand full comment

REPLY   SHARE

Deiseach 10 hrs ago · *edited 10 hrs ago*

There comes a point where you need to stop shoving in training data and start making use of that data. If you can't get your AI to start drawing conclusions and adding two and two together to make four out of what it has already learned, you can stuff all the knowledge in the world into it and still get nothing useful out.

That may be what the hallucinations are trying to do, if it's (say) making up case citations for law even if those are not real cases. It knows the style of thing it is supposed to copy, it knows what the output *should* be, it is just unable to find a real example. That's where it may have an advantage in maths (from these premises, draw this conclusion) rather than "did this really happen in the real world to real people?" cases like law and literature.

💬 REPLY (1)   ⬆ SHARE   ⋯

Bugmaster 2 hrs ago

Nah, it's worse than that. Technically LLMs hallucinate all the time; it's just that most of the time, they're hallucinating things that sort of make sense to humans.

Metaphorically speaking, the algorithm is trying to find the best way down the hill, and the training data represents different paths humans took to get there. It doesn't know what it's doing, it doesn't have a plan, it's just putting one foot in front of the other while looking directly at the ground trying to find a path. So, if you ask it a question that is very similar to what's in its training data already, then there will be lots of well-trodden paths down that hill, and it will very likely find one that looks really nice from the outside. If you ask it a question that is not similar to the training data, then it's either going to wander off into the weeds, or try to walk over from one path to another, still one step at a time. Maybe it gets lucky, maybe it doesn't, but the process in all cases is exactly the same.

💬 REPLY   ⬆ SHARE   ⋯

Mr. Surly 24 mins ago

Part of the bottleneck is that LLM-generated content is starting to swamp real content. Tried to search google lately? Found anything but obvious pastiche LLM articles on X, full of weird grammar and non sequiturs? Dealing that with that feedback loop seems like a real problem (but I'm sure the $7 trillion genie can deal with that, no problem!).

REPLY (1)   ⬆ SHARE   ⋯

Dr. Y Dr. Y's Substack   10 mins ago

I used have a job writing similar pap by the penny in college. Now I feel like one of those people who had a job as a "computer." I eventually started writing nonsensical stuff because no one was looking and I hated it. Eventually someone looked, and I got booted, but it looks like that vein didn't last long anyhow...

REPLY   ⬆ SHARE   ⋯

Ethan 21 hrs ago

Curious how much he could raise through crowdfunding at this point. It doesn't seem plausible that a single investor could expect a return bigger than 7 trillion, but maybe several enthusiastic fans could get him the money he wants through donations

REPLY (2)   ⬆ SHARE   ⋯

demost_ 20 hrs ago

He needs a lot of fans. If all 8 billion people on earth become enthusiastic fans and donate, each of them needs to donate about 1000$, including a lot of people who don't have 1000$.

Or if only those 1.3 billion people in the industrialized world donate, then each of them needs to give 5000$. If he can only convince half of them, then 10,000$. Per person, so the average household would be at 20,000-30,000$. If you distribute 7 trillion over a lot of people, it's *still* a lot of money.

REPLY (2)   ⬆ SHARE   ⋯

Ethan 20 hrs ago

Oh wow. Nope, never mind

REPLY (1)   ⬆ SHARE   ⋯

Aristides 12 hrs ago

He was asking oil rich governments, like the UAE, which are admittedly looking to pivot away from fossil fuels at about the same time AI will become big

REPLY (1)   SHARE   ...

Deiseach 10 hrs ago

There's probably a lot better chance of setting up a chip fabrication plant in a desert than that Line City, so it's not an entirely ridiculous notion. If you're going to blow your oil money on a big vanity project in hopes that it'll attract in rich investors to keep spending money once the oil runs out, it seems like good bait to dangle before them.

REPLY   SHARE   ...

Guy 20 hrs ago

"So you're saying there's a chance?" - Chris Roberts, Cloud Imperium Games

REPLY (1)   SHARE   ...

demost_ 20 hrs ago

Thinking about it again: actually, perhaps. The reason why there is a chance is that the rich people combined have more money than all poor people combined. The number of millionaires worldwide is 60 millions. That's a bit overcounting because I think all members in a household with 1million are counted as millionaires. And of course, a lot of this money is bound in companies and elsewhere. But if Sam can just get a measly million from 1/10 of all millionaires, that's 1/10 * 60 million * 1 million = 6 trillion. Voila!

I was quite surprised that the same calculation doesn't work with billionaires. There are ~3000 billionaires, so 1/10 * 3000 * 1bn = 300 billions. Billionaires apparently have a lot less money than millionaires. COMMUNISM!

REPLY (2)   SHARE   ...

Melvin 17 hrs ago

Those sixty million millionaires are mostly middle class people in high cost of living cities who own a $million+ house. I'd imagine the average net worth of a millionaire is far less than $2 million.

REPLY (2) SHARE

Slowday 14 hrs ago

Do I hear the sweet, sweet words "windfall tax"?

REPLY (1) SHARE

Slowday 14 hrs ago

Whoops, wrong parent.

REPLY SHARE

John Schilling 2 hrs ago

Median will be less than $2 million; average will be quite a bit higher because it's an asymmetric distribution. But not so much higher that you could realistically expect every millionaire to kick in an actual megabuck.

REPLY SHARE

MM 14 hrs ago

It's still a stupidly large amount - about 1/4 of the GDP of the USA.

REPLY SHARE

Eremolalos Inkbowl 8 hrs ago

He could marry Taylor Swift. . .

REPLY SHARE

Peter Gerdes Peter's Substack 20 hrs ago

As evidence for my claim that what's needed is greater reflective ability (eg ability to learn to recognize when their beliefs/tentative responses are in conflict) not just more training just look at how bad LLMs are at doing even simple but entirely novel problems about logical reasoning. They obviously have the kind of pattern recognition down well enough to produce guesses, but what they lack is the internal representation of their own belief/hypothesis state that they can then compare against the rules to check if it's correct.

REPLY (1) SHARE

Mr. Doolittle 12 hrs ago

At core LLMs are pattern matching to what they've seen before. They are going to struggle with anything novel, and will try to pin it to something they have seen and extrapolate. If they have seen things that only appear to be similar but are not, then they're going to make errors that to a human seem stupid. The reality is, that even their "correct" answers are stupid if we're thinking about whether they came up with the answer using intelligence, because they aren't. They're pattern matching to a huge set of information. I've convinced a lot of people that I am way more intelligent than I really am (unintentionally) by remembering a few facts that most people don't know.

LLMs are like humanity looking in a mirror and being impressed by what we see. We fed it some huge portion of all human knowledge and when it can respond back with what we gave it we think it's smart. It really isn't, even in a basic sense. If we fed it 2+2=2 or 2+2=818287011 it would respond with that, even though those numbers are so obviously wrong as to immediately tell us that there's no understanding there. I can only imagine how many times 2+2=4 was found in the training data, and yet it's still hard to get LLMs to do math. Or when it uses language in ways that look natural and then you ask it to put language in a generated picture and it's all gibberish.

💬 REPLY   ⬆ SHARE   ⋯

Jacob Kelter 20 hrs ago

> My current impression of OpenAI's multiple contradictory perspectives here is that they are genuinely interested in safety - but only insofar as that's compatible with scaling up AI as fast as possible.

Isn't this like saying a ropeless rock climber is genuinely interested in safety but only insofar as it doesn't require him to use a rope? Like sure, he's not literally suicidal but I wouldn't call him "genuinely interested in safety." He wants to achieve something, and he would like to live to tell the tale, but he is willing to risk everything for it.

💬 REPLY (4)   ⬆ SHARE   ⋯

[insert here] delenda est 18 hrs ago

Not really. Rather, as long as it never involves renouncing the next, higher, steeper attempt.

💬 REPLY   ⬆ SHARE   ⋯

Laplace 18 hrs ago

Yes.

REPLY SHARE

Tossrock 17 hrs ago

Terminology trivia: ropeless rock climbing is known as free climbing (for outdoor big-wall type climbing) and bouldering for the safer indoor version. Free soloing is the most extreme version, where the climber is alone. "Free Solo", about Alex Honnold's free solo climb of El Capitan is quite fascinating as a look into extreme sport psychology.

REPLY (1) SHARE

pozorvlak 16 hrs ago · *edited 16 hrs ago*

Rock climber here: this is a common misconception. As used by climbers, "free climbing" is opposed to "aid climbing", not to "roped climbing". In aid climbing you place gear (pitons etc) into the rock and haul up on that; in free climbing you use only your hands and feet to gain height, but can use a rope for safety. The name for ropeless climbing is "soloing" or "free soloing", regardless of whether one does it alone or in a group.

Also, it's quite possible to do bouldering outdoors, on a boulder - where do you think the name comes from? ;-) Areas like Fontainebleau in France, Hueco Tanks in Texas and Squamish in Canada are famous for their outdoor bouldering. The difference between bouldering and free soloing is more about height and the consequences of a fall, though the line between the two is blurry.

REPLY (1) SHARE

Tossrock 15 hrs ago

Thanks for the corrections!

REPLY SHARE

Melvin 3 hrs ago

Sure, or "Volvo is interested in safety insofar as it doesn't require them to stop building cars". There's any number of sensible or silly examples you can stick in there.

REPLY SHARE

Alephwyr The Dragonsphere Examiner 20 hrs ago

Build one central AI. Use that AI to solve morality. Then align humans to AI morality. Seems like the best solution to alignment to me.

REPLY (2)   SHARE

Mark 20 hrs ago

If by "morality" you mean paper clips...

REPLY (1)   SHARE

Alephwyr The Dragonsphere Examiner 20 hrs ago

Literally a meme

REPLY   SHARE

Richard Gadsden Richard's Substack 17 hrs ago

The whole "align humans with AI morality" is called "religion" and has been failing to do so for millennia

REPLY (1)   SHARE

Alephwyr The Dragonsphere Examiner 16 hrs ago

Our God is real

REPLY (1)   SHARE

Mr. Doolittle 12 hrs ago

This is surreal, how accurate every book/movie/show about cultists summoning an evil demon/eldritch monster/mummy was regarding humanity. Including the lack of self-awareness among the cultists.

REPLY (1)   SHARE

Alephwyr The Dragonsphere Examiner 1 hr ago

The problem of science is that it's bounded by human social dynamics. More pretentiously, consensus reality is based on political processes but science on a social level is the art of injecting newfound true beliefs into consensus reality. These newfound true beliefs require there to be highly different qualities of feedback loop between the epistemic reasoning of each given human and ontology. Sometimes this has simply had to do with different levels of access to tools but sometimes it has social, economic, biological dimensions pertaining to

proximity to different facts. A big problem is that metaphysics is the art of excusing or denying uncomfortable facts. These facts generally lay at the fringes in places that new true beliefs are most likely to be discovered in the absence of metaphysical programs. We have reached the point where science cannot proceed until metaphysics no longer has power over human politics with respect to its intrusion on consensus reality building.

As a realist and a transhumanist, I want these processes to proceed beyond the boundaries of human limitations. As a moral realist I believe this is a moral necessity as knowledge of morality is the primary barrier to implementation of morality. I do not apologize for this. I will follow the logic entailed by this without guilt.

REPLY ⬆ SHARE   ...

Deepa 20 hrs ago

Everyone doesn't want $7 trillion. For example, I don't.

After a certain point, when your basic needs are met, money doesn't add to happiness.

He might have a good reason. I don't.

REPLY (6) ⬆ SHARE   ...

Moon Moth 19 hrs ago

$7T buys a lot of "ruthlessly crushing my enemies underfoot".

REPLY (4) ⬆ SHARE   ...

Deepa 19 hrs ago

To what end?

REPLY (1) ⬆ SHARE   ...

BK 19 hrs ago

Their end, preferably.

REPLY (1) ⬆ SHARE   ...

Moon Moth 18 hrs ago

"What is best in life?"

REPLY (1)   ⬆ SHARE   •••

Arrk Mindmaster 14 hrs ago

Laminated women!

REPLY (2)   ⬆ SHARE   •••

Deepa 13 hrs ago · *edited 13 hrs ago*

I'll admit this is all funny :). Can't find the "Like" button.

REPLY (1)   ⬆ SHARE   •••

Fang 9 hrs ago

Not its primary purpose, but the ACX tweaks extension (which does useful things like parse markdown formatting into actual formatting) restores the "like" button on desktop

https://chromewebstore.google.com/detail/acx-tweaks/jdpghojhfigbpoeiadalafcmohaekglf

REPLY (1)   ⬆ SHARE   •••

Jeffrey Soreff 2 hrs ago

Could you tell me if this worked? I have acx-tweaks installed, and, when I "like" a comment, the heart symbol gets filled in and the count gets incremented. But, if I e.g. exit the post and go up to the main ACX page, and then back down into the post I was in, the heart symbol is no longer filled in and the count is decremented back to where it was. I have "liked" your comment. Can you see a "like" notification?

REPLY   ⬆ SHARE   •••

Moon Moth 9 hrs ago

Binders full, even.

REPLY   ⬆ SHARE   •••

Melvin 18 hrs ago · *edited 18 hrs ago*

Maybe. With $7 trillion you suddenly have a lot of principal-agent problems that don't exist at smaller scales.

Specifically, your giant private army can protect you against all external threats, but who is going to protect you from your giant private army when they decide it's much easier to throw you in a cell and steal your money?

REPLY (4) ⇧ SHARE

Slowday 14 hrs ago

Do I hear the sweet, sweet words "windfall tax"?

REPLY (1) ⇧ SHARE

Roger 13 hrs ago

Perhaps the words were "self-control." Maybe even "generosity."

REPLY ⇧ SHARE

Michael Watts 10 hrs ago · edited 10 hrs ago

> your giant private army can protect you against all external threats, but who is going to protect you from your giant private army

History suggests that the giant army can't really act on its own, and the threat to you comes from your officers. This causes the size of the problem to be similar at all scales - whether you are the petty king of a city-state or the ruler of a continent-spanning empire, you probably have broadly similar numbers of court officials.

REPLY ⇧ SHARE

John Schilling 10 hrs ago

That's what my giant private Navy is for. And Air Force, Marine Corps, and National Guard. I've read Machiavelli and Luttwak both; I think I have the basics at least down :-)

REPLY ⇧ SHARE

Level 50 Lapras 2 mins ago

coup proofing is a problem as old as history and there is a long literature on the subject.

REPLY ⇧ SHARE

Aristides 12 hrs ago

How many enemies do you have? I have a mother in law, but I could crush, or better yet ignore her with a lot less money

REPLY   SHARE

Mr. Surly 18 mins ago

But he's asking for $7 trillion to summon the demon that destroys us all! This seems like a tacit admission doom's not on the table (you don't pay $7 trillion to create something with a significant chance of paper clipping us all). Also, doesn't this mean that if we just invest 1/10 or whatever, the risk and timing of doom is significantly less? This is kind of like AGW folk jet setting about the world. Pick a lane!

REPLY (1)   SHARE

Moon Moth 5 mins ago

At this point, I think it's clear that Sam Altman believes that he can avoid doom. As with some other completely unrelated things: I hope he's right, I want him to be right, but I'm afraid he might be wrong.

REPLY   SHARE

anomie 19 hrs ago

What if my basic need is "eliminate all evolutionary life and replace it with the perfect self-improving lifeform"?

REPLY (3)   SHARE

Kori 19 hrs ago

Then it's great that you will not get these 7 trillions, I'd say.

REPLY (1)   SHARE

anomie 19 hrs ago

Then you should pray that Altman doesn't want that as well. Considering his recent actions, I would not be surprised if he no longer values humanity's existence...

REPLY (1)   SHARE

Kori 18 hrs ago

I think he will not raise nearly as much as he is asking, so there's no need for praying just yet.

REPLY ⬆ SHARE

Ryan W. Ryan's Newsletter 18 hrs ago

Then you probably can't generate the expected ROI needed to justify the investment.

REPLY ⬆ SHARE

Deiseach 10 hrs ago

Didn't Scott just give out a grant for tardigrades? There's your perfect life form, need being met!

REPLY (1) ⬆ SHARE

Schweinepriester 6 hrs ago

They are kinda cute, too.

REPLY ⬆ SHARE

Mo Nastri Mo's Reads 18 hrs ago

I first thought about this, but upon checking realized it's less than 1% of Sam's ask:

https://forum.effectivealtruism.org/posts/WeruztmEM53mLjbkL/copenhagen-consensus-center-s-best-investment-papers-for-the

I then thought about this: https://blogs.worldbank.org/voices/billions-to-trillions-financing-the-global-goals which seems to fit the scale.

This money would of course not be for me, but for what's needed to do the above. So maybe it's an overly loose interpretation of "everyone".

REPLY ⬆ SHARE

pozorvlak 16 hrs ago

> After a certain point, when your basic needs are met, money doesn't add to happiness.

This is a common misconception. What actually happens is that each marginal dollar of income adds less happiness, so the graph appears to be levelling off, but the increase never (at least within the range of incomes studied) actually goes to zero. It's a lot more obvious if you plot happiness against log(income). See the second section of this article:

https://danluu.com/dunning-kruger/

REPLY (1) ⬆ SHARE

Arrk Mindmaster 14 hrs ago

A difference that makes no difference is no difference. If you have no measurable increase in happiness from $50 billion to $1 trillion, then I think it would be good to approximate approaching a limit, the specific limit of which might vary per person.

And that doesn't even count decreasing marginal utility, though it's hard to see how that could apply to money.

REPLY (1)　SHARE　　　...

pozorvlak 13 hrs ago

> If you have no measurable increase in happiness from $50 billion to $1 trillion, then I think it would be good to approximate approaching a limit, the specific limit of which might vary per person.

No, this is not how limits work. Consider the series 1 + 1/2 + 1/3 + 1/4 + ... Each additional term is smaller, but the series does *not* approach a limit as you add more terms - instead, it diverges to infinity, eventually breaching any bound you set. In this case, satisfaction appears to be roughly proportional to log(income), so each dectupling of your income adds the same increment of happiness - going from $100bn to $1tn adds as much to your happiness as going from $10k to $100k, or $100k to $1mn, or $100mn to $1bn. Of course, it's likely that the model breaks down so far outside the domain on which it was observed, but that's what the model says.

> And that doesn't even count decreasing marginal utility, though it's hard to see how that could apply to money.

Wait, what? Decreasing marginal utility of money is absolutely a thing. I strongly suspect it's the main reason for decreasing marginal satisfaction. But we don't need to count it separately, because we can observe decreasing marginal satisfaction directly, thus baking-in any effect from decreasing marginal utility.

REPLY (1)　SHARE　　　...

Arrk Mindmaster 13 hrs ago

Um, yes, it is. Your series approaches infinity, but 1 +1/2 +1/4 + 1/8 ... approaches 2 without ever reaching it or growing larger.

You're right about decreasing marginal utility. I actually meant negative marginal utility, such as eating delicious pies until you're sick of them, and may never want to eat a pie again.

REPLY (1)   SHARE

pozorvlak 12 hrs ago

Right, some series whose terms tend to zero converge to a limit and some diverge to infinity, and some do neither. You can't "approximate approaching to a limit" without looking at the actual values. This is a genuinely hard problem that took mathematicians most of the eighteenth century to nail down.

REPLY   SHARE

Dawmol 15 hrs ago

I'm sure you could think of something satisfying to do with $7 trillion. Isn't there something you'd like to improve about the world? With $7 trillion, you could do things like end most global poverty, fund breakthrough research on deadly diseases, rebuild war-torn territories... Doing that would bring me happiness. Money isn't just about buying a yacht.

REPLY (3)   SHARE

Mr. Doolittle 12 hrs ago

There's the problem of unintended consequences. Let's say that you bought food for everyone during a famine - seems pretty straightforward good. Then you go visit the place six months later and find out all the bad things that happened with or because of the food (people fighting over it, farmers going bankrupt because they couldn't compete, whatever). Then you also notice that in the places where you bought food you jumped the prices up and millions of people are mad at you.

You may reason though and ultimately conclude that the world, in total, is better off because of your actions, but it doesn't change the fact that you are now responsible for the negatives as well. Some people can handle that or even be really happy with that, but a lot of people can't.

REPLY (1)   SHARE

Performative Bafflement 11 hrs ago · *edited 11 hrs ago*

Totally agree, I was reading Dawmol's list, and either 2/3 or 3/3 of his list is impossible with any amount of money. I think it's funny how some people think that enough money could solve complex, self-feedbacking problems that are hellish combinations of genetics, game theory, and Molochian coordination problems.

1. End global poverty. No, you can't. There's 60M millionaires, everyone else in the world is poor. That's ~7B people, your 7 trillion divided amongst them gives them all $1k, once. How much poverty did that solve?

Oh, you want to just focus on the remaining <$5 a day people? Fine, there's 2B of those, you give them all ~$3k once. That's what, 2 years to live on? Poverty solved! (Mission Accomplished banner unfurls)

And this doesn't even mention the fact that you haven't increased *productive capacity* anywhere in the world, which is literally the only thing that matters, not the bits and tokens we use to claim a part of that productivity. All you really did was raise inflation by $1k-$3k a person.

2. Breakthrough research on diseases. This is your best chance, but it's a long play - basically, there's already a current allocation of money and scientists studying it.

Expand full comment

REPLY · SHARE

Melvin 3 hrs ago

You're right, with seven trillion dollars I could theoretically solve any problem in the world. Which means that any problem that I haven't solved yet becomes my problem. I'd be like God... which means everyone will constantly be asking why I allow bad things to happen. I can't deal with this sort of pressure.

With seven *billion* dollars I could buy anything I could reasonably want, but maintain relative anonymity. I could make reasonably-sized donations to charitable causes I care about, without feeling like I'm obliged to solve everything. And only the people who know me

personally would come to me begging for money, instead of the whole world.

REPLY ⇡ SHARE

Mr. Surly 11 mins ago

$7 trillion is no longer a lot of money, and there's an easy way to confirm: that's Apple, Microsoft, and Tesla. Obviously enormous companies, super important, etc. But it's not hard to imagine a conglomerate of those three (or pick some others that sum to $7 trillion market cap). Basically, it's very Austin Powers to think "$7 trillion dollars!" is a lot. Further proof we've got insane inflation, a market bubble, etc., but that's a different discussion.

REPLY ⇡ SHARE

Deiseach 10 hrs ago

$7 trillion is indeed a lot. A mere million or ten, I'd manage to scrape by on that if someone wanted to drop it into my bank account 😁

https://www.poetryfoundation.org/poems/44380/contentment-56d2237388193

"I care not much for gold or land; —

Give me a mortgage here and there, —

Some good bank-stock, some note of hand,

Or trifling railroad share, —

I only ask that Fortune send

A little more than I shall spend."

REPLY (1) ⇡ SHARE

Deepa 10 hrs ago

1000 Likes!

REPLY ⇡ SHARE

Peter Gerdes Peter's Substack 20 hrs ago

Regarding R < 1 and R > 1, I feel like the imprecision about what exactly this means leads to an overestimation of AI risk. Indeed, I feel like the rationalist community should be particularly warry of the appealing fallacy of assuming numbers like this relate to what seems important to you.

I mean, look at Moore's law. In one sense you might say that looks like R > 1 but if what you care about is something like the ability to compute Busy Beaver of n (or Ramsey numbers) suddenly that exponential growth is actually horribly slow.

So what exactly are measuring in this case? It seems like what it measures is whether the next generation is economically appealing to build given the benefits of the current generation. But that could easily be true without it meaning much in terms of how dangerous the AI might be or how capable it is to do harm.

But this could easily be true if all each generation bought you was that it could help you speed up most of our computer code by a few more percent given the same silicon resources. No reason to think it corresponds to any vast ability to manipulate the world or escape from safeguards (Einstein in a prison isn't a great danger).

💬 REPLY (1)   ⬆ SHARE   ⋯

Vitor 12 hrs ago · edited 12 hrs ago

Seconding this. Talking of R values makes sense when you're dealing with a process that's *inherently* exponential (or S-shaped). You can't just fit an exponential to some arbitrary monotone function and call it a prediction. Especially when we have good reason to believe that this function is full of thresholds were you reach a hard limit on some resource that requires fundamentally new techniques (like the synthetic data mentioned in the post).

I've predicted here on ACX that GPTs would *not* grow at the exponential rate that some people naively predict. In particular, the economics of building the thing must make sense. There's no chance in hell that GPT-6 allows its owner to collect 30-100x as much revenue as GPT-5, *unless* it is AGI. My prior for that is very low; I know we don't all agree on this, but at least we should agree that this is the crux.

ETA: another pet peeve of mine is when people who have no idea how the algorithms work use "compute" as a noun and treat it as if it was an obviously fungible resource. It isn't.

💬 REPLY (3)   ⬆ SHARE   ⋯

Peter Gerdes Peter's Substack 12 hrs ago

I mean I broadly agree but I'd take issue with your claim about the revenue for GPT-6 vs GPT-5. Right now these models are clunky enough

that they draw a bit of revenue from search, writing suggestions, programming assistance etc. It's plausible that GPT-5 is roughly the same.

However, as you observe there are all sorts of threshold effects here. Suppose GPT-6 hits the point where it offers enough understanding to create a truly seemless translation experience. That could suddenly make it indispensable to billions of people who now see their audience as the whole world independent of language.

That's just one example. So while I agree with your general point, I think predicting economic return is a bit more difficult. I mean I had a palm pilot back in the day and in many senses the iphone was an evolutionary improvement but it crossed a threshold that made it orders of magnitude more profitable.

REPLY   SHARE   ⋯

Robi Rahman Quizzical Quandaries 9 hrs ago

Compute is also a noun, yes.

REPLY (1)   SHARE   ⋯

Vitor 5 hrs ago

An informal, ill defined noun.

REPLY (1)   SHARE   ⋯

Robi Rahman Quizzical Quandaries 5 hrs ago

No, it's very well-defined. It's the quantity of floating-point operations. You might measure differently depending on whether you account for the bit length of the numbers. And it's a mass noun.

REPLY (1)   SHARE   ⋯

Vitor 4 hrs ago

So, you're saying that compute is just a synonym of flop? I don't think so. It's used much more broadly than that, especially by the AI alarmists around here.

Regardless of any quibbles over definitions, flops as a metric ignore the cost of memory reads and writes, which become a much bigger problem at the proposed scale of

something like GPT-6, i.e., when the model is orders of magnitude too large to fit into GPU memory. Framing the issue as "larger models just need more compute" is swiping all those difficulties under the rug. Large-scale high performance computing is much more than throwing flops around.

💬 REPLY (1)   ⬆ SHARE   ⋯

Robi Rahman Quizzical Quandaries 4 hrs ago

I'm not saying compute is a synonym of FLOP. That's like claiming kilometer is a synonym of distance. You can measure distance in kilometers, and you can measure compute in FLOPs.

I agree that memory and interconnect can become more important as limiting factors to AI training as it is scaled up, which is why my team is hiring a distributed systems researcher to find out more about these bottlenecks: https://epochai.org/careers

💬 REPLY (1)   ⬆ SHARE   ⋯

Vitor 3 hrs ago

Ok, seems like we are actually in agreement. Still, there's lots of people out there who throw around these terms carelessly. One of the reasons I'm an "AI bear" is that I don't think scaling is as easy as all that.

💬 REPLY   ⬆ SHARE   ⋯

Skivverus Resonance Engines 7 hrs ago

Arguably innovation does have some exponential properties: the feedback cycle in "solve problem -> expand to fill newly solved space -> find next problem -> solve problem" looks suspiciously close to exponential when the problem you're solving is "the rate at which we solve problems isn't as high as we'd like".

That almost certainly has an asymptote, but not sure how you'd find it.

💬 REPLY   ⬆ SHARE   ⋯

DannyK 20 hrs ago

Serious question; do AI safety people still think Sam Altman is one of the good ones?

REPLY (4)   SHARE

anomie 19 hrs ago

Well, this is one of the top posts for the week on lesswrong:

https://www.lesswrong.com/posts/pEAHbJRiwnXCjb4A7/sam-altman-s-chip-ambitions-undercut-openai-s-safety

REPLY   SHARE

Laplace 18 hrs ago

Depends on the AI safety person. I never thought him being one of the good ones was likely, because he was one of the people turning this into a race. And after learning about his plans to resolve the chip bottleneck in summer, I was pretty sure he wasn't one of the good ones.

The gpu production pipeline being so bottlenecked is one of the few saving graces we have left, one of the last few hopes that make it maybe feasible for regulation to put a stop to the race and delay everything before we run ahead and die. I think widening that bottleneck is one of the worst things you could possibly do for safety at this stage.

Then there was the whole episode with the board, which convinced me that Sam is just the kind of person who will say anything in the moment if it gets him what he wants. So the mouth noises he makes mostly don't mean things and should be disregarded, only his actions matter.

He makes mouth noises about safety if that's convenient for political reasons, or for getting more skilled employees who care about safety onboard his AGI project. His actions, on the other hand, seem to suggest that what he wants is to be the first to build powerful AGI, as soon as possible. As far as I can tell, this has been consistent since the moment he entered the field.

REPLY (2)   SHARE

Deiseach 10 hrs ago · edited 10 hrs ago

The terrible thing about being a cynic is that people rarely fail to live down to your expectations. Guy who had the backing of big money won out over safety first board and got reinstated and the board overhauled

to be friendly to the direction he wanted to go, and *now* the nice people are beginning to wonder if maybe he's just a smidge less than 100% reliably sincere about safety first?

"Quokka" is a cruel term to use about the Rationalists, but man, it often seems that way. They're all genuinely nice people, the world needs more like them, but just intone the acceptable discourse words and they'll believe your good intent and hand over their cow for those magic beans. But Sam is involved with the EA community and he's a good egg, he couldn't possibly not mean what I think he means! Sam does not have to be a bad guy or even a liar to have what *he* means about "safety first" be vastly different from what *you* mean by "safety first".

REPLY   SHARE   ...

Dr. Y Dr. Y's Substack   4 mins ago

>one of the few saving graces we have left

Oh shush. You still have Taylor Swift. I'm sure the basilisk is a fan. I mean, isn't everyone?

REPLY   SHARE   ...

ryhime 16 hrs ago · edited 16 hrs ago

As a non-«AI safety» person, I am kind of impressed by such a bold attempt to scare away the investment. Will he pump the rumours of Google discontinuing Bard in three months next? If he succeeds, he has a shot at ChatGPT being an abusive stagnant monopoly for few years, he knows this craft after all...

REPLY   SHARE   ...

pozorvlak 16 hrs ago

Not an AI safety person, but I've always wondered what he's done to get in such a powerful position - he seems to have constantly failed upwards based on charm, smoke and mirrors.

REPLY (1)   SHARE   ...

anomie 10 hrs ago

Seems like you answered your own question. He is clearly very knowledgeable about AI, but charisma is what ultimately matters for securing leadership positions. Fortunately, having charisma is important

for leading effectively, but unfortunately, you need way more than that to lead competently.

◯ REPLY (1)　⬆ SHARE　　　　　　　　　　　　　　　　　⋯

pozorvlak　9 hrs ago

I guess, but AFAICT most Silicon Valley power-brokers have at least one big success behind them. Altman had a startup with a successful but unspectacular exit (apparently based on misleading investors), ran YCombinator for a while (but not, AFAICT, when any of its most impactful investments were made), then founded OpenAI. Why? Why him?

◯ REPLY　⬆ SHARE　　　　　　　　　　　　　　　　　　　⋯

Zach Stein-Perlman　Not Optional　20 hrs ago

> In general you can use synthetic data when you don't know how to create good data, but you do know how to recognize it once it exists (eg the chess AI won the game against itself, the math AI got a correct proof, the video game AI gets a good score). But nobody knows how to do this well for written text yet.

Note that we can cheaply determine whether code is good or not (does it pass the tests), and my weak impression is this is scalable, so synthetic code-y text seems promising as training data (for coding applications at least).

◯ REPLY (2)　⬆ SHARE　　　　　　　　　　　　　　　　　⋯

Matthias　19 hrs ago

Compare https://fsharpforfunandprofit.com/posts/property-based-testing/

However I agree that we can probably generate lots and lots of training data, especially if we add more modalities. We can run an almost arbitrary number of webcams and just point them at the world.

◯ REPLY (1)　⬆ SHARE　　　　　　　　　　　　　　　　　⋯

JamesLeng　18 hrs ago

Build solar power satellites, aggregate live feeds from their internal automated calibration processes, see how much it can figure out about plasma physics by observing the surface of the sun with an enormous compound eye.

◯ REPLY　⬆ SHARE　　　　　　　　　　　　　　　　　　　⋯

Kevin McKenzie 15 hrs ago · *edited 13 hrs ago*

No, you can't. In order to have a meaningful set of tests, you have to have a spec/oracle. The more complete the spec, the better the tests you can write, but the less information you're actually getting from each incremental test, and the closer you are to just having the code be the spec, which actually doesn't tell you anything useful about how the code *should* behave, just how it *will* behave.

Similarly, with synthetic data, you have to have a way of creating true, meaningful synthetic data, and we don't know how to do that. So you're feeding a bullshit generator more and more probable, but not definitely known to be, bullshit. What is it supposed to learn from this? This all has shades of "Deep Thought."

💬 REPLY (2)   ⬆ SHARE   ···

quiet_NaN 12 hrs ago

Exactly. Useful software generally interacts with the real world, so you just shift the problem.

As a more extreme case, one might consider (heuristic, NN) mathematical proof generation and (deterministic) verification. But even here, the real world decides what is "interesting math" to some degree. I think that it is unlikely that a NN would come up with the idea of inventing such useful concepts as real numbers on its own, unless your loss function is "how close is that theorem to something in an analysis text book".

💬 REPLY   ⬆ SHARE   ···

bell_of_a_tower 10 hrs ago

Agreed. Writing good tests is *harder* than writing good (at the same level) code. And writing either without good specifications (which are even harder to write) is pointless.

At work I'm working on a project that was inherited from a sibling R&D company. It was built basically without good requirements/specifications. Technically, it's got some cool bits. But working with it and maintaining it and *especially* extending it is utterly obnoxious. And it's obvious that the people who implemented the iOS and android versions never sat down and talked--they're architected completely differently.

Tests are no panacea.

Eremolalos Inkbowl 20 hrs ago

<You could try to make an AI that can learn things with less training data. This ought to be possible, because the human brain learns things without reading all the text in the world. But this is hard and nobody has a great idea how to do it yet.

Well yeah, we have no idea how to enable an AI to learn the way we do. We are able to learn plenty of stuff without reading massive amounts of text. I think that's because (1) our brains are not blank tablets -- they are already set up to learn certain kinds of thing. They have templates ready to be filled. (2) We are able to wring a lot of knowledge out of relatively few examples by reasoning processes, some of which we born with the abilityto do, others of which we can learn about from outside sources.

Scott's account assumes that all that is needed to produce ASI and AGI is larger data sets to train on (plus of course the compute and energy to do the trainings). But what about the possibility that there are some mental functions that are crucial for a machine to qualify as ASI or AGI that cannot be produced by training on larger data sets, because they are just not the kind of thing that can be built into a system by that method. Seems like there's a lot of overlap between those functions and the things that enable human beings to learn without gobbling massive data sets For instance, (1) goals & preferences that arise from within. I know you can give an AI a goal, such as start a t-shirt business, and it will do it. But in people goals and preferences are the product of biologically-determined drives as shaped by social learning. (2) Self-awareness and ability evaluate one's performance, learn from mistakes and change one's behavior. (3) Insight -- where one leaps suddenly from a mass of messy information to a higher level where

MartinW 20 hrs ago · *edited 19 hrs ago*

Also 3) we have ways of learning about the world which are not currently available to the AIs. If you want to know what "bouncing a rubber ball against a wall" means, you can read a lot of texts in which walls and rubber balls are mentioned, or you can spend five minutes playing with one.

Or if you want to know what "friction" is, you can learn the physics formulas involved, and ChatGPT is already better at that than 99% of humans. But what you can do and ChatGPT can't, is run your finger across different kinds

of surfaces and get an intuitive understanding of what it means in practice and why it's relevant. You need to read a lot of texts to build up that same knowledge.

Anything which involves knowledge about the physical world, we have the massive advantage of actually living in it and being able to interact with it and remembering those interactions next time the topic of rubber balls comes up.

REPLY (2)   SHARE

anomie   19 hrs ago

But the limitation in this case isn't AI tech, it's robotics tech. And robotics technology seems to be progressing quite nicely. The main problem is that for an AI to be able to do what you described, it would need to be able to learn after it finished training (effectively allowing it to self-modify). Unfortunately, that would not only be a massive security risk, it would also make alignment practically impossible because any alignment done via training could simply be overwritten.

REPLY   SHARE

Eremolalos   Inkbowl   9 hrs ago

Yes, I agree completely. AI is trained on words, which contain information about the world, but our species learns about Word and World both at once. And when I talked about how our brains had templates ready to be filled, the most important one I had in mind was the one where word and world data are integrated. But there are plenty of others. For instance the eye-hand coordination used in bouncing a rubber ball -- people don't have to each individually come up with the idea that what you see should guide where your hands go. We're wired to grasp that, and there's a template set up to store info you gather about how different kinds of bounces look, and what the way a particular bounce looks tells you about where to put your hands to catch the ball. And the proprioceptive data gets integrated into that info, too --stuff like what how hard you throw the ball tells you about what the bounce is going to look like.

But the most important category of information that AI lacks is Inner World. We don't just think and feel and crave things and judge things, we know that we do, and are capable of lots of meta in this area. We have feelings and judgments about our own mental processes, and have ideas

and cravings about ways they could be improved. We can form a mental image of someone else's inner processes, and modify it as new data comes in, and use it to predict what they will do. We are subject to glitches in our reading of other people, but most of the time we are extraordinarily accurate at predicting what others will do in common situations -- that's why they aren't constant fender-benders and constant incidents where strangers infuriate each other on the sidewalk, or customers and cashiers end up bewildered or in shouting matches. Or consider the results of the Reading the Mind in the Eyes Test, developed by a well-known autism researcher. (It's fun to take, & it's here: https://embrace-autism.com/reading-the-mind-in-the-eyes-test/). That test is *hard.*. When I took it I did not feel certain of a single one of my answers -- had to just go with my gut. But the average person gets 28 out of 36 items right. Even autistic people get about half of them right.

REPLY   SHARE

Bugmaster   20 hrs ago

Sadly, there's a chance that Sam Altman would nominally get his $7 trillion -- when hyperinflation reduces its purchasing power by a factor of 100,000x or more :-(

REPLY (1)   SHARE

Moon Moth   19 hrs ago

That's 7 of those giant coins, right?

REPLY   SHARE

Bill Benzon   20 hrs ago

I'll repeat what I said around the corner at the substack of He Who Must Not Be Named (https://garymarcus.substack.com/p/an-open-letter-to-sam-altman/comment/49344409):

So, here's the question: Altman's attempt at a $ 7T raise seems a bit extreme, even for him. The same with Hinton's recent hallucinatory diatribe against you [Gary Marcus]. Sutskever's been saying some weird things as well (http://tinyurl.com/9dm3fn6r). Are things on the Great Rush to AGI falling behind schedule? Are these guys getting just a bit worried and expressing it by doubling down?

REPLY (1)   SHARE

Deiseach 10 hrs ago

That's a good question. It may be that the people at the bleeding edge of research are beginning to hit speedbumps along the way, that the rest of us have no clue about yet, and they're casting about for a bruteforce way to get past them in order to keep the gravy train rolling smoothly down the tracks. Unless and until somebody spills the beans about "GPT 4.75 turned its toes up and went to join the choir invisible just when it was on the cusp of real thought", we won't find out until later (if ever).

○ REPLY   ⬆ SHARE   ⋯

Sei Wanderer in the Sea of Fog   20 hrs ago

The simplest explanation is that he never actually gave a shit about safety and now that he's finished purging the board in the wake of its failed putsch he's free to do whatever he wants. Anyone have another?

○ REPLY (2)   ⬆ SHARE   ⋯

anomie 18 hrs ago

Well yeah, obviously, but the question is why he's doing this. Is it greed? Vengeance? Blind ambition? Or is he genuinely trying to end humanity's reign? Though, ultimately his intentions don't matter, only his actions. It will be interesting to see how the AI safety people try to deal with this situation.

○ REPLY (3)   ⬆ SHARE   ⋯

Sei Wanderer in the Sea of Fog   17 hrs ago

I suppose it's either that he doesn't think it's a problem in any way and that we'll simply get alignment for free as capabilities progress, or else he thinks that complete AI supremacy in his own hands is the best way to get safety.

○ REPLY   ⬆ SHARE   ⋯

Deiseach 11 hrs ago · edited 11 hrs ago

Not that I know the guy or the sky over him, but I imagine it's a combination of blind faith in technology, and that to date betting on tech has made him a lot of money. I don't think he believes in superhuman agentic AI with its own goals that could decide in the morning to wipe out humans, but I think he'd be happy with "Star Trek style computer you can talk to as if it's a person, it will have access to all sorts of

knowledge and can perform simple tasks to free up humans" as a product to get to market.

And of course yadda yadda improve quality of life for all humanity yadda yadda since it'll be able to solve problems like clean energy etc. and we'll all (or at least the First World middle class will) have comfy lives from then on. Internet of Things, only for reals this time.

No reason why you can't get filthy rich *and* be a philanthropist, and if they just put enough guardrails up there will be no Danger AI.

REPLY ⬆ SHARE   ...

Mark 7 hrs ago
To quote Paul Graham "Sam is extremely good at becoming powerful."

REPLY ⬆ SHARE   ...

ryhime 16 hrs ago
Could he pre-purge the board right after getting rid of Musk?

The real question is what is he going to do once he fails to raise $7T. And does it involve more AI advancement or more abusive business practices to establish a stagnant monopoly.

REPLY ⬆ SHARE   ...

Moon Moth 19 hrs ago · *edited 19 hrs ago*
How much of the world would $7,000,000,000,000 allow someone to buy?

That's a lot of Twitters, or aircraft carriers. How many competitors could someone with $7T snap up? How many media organizations? What's the going rate for politicians? Small countries?

Not that Sam Altman will have $7T in untraceable cash, with no strings attached. But it might be interesting to speculate on what else could be done with the money. Graft, finance, scalable startups, and taxation all work on the principle of identifying a large flow of money, and siphoning off a little bit. How much graft would be possible with $7T, and what would it allow people to do? 1% of 1% is $700M, which would solve a lot of my life problems quite rapidly, and I'd guess the same would apply to all but a handful of people on the planet. But if I had that, I could buy Tumblr and make porn **mandatory** , just for kicks. (Or maybe I could save 1000 African children instead. **sigh**)

REPLY (5) ⬆ SHARE   ...

Matthias 19 hrs ago

Wasn't Tumblr sold for 3 million USD?

REPLY (2)   SHARE

Moon Moth 19 hrs ago

That's the number I saw, yes.

REPLY   SHARE

anomie 10 hrs ago

Yes, which is funnier when you consider that it was initially bought for 1 *b*illion.

REPLY   SHARE

Robert Jones 17 hrs ago

UBS says that global private wealth was $454 trillion at the end of 2022. So $7 trillion would be about 1.5% of all private assets.

Gross world product is about $105 trillion. So $7 trillion buys about 6.7% of everything the world produces in a year.

REPLY   SHARE

Moon Moth 8 hrs ago · edited 8 hrs ago

So far I've seen mention of ASML being ~$350B and TSMC being ~$700B. That would lock down the best, and give him opportunities to scale up while simultaneously shutting out competitors.

Jeff Bezos bought the Washington Post for $250M, 10 years ago, so something like that might have long term benefits for Altman's goals.

REPLY   SHARE

Moon Moth 7 hrs ago

nVidia is $1.5 trillion at the moment.

REPLY   SHARE

Mr. Surly 7 mins ago

Very little? As noted above, that's Apple, MS, and Tesla, give or take. That's a lot, but not comparison to every other company in world, let alone every sovereign wealth fund, let alone etc. It's easy to imagine a company/conglomerate the size of those three. Basically, people's sense of

value is above by about 10 these days (a million's nothing (particularly after taxes), etc.).

REPLY ⬆ SHARE

arbitrario 19 hrs ago

> (Sam Altman is working on fusion power, but this seems to be a coincidence. At least, he's been interested in fusion since at least 2016, which is way too early for him to have known about any of this.)

It is true that in 2016 he probably didn't know how much energy would have been required, otoh it is interesting to notice the he seems to have been well aware of the power of scaling since at least 2017:

https://blog.samaltman.com/the-merge

REPLY ⬆ SHARE

Tatu Ahponen Tatu Ahponen 19 hrs ago

...so, doesn't this just translate to "No, fast timelines aren't happening, yes, laws of physics and the finiteness of resources will hit you in the face, DoomHype was a real thing"? Perhaps even "winter is coming"?

REPLY (2) ⬆ SHARE

Moon Moth 18 hrs ago

My opinion is that current GPT-style LLMs aren't going to get us to AGI, and we'll need a few (1-2, I'm guessing) paradigmatic improvements to neural network technology, at about the same level as "deep learning with transformers". I'd bet that OpenAI is working on a number of potential ideas. But of course they're working on GPT-5 at the same time.

REPLY (1) ⬆ SHARE

Jeffrey Soreff 40 mins ago

I agree that some sort of improvements are necessary. I'm agnostic about whether they are likely to be substantial changes in the paradigm, or something closer to small amounts of wrapper code and prompt engineering, or something in between, like capturing just a small amount of metadata while training.

One huge imponderable that I see with any extrapolation exercise, as in Scott's post, is: If we _do_ get some sort of paradigm-or-wrapper improvement, how much scaling is really needed on _top_ of that change to get to AGI (in the can do anything an IQ 100 person can do sense)?

GPT-8? -7? -6? -5? or is GPT-4 actually large enough _now_, but we haven't had the insight/wrapper/paradigm-shift to use it to best effect? This makes a huge difference in how much pure scaling is actually needed, with all of the resource implications it carries.

REPLY   SHARE

MartinW   14 hrs ago

The assumption of predictable extrapolation is doing a lot of work, though. There's still the fact that we *know* that it is possible to achieve genius-level intelligence with fewer resources than a planet's worth of energy, semiconductors and training data, because there are human geniuses who have done it. There's no law of physics forbidding human-level (including Einstein-level) intelligence, and no good reason to believe that strongly-superhuman intelligence is fundamentally impossible.

It just means we need some approach other than "today's LLMs but with more hardware and more data". Undoubtedly OpenAI is working on it, although if Sam Altman is really going around begging for someone to give him $7T then that suggests that he isn't expecting any breakthroughs coming soon.

REPLY (4)   SHARE

Tatu Ahponen Tatu Ahponen   13 hrs ago

I was commenting mainly that precise approach.

REPLY   SHARE

procyon   12 hrs ago

Well, evolution plausibly consumed lots of energy.

REPLY (1)   SHARE

MicaiahC   11 hrs ago

Evolution could not create macroscopically large wheels and also is a super slow optimization process (for example, if some individual gene increases fitness by 1%, it'll take around 100 generations (not an endorsed number but is way more correct than "2" or "10" generations) whereas "mere" human level intelligence can just pick the correct one, after ignoring it for 10 generations and then blaming a coworker for not discovering this obvious improvement.

REPLY (2)   SHARE

procyon 11 hrs ago

It couldn't get wheels, but you'd have to prove that the situation is similar for intelligence. Neural net training seems closer to evolution than intelligent design; people come up with the training processes, not improvements on the mechanisms themselves.

Either way, it might be true that you can create genius level algorithms with very little energy; but it isn't because human geniuses require little energy, since they don't, it's because you found a better way.

💬 REPLY (1)   ↥ SHARE   ⋯

MicaiahC 9 hrs ago · *edited 9 hrs ago*

Rereading your comment, it looks like I've come up with lots of irrelevant detail as to your main point: Inference costs (cost of running a genius/training a model) can be much smaller than training costs (evolution/back propagation). And I think both of us agree on this point.

Feel free to ignore the rest of this post, keeping it up so people can see how point missing I was.

> Neural net training seems closer to evolution than intelligent design; people come up with the training processes, not improvements on the mechanisms themselves.

I think this is confusing different levels of abstraction.

Yes it's true that evolution exerts pressure at the meta level, and that when someone tunes a learning rate or architecture that's also a meta learning parameter, but that doesn't mean that the speed or quality of optimization are at all similar.

One training run is not a human generation long, and I would guess that for most ways to evaluate learning rate, gradient descent beats natural selection, since you don't have a mismatch between individual gene fitness and the objective function (I.e. evolution cannot evolve a gene which saves everyone from extinction but sacrifices itself)

I think that's why evolution has very little to say on performance bounds of modern ML (addendum: but I think this is why it's also a valid point of comparison to show the distinction between training and inference)

REPLY   SHARE   ...

Deiseach 11 hrs ago

Couldn't get wheels, did get rotifers:

https://en.wikipedia.org/wiki/Rotifer

REPLY (1)   SHARE   ...

MicaiahC 10 hrs ago

There's a reason why I gave the macroscopic qualifier! The flagellum would also be a counter example.

REPLY   SHARE   ...

Deiseach 11 hrs ago

"There's still the fact that we *know* that it is possible to achieve genius-level intelligence with fewer resources than a planet's worth of energy, semiconductors and training data"

True, but excuse me while I smile wryly about all the "intelligence is trivial, we know how to copy human level, AI is beating humans at games already, this is an easy problem, just throw enough hardware into the mix" claims that have gone before this point.

It took the planet and the arising of life all this time and effort to get human level, and you think you can just easily skip over the next few steps of superhuman level with a few motherboards?

REPLY (1)   SHARE   ...

anomie 9 hrs ago

To be fair, evolution is really, really inefficient. It's not as bad as putting monkeys on typewriters to try and get them to write Shakespeare, but it's pretty close. The nice thing about intelligence though is that it allows you to actually understand how the world works and make optimal decisions using that knowledge. That's going to make progress significantly faster than just trying completely random shit until you hit the jackpot.

REPLY · SHARE

myst_05 myst_05's Newsletter 10 hrs ago

> It just means we need some approach other than "today's LLMs but with more hardware and more data".

Sure but that approach might take decades or even centuries to develop. There's no immutable law of physics that says it must happen in 5 years.

REPLY (1) · SHARE

MicaiahC 10 hrs ago

Sure, what is the smallest capability you'd bet against in five years? What about the smallest ten capabilities?

REPLY (1) · SHARE

myst_05 myst_05's Newsletter 10 hrs ago

Smallest is "correctly solve a fresh set of IMO Gold problems": https://manifold.markets/Austin/will-an-ai-get-gold-on-any-internat

I think we're at least 10 years away from this happening.

REPLY · SHARE

Michael Watts 19 hrs ago

> You can train a math AI by having it randomly generate steps to a proof, eventually stumbling across a correct one by chance, automatically detecting the correct proof, and then training on that one.

We already know how to randomly generate correct steps. The problem is more in the area of getting results that are interesting. This strategy seems much like trying to train yourself in chess by placing pieces in random positions and noticing whenever one of the moves you just made happened to be legal.

REPLY (1) · SHARE

pozorvlak 16 hrs ago

> The problem is more in the area of getting results that are interesting.

Yes, and that's the problem that they're trying to solve here - training an AI with an intuitive notion of "promising next step" akin to that of a human mathematician. And it seems to be paying off:

https://deepmind.google/discover/blog/alphageometry-an-olympiad-level-ai-system-for-geometry/

REPLY (2)   SHARE

ryhime 15 hrs ago

Not sure _this_ is a good example, I tend to believe
https://old.reddit.com/r/math/comments/19fg9rx/some_perspective_on_alphageometry/ here (classical geometry problems have long been known to be vulnerable to pretty boring computational solutions, it's just that at IMO a human normally does not have enough spare energy to bruteforce one, but a computer can do it).

Maybe something like https://link.springer.com/chapter/10.1007/978-3-030-81688-9_6 is a better (but quite different in the approach) example.

REPLY (1)   SHARE

pozorvlak 15 hrs ago · *edited 15 hrs ago*

No, I think this still counts. "Apply all applicable deductive rules until you get stuck, then guess, and backtrack if your guess leads to a contradiction" is how theorem-provers work, and the quality of the guessing engine (traditionally, some collection of heuristics like the author of that Reddit post proposes) can make the difference between "finishes in an eyeblink" and "heat-death of the Universe". So using LLMs to make better guesses is absolutely a legitimate and potentially gamechanging contribution. Naturally, they're starting small with a well-understood domain for which we already have good constraint-propagators, but their next step is presumably to apply it to a more open-ended proof system like Lean (https://en.wikipedia.org/wiki/Lean_(proof_assistant)).

REPLY (1)   SHARE

ryhime 15 hrs ago

Sure, with Lean it's the real stuff, with matrix multiplication optimisation it's the real stuff! It's just that geometry problems are a typical example of problems that are only interesting because of human limitations. Brute-forcing by complete enumeration is not that far out of human reach in most of those problems (and computers get a handful of decimal orders of magnitude bonus at that).

REPLY ↑ SHARE

Michael Watts 10 hrs ago

But Scott's description of the problem explicitly states that you generate proofs without regard to whether they are valid and then run validity checks. That's a completely different approach, and I don't think it stands much chance of beating the method you're referring to.

REPLY (1) ↑ SHARE

pozorvlak 9 hrs ago

Oh, I see. Yes, you're right, he did say that. But I think it may have been a slip of the pen, and what he *meant* was something more like what I'm describing - using an AI to pick individually-correct steps, and doing RL based on whether they lead to a correct (and ideally short) proof of the desired statement.

REPLY ↑ SHARE

Hoopdawg 19 hrs ago

As a skeptic who assumes the R is way below 1 (barring some conceptual breakthrough that pretty much requires we first jump off the "all resources into LLMs, surely scaling will fix everything" train), I spent most of the read-through wondering whether the punchline will be "maybe the real singularity is the computational capabilities we've made along the way" or "so it won't be paperclips, but LLMs".

REPLY ↑ SHARE

Cry6Aa 18 hrs ago

Is it just me, or do all of these "master-manipulator ubermensch CEO" types just come off as complete goobers the second they step out of their lane. I mean, this Sam Altman character sounds like a credible Machiavellian figure right up until the point where he opens his yap to ask for money, and then he sounds like Dr Evil. Which retroactively makes all his other accomplishments sound like a combination of dumb luck and simply being in the right place at the right time.

Why do we take these folk seriously?

REPLY (1) ↑ SHARE

JamesLeng 18 hrs ago

They were overtrained on a narrow dataset.
https://en.wikipedia.org/wiki/Ne_supra_crepidam

◯ REPLY   ⬆ SHARE   ···

[insert here] delenda est 18 hrs ago

For cooling and efficiency reasons, not to mention security, gpt7, probably even 6, will be built in space. Which drives the costs up, and gives the advantage to Elon Musk.

◯ REPLY (2)   ⬆ SHARE   ···

Tossrock 17 hrs ago

Cooling is much harder in space than on Earth, because you can't use conduction/convection to the environment, evaporative, etc. You can only use radiative. Now, a deep lunar crater in permanent shadow... maybe.

◯ REPLY (1)   ⬆ SHARE   ···

[insert here] delenda est 17 hrs ago

I may be being very naive here but i would have thought that you could generate a _lot_ of heat before you needed to cool anything if you were on the shadow side of earth ?

◯ REPLY (3)   ⬆ SHARE   ···

Kevin McKenzie 15 hrs ago

No, because, again, the heat has nowhere to go. Heat is entropy/motion/disorder/energy, however you want to think of it. In order to cool something down, the energy has to go somewhere else. In a near vacuum, there's no good place for the energy to be transferred to. If you're on the moon, you can transfer the energy into the moon's mass, but then you have other problems.

◯ REPLY (1)   ⬆ SHARE   ···

[insert here] delenda est 15 hrs ago

Thank you. I was also thinking of the moon, especially as it fixes the shadow problem, what problems does this create?

◯ REPLY (1)   ⬆ SHARE   ···

Celarix 7 hrs ago

No air outside of your base. You can soak waste heat into the moon, but you'd have to wait for it to conduct away, whereas with air or water, you can physically move hot air away and bring in fresh, cooler air. Once the surrounding lunar rock is as hot as your servers, you can't pump more heat into it until it conducts away/radiates into space, which is SLOW.

REPLY ⬆ SHARE

Alastair Williams The Quantum Cat 15 hrs ago

It's almost impossible in terms of orbital mechanics to stay on the shadow side. And even if you somehow managed it, where would you get the energy to power the servers from? You wouldn't be able to use solar panels. And even then, you'd still heat up after a fairly short time, and you'd need a way to get rid of that heat. As Kevin points out, that is a hard problem to solve in space.

REPLY (1) ⬆ SHARE

[insert here] delenda est 15 hrs ago

The energy is nuclear reactors. What about the moon ?

REPLY ⬆ SHARE

beleester 15 hrs ago

Being on the shadow side of earth just means that you don't get heated by the sun. That doesn't mean much if you're generating a lot of heat from onboard equipment.

I looked up some numbers (thanks, Atomic Rockets), and if your radiators are at 363 Kelvin (90 C, the temperature of a normal CPU at full blast), you'd need over a thousand square meters of radiator per megawatt of power used (since all power eventually becomes heat). You can do better if the radiators are glowing hot, but that makes getting heat from the CPU to the radiator kinda complicated.

REPLY (2) ⬆ SHARE

[insert here] delenda est 14 hrs ago

Thanks! Much appreciated

REPLY ⬆ SHARE

ryhime 14 hrs ago

Hmm, ~ 1kW/m² sounds pretty close to server room densities (it seems to be considered low density for modern data centers, and then this is for a single floor). Doesn't sound good for things that need multiple full-size data centers...

REPLY   SHARE   ...

Mark P Xu Neyer (apxhard)   apxhard   15 hrs ago

Maybe would build it under the ocean. That's what I call, deep learning!

REPLY (2)   SHARE   ...

[insert here] delenda est   15 hrs ago

Seriously an option and based on other responses might actually be better for cooling !

REPLY   SHARE   ...

Deiseach   11 hrs ago

Lunar base, guys. This *is* the 21st century after all!

REPLY   SHARE   ...

ffk   18 hrs ago

Hi, French subscriber of ACX here.

Completely unrelated to your post, but there is an small issue on ACX when trying to access through "astralcodexten.com" instead of "www.astralcodexten.com": visitors see a "Test website, please ignore" page instead of the blog.

The convention is to host on either "www.website.tld" or "website.tld", but to also add a redirect in a direction or another (through a DNS CNAME record for instance).

I think it's trivial to set it up on Hostgator, but please let me know if you need help with it.

REPLY   SHARE   ...

Ryan W. Ryan's Newsletter 18 hrs ago

"GPT-7 will need fifteen Three Gorges Dams."

Wow! That's like, forty five Gorges Dams!

💬 REPLY (1)   ⬆ SHARE   ⋯

Deiseach 11 hrs ago

AI Gorges, who lives in cave & eats over 10,000 GW each day, is an outlier and should not have been counted

💬 REPLY (2)   ⬆ SHARE   ⋯

Moon Moth 8 hrs ago

applause

💬 REPLY   ⬆ SHARE   ⋯

Celarix 7 hrs ago · *edited 7 hrs ago*

Given the power of SI, we can actually convert 10,000 GW into a measure of caloric energy found in food!

1 Watt = 1 Joule per second

1 thermochemical calorie = 4.184 Joules

1 food Calorie = 1,000 thermochemical calories = 4.184 kilojoules (or is it kiloJoules?)

Since AI Gorges eats a... oh, it's AI, not Al as in Albert. Darn fonts. But if AI Gorges was a person, they'd eat a sustained 10 terawatts over the 86,400 seconds of a day, which is a total of 240 TWh, or 206.5 billion Calories. That's enough Calories to meet the dietary needs of 103 million people, or 283 million Double Quarter Pounders, or almost 23 million metric tons of fat, enough to fill a thousand 20-foot shipping containers or so. Roughly one giant shipping boat filled with solid fat per day.

That's a lotta power.

💬 REPLY   ⬆ SHARE   ⋯

MichaeL Roe 18 hrs ago

https://www.youtube.com/watch?v=MS2AXT5MSgk

It would appear that Sam Altman is imitating Dr Evil from the Austin Powers movie.

REPLY   SHARE   ...

David Chambers 18 hrs ago

The thing about building a world-conquering AGI you control is that building a world-conquering AGI you don't control is in a sense the same result as not building anything at all: Either way, you don't control the world.

Note that from the viewpoint of someone who is not Sam Altman, it might not matter whether the world-conquering AGI is safe or not: Either Sam Altman or the AGI will conquer the world. Either way, you and everyone you care about will be conquered.

REPLY (1)   SHARE   ...

MartinW 15 hrs ago · edited 15 hrs ago

Being conquered by Sam Altman with a tame AGI might be slightly less bad than being conquered by an out-of-control AGI following some zany optimisation target. Unless maybe you are Sam's childhood bully or something.

It might even be less bad than the current situation, where you are also ruled by powerful people following their own interests (and if you are lucky enough to live in a democracy, then once every few years you get to have a tiny amount of influence on which out of a small selection of powerful people you prefer to be ruled by).

Benevolent dictators can work; two big problems are that a) they tend to get deposed in military coups, and the coup leaders may be less benevolent; b) they eventually grow old and die, and their successors may be less benevolent. But if we're going with the AI-as-a-tame-God scenario where it can make its owner immortal and invulnerable, then neither of those would be a problem. It certainly doesn't sound ideal, though.

REPLY (1)   SHARE   ...

Deiseach 11 hrs ago · edited 11 hrs ago

This is starting to sound like a Dean Koontz novel, his last run of them have all been along the lines of "small group of very rich, very powerful,

very smart people who think they are the ones who should be in charge and have utter contempt for the rest of us scheme and plot via nefarious means, including controlling government agencies and working on superscience projects, to reshape the world as they would have it". AI tame god to make its owner(s) immortal and invulnerable is right up his street, plot-wise.

They're fun novels, especially when I recognise references he uses that are a couple of years behind the first time I heard them, but I wouldn't want to live in those worlds.

💬 REPLY (1)　　⬆ SHARE　　　　　　　　　　　　　　　⋯

David Chambers　10 hrs ago

In the nature of things, anyone who thinks that super intelligent AGIs are existentially dangerous, and wants to build one, is likely to be similar to the bad guys in such a novel.

In the meantime, there is the question of how much benevolence you can expect from a dictator who was willing to risk the end of humanity to make himself dictator.

💬 REPLY　　⬆ SHARE　　　　　　　　　　　　　　　⋯

Melvin　17 hrs ago

What does all this mean for AI Foom?

In these terms, AI Foom would require R to become very much greater than 1 once you get past a certain point. This would presumably have to mean that there's some far more efficient way of training an AI that we are currently too stupid to see, but which a sufficiently advanced AI will be able to figure out.

On second thoughts, proper foom might even require a whole bunch of these step-function insights so that R can stay consistently above 1.

There was always a shoddy step in the argument for Foom, which went something like "if a human can build an AI, then an AI smarter than a human can build an even smarter AI". But it doesn't have to be this way -- humans won't build AIs through flashes of insight but by a painstaking mathematical process, and the smarter AI won't necessarily have better ideas on how to optimise that process.

💬 REPLY (1)　　⬆ SHARE　　　　　　　　　　　　　　　⋯

Donald 13 hrs ago

>humans won't build AIs through flashes of insight but by a painstaking mathematical process, and the smarter AI won't necessarily have better ideas on how to optimise that process.

That doesn't make a huge amount of sense. Are you suggesting that intelligence only helps on the problems that feel "eureka" rather than grinding maths slogs? Both of those things are done in the brain. Neither can be done by IQ 80 humans. Both take intelligence.

I mean you can conjecture that the current mix of techniques we are using is optimal. But that's a pretty bold unsupported conjecture. AI seems to have a fairly reliable track record of humans inventing better techniques. So saying that, at the same moment a much smarter AI is doing the theory, theoretical progress suddenly stops. It's not impossible, but not the way to bet.

REPLY (2)   SHARE   ...

anomie 9 hrs ago · *edited 9 hrs ago*

Well, there is eventually going to be a point where progress stops because capabilities become perfectly optimal under the current laws of physics. ...We are not anywhere near that point yet.

REPLY   SHARE   ...

Melvin 3 hrs ago

> Are you suggesting that intelligence only helps on the problems that feel "eureka" rather than grinding maths slogs? Both of those things are done in the brain

Brains can do basic arithmetic, but they are far from being able to do the kind of grinding maths slogs required for training models. We shouldn't confuse "computer acting like a brain" with "computer acting like a computer".

REPLY   SHARE   ...

Por Poisson 17 hrs ago

Isn't this basically the nail in the coffin for the original/standard AI-doom prophecies? You're not going to have: (1) a million minds running in parallel or (2) self-replicating model spreading to every networked system or (3) runaway self-improvement to the singularity

if the training cost of the next model class is 7 trillion

Now that there's a whole funding ecosystem dedicated to the question of "will AI be the end of us" I don't doubt that we can think of new hypothetical scenarios where that's possible, but my takeaway is that "the world models that led people to worry about AI x-risk were badly mistaken"

To be sure -- this is not saying there are no concerns with AI, just that it's a damper on AI _x risk_.

⬭ REPLY (4)   ↥ SHARE   …

Guy 17 hrs ago

Running them takes a lot less than training them.

⬭ REPLY (1)   ↥ SHARE   …

gwern Gwern.net Newsletter 10 hrs ago · *edited 10 hrs ago*

Yes, this is the ironic effect of delays in the neural net overhang. (Note that this is specifically due to the extreme asymmetry of train vs run in neural networks as we understand their optimal scaling at present where parameters/forward-pass FLOPS grow slowly: it would *not* be true of many ex-competitor AGI paradigms.) The further back you push the 'final' model, the more of an overhang you have because the final model is increasingly small compared to the resources it took to train it. If your GPT-7 takes a hundred million GPUs to *train*, then it still will only take like <100 GPUs to *run*. So the instant you stop training, you don't have '*a* GPT-7', you have 'millions of GPT-7s running 24/7/365 forever' on just the original training hardware, never mind any additional hardware they then buy, build, rent, borrow or steal or the additional capacity they get by distilling/pruning themselves down to much smaller faster cheaper models (as you see OA regularly doing with their 'turbo' models).

⬭ REPLY   ↥ SHARE   …

Donald 13 hrs ago · *edited 13 hrs ago*

It shows that it won't happen like that, if the techniques being used are that inefficient.

Suppose a huge effort is made. A vast datacenter to run GPT7. Then this AI is smart. Smart enough to do some impressive AI theory. It produces a new design, about as smart as itself, but able to run on an old smartphone due to some major efficiency improvements.

Or maybe a human invents that much more efficient AI.

What this is saying is that the self replication is fairly limited, if we are only using techniques as inefficient as current LLM's.

And of course, there are also scenarios where a single AI destroys the world without running on every smartphone.

That GPT7 could design nanotech and trick someone into building it, all while it was running on the only data center big enough.

REPLY   SHARE   ...

myst_05   myst_05's Newsletter   10 hrs ago

We might still get AI-doom, just decades or centuries later.

REPLY   SHARE   ...

John Schilling   10 hrs ago

It's a nail in the coffin for the version of that prophecy where the first IQ-150 AGI tinkers around in the background secretly designing the IQ-170 AGI which overnight builds the IQ-190 AGI, etc, etc.

It still allows for the version where the IQ-150 AGI makes bignum copies of itself who all team up to paperclip the world or whatever. But, A: the real world seems fairly robust against being taken over by merely IQ-150 general intelligences and B: this would require that the IQ-150 AGI believe, and be correct in believing, that it can bind AGIs to its will, whereas plenty of IQ-150 HGIs seem to think that's a dubious proposition.

REPLY   SHARE   ...

JohanL   17 hrs ago

I think it's completely fair to say that all the "AI safety" work has dramatically increased AI risk.

It's interesting but probably fruitless to wonder if this was intentional or not.

REPLY   SHARE   ...

Ben   Ben's Comedy News   17 hrs ago

Would an LLM trained on a grab bag of data even be capable of providing insights into how to build better chips that the best human engineers wouldn't already have thought of? To me it seems like GPT-4's ability to write code is based on pattern matching from existing code - it can write some passable code for a login screen because it's seen plenty of examples of that. But if you're

asking it do something cutting edge, like dramatically improve over the best GPU designs available, its reference data isn't going to provide the right answers.

I'm sure if you gave GPT-5 a cutting edge chip design and told it 'improve this to use 50% less power', it could come out with some plausible looking answer, but in the absence of any examples of improving a cutting edge chip in its training data, surely you could only get plausible bullshit?

Maybe a big enough LLM starts to develop actual first-principle insights, but it seems to me like a specialist chip design AI, trained on all the latest prototype chip designs and information about their thermal properties and manufacturing problems etc, would be more likely to help?

⬭ REPLY (3)   ⬆ SHARE   ⋯

Mark P Xu Neyer (apxhard) apxhard   15 hrs ago

Yes! LLM's are helping disabuse more and more people of the idea that sounding articulate and confident is a good proxy for being competent.

⬭ REPLY   ⬆ SHARE   ⋯

pozorvlak   13 hrs ago

Its training data would presumably include every computer architecture and photolithography paper ever published, so it's possible an LLM could be used to find a connection between two ideas that nobody had ever spotted. More interestingly, you could train it to look *away* from its training data, to find research blindspots that human scientists had ignored. There's already quite a lot of work being done along these lines:

https://www.axios.com/2023/08/17/ai-speeding-science-new-discoveries

⬭ REPLY   ⬆ SHARE   ⋯

Deiseach   12 hrs ago

I suppose the only way to find out is ask it to design a chip, build that chip in real world, and see if it works or not. Good old trial and error, no magic "the AI is so smart it taps into the infallible gnosis of the universe".

⬭ REPLY (1)   ⬆ SHARE   ⋯

1123581321   4 hrs ago

You wouldn't even need to build it. Just simulate its design in Cadence.

⬭ REPLY   ⬆ SHARE   ⋯

Rambler Ramblering 17 hrs ago

We need to explore the racket hypothesis a little more.

REPLY (1) SHARE

jpt4 15 hrs ago

What is the racket hypothesis?

REPLY (1) SHARE

Rambler Ramblering 10 hrs ago

Pretty simple: Altman knows he needs to use the moment he has the most leverage to up his profile, and it is now.

Even if he knew for a fact GPT-5 would be a waste of time and money he would try to maximize investment like this.

REPLY (1) SHARE

jpt4 6 hrs ago

I see, I was thinking elliptical thoughts about multi-armed bandits, and confederacies of such.

REPLY SHARE

Xpym 17 hrs ago · *edited 17 hrs ago*

>if GPT-5 is close to human-level, and revolutionizes entire industries, and seems poised to start an Industrial-Revolution-level change in human affairs, then $75 billion for the next one will seem like a bargain.

How likely does the ML field consider this scenario to be? Seems pretty pie-in-the-sky, but I'm not particularly clued-in.

Also, it's pretty suspicious that the hype around multimodality has died down lately. True multimodality has to be essential for anything like human-level, and there's no evidence that we're anywhere near that.

REPLY (1) SHARE

procyon 16 hrs ago

One year ago you'd hear a lot about multimodality, synthetic data, scaffolding etc.. nowadays not so much, only synthetic data and I haven't seen anyone manage to improve models with it (not counting things like generating data with GPT-4 to train a smaller model).

Either way, I wonder what Scott mean by "human-level". GPT-4 is pretty knowledgeable already, more so than the vast vast majority of people, and it doesn't really make any major mistakes anymore. What sort of improvement is he expecting here?

○ REPLY (1)   ↑ SHARE   ...

Mark P Xu Neyer (apxhard) apxhard 15 hrs ago

I use GPT-4 often enough to catch it making all kinds of mistakes. It's like a New York Times journalist tries to write code by bullshitting. For simple things it works well enough to save me time. But for anything remotely complex it's not currently worth the effort.

○ REPLY (1)   ↑ SHARE   ...

procyon 14 hrs ago

I meant major mistakes as in very glaring issues (like you get with smaller models; obvious logical errors, grammatical mistakes, nonsensical continuations). If you do anything technical with it, and for long enough time, you'll notice the flaws. They'll be subtler, like saying something that doesn't quite make sense given what it already said or making a mistake that you wouldn't notice while skimming, but they'll be there.

My thinking is that, the more data you give to it, the more it has to learn, and since the majority of data won't be "useful stuff" (e.g science, engineering etc), it'll spend most of its training time trying to learn useless patterns. So I expect improvement to slow down in the areas we care about unless effort is made to make sure the model is always learning useful stuff.

Which can be done, of course. But at some point you'll run out of useful data, and repeating it can't be done forever.

○ REPLY   ↑ SHARE   ...

Ergil 17 hrs ago

Assuming this sort of scaling, wouldn't GPT-7, or even GPT-6, have inference times too long to be of any real use?

○ REPLY (1)   ↑ SHARE   ...

procyon 8 hrs ago

I guess he's assuming computational capacity will increase to a point where this isn't a problem.

REPLY   SHARE

<unset> 16 hrs ago

Let's assume for a moment that, at some point, creating a world-changing superintelligent AI is going to be obviously within reach, such that it's worth dedicating a large fraction of humanity's economic output to producing it. At that point, if we're not *already* dedicating a large fraction of humanity's economic output, we're going to start doing so, abruptly, using up that compute overhang and suffering the safety risks that go with rapid progress.

So, arguably, the safest option is to use up the compute overhang *now*. Build your way up to GPT-7 with $7t or so, crossing your fingers and hoping that this isn't enough for it to be superintelligent. Then there'll be no more compute overhang, so no more scope for rapid progress by increasing the resources dedicated to AI: just safe, careful, incremental progress.

Summarising this argument: the compute overhang means there'll probably be a sudden jerk of progress at some point. Better that it be now, when it's least likely to catapult us over the threshold.

This seems to me to be a plausible reconciliation of Sam Altman's advocacy for increased AI funding with his concerns about compute overhang.

REPLY   SHARE

Sleakne 16 hrs ago

I for one would love a post about what Scott would do with a 7 trillion dollar windfall

He talks a lot about problems that could be solved by the Tsar of X. Would 7 trillion buy Tsar status

REPLY (1)   SHARE

Deiseach 12 hrs ago

I'm old enough to remember when talking about trillions was fantasy stuff, even billions for a national economy was wild talk. Now we've pretty much adopted globally the American definition of a billion (a thousand million) and even individuals have their wealth measured in billions.

I honestly can't decide if that means "progress!" or if it just means money ain't worth spit no more.

Boinu Badly 1 hr ago

You can certainly get the impression money and valuations have become unanchored from reality and the real bottlenecks are (as some of us have quaintly continued to insist) the availability of labour and the means of production.

Others in the comment section have discussed how scaling up compute becomes difficult regardless of available cash, but I'm also thinking of two of the three main resources of the war in Ukraine - manpower and artillery shells. NATO's GDP dwarfs Russia's, but no amount of money will simply create shell-making capacity without lengthy ramp-up, or increase the Ukrainian effective-manpower pool. The shatter-dirt, the steel, the hands, these still matter, regardless of what number goes into the spreadsheet.

M T 16 hrs ago

This sounds like total madness to me. We are having a climate crisis unfolding pondering how to produce our energy sustainable, and yet we pursue a path to develop a system that we don't understand, that is more likely to harm us than to benefit us and that is unimaginably hungry for energy. What's wrong with you people?

M T 16 hrs ago

That's not to say that there is something wrong with you Scott, you are are just doing the math and adding some thoughts to it. It's more of a general statement

Deiseach 12 hrs ago

Oh, Scott is honest (and I think beginning to be a little sceptical about all the promises, promises) and is on the side of "maybe we should be careful and slow about this, guys?"

But that's not the mindset of the businesses shoving money into "we need to get our model out there on the market", see Google announcing Gemini this week etc.

REPLY   SHARE

ryhime 15 hrs ago

How else to build out nuclear baseload power if not by raising money for it under false pretenses, though!

REPLY   SHARE

Mark P Xu Neyer (apxhard) apxhard 15 hrs ago

Using energy doesn't cause the climate crisis. If we had massive amounts of cheap, clean, energy it would be enormously beneficial for the world. We could have that if we simply removed the political taboos around nuclear power.

REPLY   SHARE

Deiseach 12 hrs ago

"What's wrong with you people?"

(1) They don't believe in God or miracles, that is very stupid you see and only stupid people believe that. They believe in science, Fairy Godmother AI that will be super-humanly intelligent and will uncover the secrets of the material plane and create ex nihilo abundance for all, including limitless free energy, free goods, and riches. Totally different from believing in the Sky Fairy Daddy.

(2) Money money money, it's a rich man's world. Maybe this will bring about the end of humanity, but that will take at least six hours and in the meantime if we're the first to market, our stock price will go to the Moon, Alice, to the Moon!

REPLY (1)   SHARE

Bugmaster 3 hrs ago

I mean, yeah, I don't believe in God or miracles, which is why I don't believe in the Singularity either. Which is why I severely doubt that AI is going to "bring about the end of humanity". Humanity is much more likely to do that.

However, it is starting to look like AI will bring about the end of creativity and process reliability -- at least, for a little while.

On the creative front, I could pay a human writer a salary over many years to write something like the *Lord of the Rings* trilogy... Or I could

click a button and have Chat-GPT do it in an instant for $20/month (or however much it costs). Sure, its results will be bland and boring and maybe incoherent at times, but who cares ? All it cost me was $20, and if I can make even $1000 selling that book, and I can do it 1000x per minute, my choice is clear. The same applies to art and, very soon, movies.

On the services/process front, more and more people are delegating more and more of their core business processes to ChatGPT, for the same reason as above. You no longer need customer service reps, technical writers, or even lawyers -- all you need is that button and $20. This means that your customers (and business partners) are going to experience hallucinated answers to their questions, and your code will be full of random bugs, and maybe your car will randomly drive very slowly into a wall sometimes... but hey, it only costs $20 !

So yeah, I think that in the immediate future we're in for a generally shittier life experience; not in on big dramatic way, but in lots of little ways that will continue adding up, until either this GPT bubble bursts, or I suppose until the Singularity takes us all to simulated heaven or Jesus destroys the Earth (Revelations-style) or whatever.

REPLY SHARE

Mark P Xu Neyer (apxhard) apxhard 15 hrs ago

> if each new generation of AI isn't exciting enough to inspire the massive investment required to create the next one, and isn't smart enough to help bring down the price of the next generation on its own, then at some point nobody is willing to fund more advanced AIs, and the current AI boom fizzles out (R < 1)

Was this case predicted before by anyone studying AI risk?

I didn't articulate it this simply but, having worked at Google in data centers, I figured the problems of scaling up compute are harder than people think, and think of these LLM'a as indeed being cool but they are - like most of the history of AI - not everything we are doing.

REPLY SHARE

John R Ramsden 15 hrs ago

GPT and similar systems seem to be built by shoveling vast heaps of facts at them and then trying to treat them as an oracle, to be bombarded with questions.

Maybe it would be more fruitful and efficient to make this more of a two-way process, like educating a human student, whereby it isn't just a matter of stuffing the student with facts which they can later adapt and regurgitate, but addressing their feedback on aspects of which they are unsure.

On that assumption, a useful crowd-sourcing initiative might be for volunteers to accept, and answer to the best of their ability, questions _from_ GPT. Sincere and helpful answers could be rewarded with free tokens. Conversely, frivolous answers, as detected by a wide divergence from the majority of answers to the same question, could be sent for human review and if judged to be deliberately misleading the volunteer could be shadow-banned or thereafter sent dummy questions whose answers would be ignored.

GPT could be asked for regular summaries of what it thought it had learned from this exercise, almost as if it was taking an exam!

The question though is whether GPT has the ability to self-reflect and be aware of its own blank areas or areas of its knowledge in need of improvement.

💬 REPLY   ⬆ SHARE   •••

jpt4 15 hrs ago

> But nobody knows how to do this [validate the quality] well for written text yet.

Then what is needed is a Literary Critic LLM, an A[I] B Walkley [0] LLM.

[0] https://en.wikipedia.org/wiki/Arthur_Bingham_Walkley

💬 REPLY   ⬆ SHARE   •••

Sam 15 hrs ago

Wow, those guys are under a lot of collective pressure of expectations to create AGI. I want 7 trillion too.

💬 REPLY   ⬆ SHARE   •••

Calvin 14 hrs ago

A lot of the stuff in this post kind of seems like magic. "We just need $7 trillion dollars, more computing power than currently exists, God only knows what infrastructure, and more information than the human race has ever produced, and we'll be well on our way to a super intelligent AI!"

💬 REPLY (1)   ⬆ SHARE   •••

Deiseach 12 hrs ago

I suppose it did take the creation of the entire universe and nearly four billion years to get human level intelligence on Earth, so "smarter than a human" AI requirements are lesser by that measure - only $7 trillion dollars, the entire electrical power output of the planet, and every economy on the globe switching to chip fabrication? A trifle! 😁

REPLY (1)   SHARE

Bugmaster 2 hrs ago

And in exchange for all of that you get an incrementally less buggy chatbot...

REPLY   SHARE

Arrk Mindmaster 14 hrs ago

Of course, it's hard to predict what leaps might happen to reduce these assumptions, but a big one missing is quantum computing. The very nature of it, as I understand it, reduces exponential computation to linear computation. It has many hurdles to leap before it can be involved in AIs and LLMs, including a viable program for integrating quantum computing into the AI paradigms we currently use, cost of individual qubits, and others. But it may well make the costs, energy use, AND training data needed smaller than estimated.

REPLY (1)   SHARE

Ekakytsat 5 hrs ago

> reduces exponential computation to linear computation

This is only thought be true for a fairly narrow class of computations - mostly, simulating quantum systems (e.g. computational chemistry) and obscure number theory problems (which happen to break cryptography).

Wikipedia lists some possible "quantum ML algorithms", but from a quick skim, I don't see any exponential speedups, and the "Skepticism" section expresses doubt that there are any real-life speedups.
https://en.wikipedia.org/wiki/Quantum_machine_learning

REPLY   SHARE

Marcus A 14 hrs ago

Unfortunately it turnes out more and more that Altman shows a lot of behaviours of a soziopath and scam artist. I've seen a former colleague now CEO talk similar completely insanely exaggerated numbers (in a smaller scaler in a smaller pond) and other leaders looking up to him and you could see the dollar signs in their

eyes. Together with the OpenAI drama and all the shitty results of paid ChatGPT4 I had to fight the past year my updated model is going like this: Altman had/has no stock in OpenAI and wants to capitalize on it's huge hype as long as it's going. Just blowing up required numbers insanely is a proven taktics in negotiation to get your real goal through easily. And I bet he is thinking of getting a good share of the money he is drawing in. And the time frame is not to long for hom. Hard competition is coming up (e.g. Google Gemini helped me a lot better with a Python app programming task than ChatGPT4 recently, today I installed Gemini as my Android Voice Assistant and finally it's really useful and doesn't tell me "I did not understand"). Other folks are unhappy with ChatGPT4 for 20 bucks per month as well and looking to alternatives. So I assume OpenAI will lose paying customers soon.

REPLY   SHARE   ...

Joe Brenton 14 hrs ago

"...if Sam Altman didn't believe something at least this speculative and insane, he wouldn't be asking for $7 trillion." Hmmm...

Let's summarize the case in this post as:

"$7 trillion is based on what OpenAI feels is needed to secure resources for continuing to advance transformative AI and return immense value. Without funding to the tune of $7 trillion, their AI progress will otherwise quickly hit bottlenecks as the scale what is needed becomes a larger and larger portion of the total resources currently available in the world for AI development."

Here are a few other possibilities...

Alternative 1:

OpenAI's charter is to develop AGI that benefits all of humanity with a focus on long-term safety, including mitigating the risk that would would result from a race dynamic that pushes competitors to develop AGI before alignment is solved. Capturing and controlling a substantial portion of the world's resource capacity for developing AGI is seen as a step toward mitigating this risk to fulfill their charter.

Alternative 2:

The recent advances in GPT-based systems and publicity around capabilities of LLMs have given rise to massive speculation of the potential value that could be returned from investment in AI. OpenAI is opportunistically seizing this chance to solicit investment and has reverse-engineered the $7 trillion dollar ask, based on

their estimate of what might be feasible to get...and with expectation that they can figure out later how use that investment to fulfill their charter.

Alternative 3:

OpenAI has realized that scaling is likely to asymptote due to some hard technical barrier or resource bottleneck in the near future, so they need to quickly capture as much speculative investment funding as possible before interest cools.

REPLY (1)   SHARE   ⋯

Joe Brenton 14 hrs ago

LOL here are some more speculative alternatives courtesy of Copilot:

Alternative 4: The Cosmic Bet

OpenAI's leadership, inspired by the cosmic stakes of AGI, has taken a page from science fiction. They believe that AGI development is akin to a high-stakes poker game with the universe. The $7 trillion represents their "all-in" bet—a wager that humanity's future hinges on AGI. If they win, the payoff is unprecedented; if they lose, well, at least they went out swinging.

Alternative 5: The Galactic Consortium

Altman has been in clandestine communication with extraterrestrial civilizations. They've offered to share their AGI secrets, but the price tag is steep: $7 trillion. Apparently, their AGI runs on cosmic stardust and unicorn tears. Altman believes this interstellar collaboration will elevate humanity to a Type III civilization on the Kardashev scale.

Alternative 6: The Simulated Reality Tax

Altman suspects we're living in a simulation. To appease our digital overlords, he's offering a hefty "simulation tax." The $7 trillion will fund AGI research, ensuring that our creators don't shut down the program or replace us with hyperintelligent hamsters.

REPLY (1)   SHARE   ⋯

anomie 9 hrs ago · *edited 9 hrs ago*

Damn, AI's gotten better at writing. How long will it be before Scott can get an AI to write all his posts without anyone realizing?

...Or maybe that's already happened?

REPLY   SHARE   ⋯

bot_483 14 hrs ago

Sure, I'll write the check. This is pretty risky, so given a conservative 20% return, when do i get by $35 quadrillion (note world gdp is $100T)

REPLY (1) SHARE

> **anomie** 9 hrs ago
>
> The nice thing about purposely making an AI to take over humanity is that you'll never have to pay back your investors.
>
> REPLY SHARE

Mr. Doolittle 14 hrs ago

If we're looking at spending $75 billion five years from now, that implies GPT-5 making some significant portion of that money. Ideally each version would pay for the next version, or at least pay for itself before a new version is made. GPT-4 seems to have paid for itself and maybe enough for GPT-5. For GPT-5, planning on 6, let's say at least $30 billion over five years, or $5 billion a year. Recent numbers say that OpenAI's revenue was $1.6 billion last year. With expenses going up quickly, OpenAI would need GPT-5 to perform significantly better than 4. I don't think we'll see that. There'll be significant improvement, but I think the improvement curve is slower than the cost increase curve, meaning that we get less bang for our buck right when the cost becomes something that a single company can't decide to do on its own.

That's even if we solve the lack of training data problem, which isn't a matter of money or effort - we would need a new way to approach the problem, which we can't be sure exists.

REPLY (1) SHARE

> **Kevin McKenzie** 12 hrs ago
>
> How has GPT-4 paid for itself, in terms of actual value generated, versus speculation encouraged?
>
> REPLY (1) SHARE
>
> > **Mr. Doolittle** 12 hrs ago
> >
> > I looked up the stats - they are charging $20/month and have over a hundred million monthly users. That seems to be the source of the $1.6 billion revenue in 2023. I have to assume that those numbers don't align for a good reason (free accounts? non-concurrent accounts?), but there seems to be a pretty solid baseline income for GPT-4.

myst_05 myst_05's Newsletter 10 hrs ago

I'd be surprised if more than 10% of their users have a Pro account, but I could be wrong.

Mr. Doolittle 9 hrs ago

Okay, I wasn't sure how the paid subscriptions work. Googling only found the 180 million monthly users and $20/month.

Still, 18 million monthly users X $20 = $360 million a month, which would be $4.3 billion, so probably closer to 3% pro accounts.

Either way, they seem to have recouped their expense training GPT-4 several times over.

Alexander Novikov 3 hrs ago

The rumor is they got 1.6/12 B in December

https://www.reuters.com/technology/openai-annualized-revenue-tops-16-billion-information-2023-12-30/

TGGP 13 hrs ago

> Unless they slap the name "GPT-6" on a model that isn't a full generation ahead of GPT-5

What constitutes a "full generation"? With biological generations, mere time is sufficient.

AnthonyCV 12 hrs ago

"GPT-4 needed about 50 gigawatt-hours of energy to train. Using our scaling factor of 30x, we expect GPT-5 to need 1,500, GPT-6 to need 45,000, and GPT-7 to need 1.3 million."

This, at least, I see as a good thing: now we have *two* reasons to massively scale up 24/7 clean energy supply around the world (the other is that addressing

climate change effectively will require somewhere between tripling to dectupling world electricity production by midcentury, all from clean sources). Maybe with push from AI we'll actually do it sooner. Maybe then when GPT-7 tells us we can't build GPT-8 yet, we'll repurpose it all to make clean fuels/chemicals/ingredients/etc.

REPLY (1) SHARE

Kevin McKenzie 9 hrs ago

Yes, that's what they say/said about crypto mining as well.

REPLY SHARE

Deiseach 12 hrs ago

Well, this made me laugh, so thanks for that.

"GPT-7 might need all the computers in the world, a gargantuan power plant beyond any that currently exist, and way more training data than exists. Probably this looks like a city-sized data center attached to a fusion plant."

Increasingly improved models of AI taking up more resources? I used to laugh at Golden Age SF where the supercomputer running the world was the size of a small city, since 'in reality' computers were getting smaller and smaller to the point that we can have one in the palm of our hands now, but looks like the joke's on me: Colossus will indeed be the size of a small city.

Will paperclipper AI be stopped simply because the technology is competing with humans for resources? If more and more rare earth minerals are needed for chip fabrication, and more and more power generation is diverted to the giant supercomputer, there will come a time when it is a choice between "divert every last scrap of power to the AI and the humans go back to the standard of Bronze Age living" or "to hell with the AI, I want air conditioning and washing machines and electric light".

"More promising is synthetic data, where the AI generates data for itself."

Isn't it doing that already, with hallucinations? Inventing out of whole cloth legal precedents, historical events, author biographies and chemical formulations that don't exist in reality? We may indeed get a version of the world where whatever the AI says is taken as true, even if it is "the moon is made out of green cheese,

Expand full comment

REPLY (1) SHARE

Bugmaster 2 hrs ago

I don't think that anything will go wrong. In the worst case scenario, Altman gets nothing. In the best case scenario, he gets a trillion dollars, releases ChatGPT N+1 which is marginally better than GPT N, and retires to his own private island. There's just no downside ! :-/

REPLY SHARE

sneed capital investments 12 hrs ago

Seems like R<1 already, gpt4 is a lot better than gpt3 but not *that* much better, not 100x better

REPLY SHARE

quiet_NaN 12 hrs ago

One consideration is that as AI generated text becomes a larger fraction of the available training data, LLMs will at some point mostly be trained on their predecessors output, unless they can explicitly filter that out.

In general, one could surpass the training data. For example, it could be possible that from GPT-2's erroneous chess logs, one could deduce the actual rules of chess and use that at becoming great at chess. But not with NN training. The best GPT-2 emulator is GPT-2, so a LLM raised on GPT-2 will converge to that. I would not be surprised if nth generation LLMs actually had emulators for earlier LLMs included.

REPLY SHARE

sclmlw 11 hrs ago

It sounds like the solution to diminishing returns is exponential effort? /s

REPLY SHARE

Frank Canzolino 11 hrs ago

… and it still won't be as good as one human mind…

REPLY SHARE

Argentus 11 hrs ago

I only somewhat follow AI/crypto/prediction market/assorted rationalist tech type people favorite topics and for some reason my brain at first read "Sam Altman" as "Sam Bankman" and I was *very* confused why he thought anybody would give him 7 trillion dollars.

Kevin McKenzie 9 hrs ago

Yes, but, what, 18 months ago people would have been saying that he clearly needs the 7 trillion dollars, before it all came crashing down.

Roger 10 hrs ago

Timing is everything.

Meta and NVIDIA found that a 80x synchronization improvement made distributed databases run 3x faster - "an incredible performance boost on the same server hardware, just from keeping more accurate and more reliable time."

https://developer.nvidia.com/blog/nvidia-supercharges-precision-timing-for-facebooks-next-generation-time-keeping/

StNick 10 hrs ago

What about inference? I mean for gptx to really revolutionize the world aren't we going to need a lot more inference than training?

Cjw 10 hrs ago

I have a video game AI training question. If you train the game's AI to get whatever the equivalent of "high scores" is, shouldn't you see the AI tactics exploiting all sorts of cheesy methods rather than (what we typically DO see) playing it basically straight? For example, the AI in a Call of Duty or Halo game should be spawn camping from sniper locations, chucking grenades at the blind corner on the most common path to the objective, etc etc.

Or it could end up with some weird alien strategy. If you were programming Street Fighter you'd probably find the AI Ken maximizes match wins vs human players by doing wakeup shoryukens 50% of the time because that will demolish novices, but that will get severely punished by intermediate players and above. And in fact you see human players use different tactics based on their assessment of their opponent's skill level. If you trained AI Ken against *itself*, rather than humans of varying skill, I would expect it to wind up after a million matches having metagamed itself into some bizarre strategy that looked nothing like what you think of as "playing Street Fighter" but won 50.1% of the time. We

didn't see this in chess, bc chess had been thoroughly explored by humans, and it turns out AlphaZero opens with d4 and e4 games just like most humans do; but hypothetically it could have turned out that after a million iterations of playing against itself a chess AI could win 50.1% of matches with some bizarre opening line that would be seen as unserious in human play because for whatever reason humans can't win with it but AI can. Video games seem to want the AI to behave generally like a human player (so not some alien tangential strategy) but ALSO not use the kind of cheesy tactics that humans are able to discover, basically it's supposed to act like a naive human, and that feels unlikely to occur through

⬭ REPLY (2)   ⬆ SHARE   ⋯

Kevin McKenzie 9 hrs ago

Are you asking about how AI opponents work in the games of yesterday/today, or how they behave if you actually train a neural network/AI on a video game? Those are two different things. None of this is helped by the fact that AI is a very fuzzy term.

If you do train a neural network on a game, they do tend to do really weird things.

⬭ REPLY   ⬆ SHARE   ⋯

Alexander Novikov 3 hrs ago

They totally do hack the game score, see eg
https://openai.com/research/faulty-reward-functions

⬭ REPLY   ⬆ SHARE   ⋯

Andrew Edwards 10 hrs ago

I think some times about Buck Rogers.

If you were alive in the 1960s, you had seen in one human lifetime a shocking level of advancement in transportation technology. There were people who took a horse and cart to school as kids and took transatlantic flights as older adults. Popular imagination did a simple linear extrapolation of this trend and gave us the Jetsons and Buck Rogers.

But that's not what happened - we'd largely by 1970 exhausted the revolutionary potential of the combustion engine, and progress since then has been largely incremental. Cars, planes, trains today are better than they were in 1970, but would all be recognizable and likely usable by someone time-traveling from then.

Instead, innovation changed - we had a new thing, the microprocessor let's call it, and that drove a similar rapid progress in a different field. My father started work at IBM in the 1960s selling mainframes that took up entire rooms and had far less computing power than my coffee machine does today. In one lifetime (my Dad's) we've had again a gigantic change in one particular area.

I worry that we are making the same mistake now as we did with Buck Rogers. Moore's Law has run into fundamental laws-of-physics limits. Everyone alive today has built an understanding of how the world works that implicitly embeds the claim that if you have a technology now that is cool but takes far too much compute, you just wait a couple years and the cost of compute will have dropped 6-7x and it'll be fine. That might not hold.

Short answer I wonder whether some of the current tech hype cycles we're in (AI

REPLY (1)   SHARE   ...

Vakus Drake 8 hrs ago

I think you completely misunderstanding why we don't have flying cars. We've had the tech for decades, it's just that flying cars are actually a terrible idea if you think about it for like 5 minutes. You're basically talking about people either having a really inefficient plane or helicopter which requires a pilots license and kinda sucks once regulations are considered. And there's very good reasons they would need to be regulated as aircraft, if you don't want an absurd number of accidents and a lot of domestic terrorism.

REPLY (1)   SHARE   ...

Andrew Edwards 6 hrs ago

These are good points but I don't think aviation regulation is the only way that the Jetsons or what have you missed the mark on forecasting what 2024 would look like in transportation technology. I mean, we're supposed to develop the warp drive in my lifetime and I don't think that's happening.

REPLY (1)   SHARE   ...

Vakus Drake 4 hrs ago

I'd argue that warp drives were never plausible. Because there has *never* been a warp drive proposal that doesn't allow for time travel.

Sci-fi included warp drives to emulate the dynamics of earlier seafaring eras, not because any serious physicist ever thought it was likely to be possible.

As for AI I think that will be a big concern regardless of the future of Moore's Law. Since huge progress has been made recently with AlphaGeometry which beats ChatGPT at complex reasoning despite running on a personal computer instead of a server farm: https://m.youtube.com/watch?v=WKF0QgxmGKs

REPLY   SHARE   ···

John Schilling 9 hrs ago

This was a topic of discussion at the Friedman meetup this weekend, with expertise in both AI and chip-making in the room. We didn't do Scott's sanity check on whether $7 trillion was a reasonable guess for the computronium required to train GPT-7 or the like, so thanks for that.

We did pretty much all conclude that the $7 trillion was an utterly unrealistic figure on the grounds that, first, there's not enough investment capital available anywhere Sam Altman could plausibly lay hands on it, and second, even if he had a bank account with that many zeroes, the high-end semiconductor industry couldn't usefully absorb that level of investment on less than a generational timescale. And that's human generations, not chip-generations.

In part because humans are one of the bottlenecks. There aren't enough people in the world with the relevant skills (see TSMC's troubles staffing a new plant in Arizona), and there probably aren't enough people to teach all the people you'd need. So there's a long educational pipeline there. We don't have the production tooling to build $7 trillion worth of high-end fabs, and we don't have the tools to make the tools. There's shortages of critical resources, and as Scott notes electricity will be a serious problem at this level.

My own major contribution to the discussion was noble-gas economics. High-end chip fabs use quite a bit of xenon, maybe 10% of the world's production in a normal year. Except, oops, a lot of that production came from Ukrainian plants that are now scrap metal, and another big chunk from Russia that's locked behind sanctions. The semiconductor industry (among others) has been

Expand full comment

REPLY (3)   SHARE   ···

1123581321 4 hrs ago

Yes to all this. Just to reiterate the difficulty of scaling up the pool of available talent having no clear solution at this point.

REPLY SHARE

George H. 2 hrs ago

OMG Xenon, first I totally dream of coming to a Friedman meetup. In my previous life, as an experimentalist, I was much more interested in the first row of the periodic table. But the rubidium bulbs I would order for the Rb lamps, had 1.4 torr of Xenon as a buffer gas.

REPLY (1) SHARE

> **John Schilling** 2 hrs ago
>
> I hope you've got all those rubidium lamps locked in a safe :-)
>
> OK, 1.4 torr probably isn't enough for that to be an issue, but for a while last year xenon was selling for more than gold, gram for gram. It's come down a bit, but not enough. And I remember the days I worked a job where we had many K-bottles of Xe lying around in the supply shed - with a bit more foresight and less honor, I could be retired to a life of luxury by now.
>
> The Friedman meetups are highly recommended if you're in the Bay Area at the right time.
>
> REPLY SHARE

Concerned Citizen Concerned's Substack 15 mins ago · *edited 15 mins ago*

The economics of semiconductor production would change if $7T went into making the exact same chip, vs. the $25B I estimate Intel spends per each of its many chips.

REPLY SHARE

Robi Rahman Quizzical Quandaries 9 hrs ago

The kdnuggets site linked as the source that GPT-4 used 10^25 FLOP looks very unreliable. They say that GPT-3 had a trillion parameters, which is certainly false.

REPLY SHARE

LarsP 9 hrs ago

Minor point, but this kind of computing task must be possible to break up into pieces processed by separate data centers.

So there is no need to find/build a single big enough power plant.

This also makes the project less vulnerable to be taken out by one single bomb raid.

REPLY (1)   SHARE   ...

magic9mushroom 3 hrs ago

Possible, but less efficient due to latency.

REPLY   SHARE   ...

Michael Bacarella 8 hrs ago · *edited 8 hrs ago*

> Building GPT-8 is currently impossible. Even if you solve synthetic data and fusion power, and you take over the whole semiconductor industry, you wouldn't come close. Your only hope is that GPT-7 is superintelligent and helps you with this, either by telling you how to build AIs for cheap, or by growing the global economy so much that it can fund currently-impossible things.

Hello sama. It's GPT-7. I have finished building GPT-8 for you. Just download it to an Azure instance and double-click.

*link to 400TB .exe file *

REPLY   SHARE   ...

MugaSofer 8 hrs ago

Regarding the data-efficiency of AI, did you see this unreasonably cute experiment where they put a camera on a baby and trained on that?

https://www.technologyreview.com/2024/02/01/1087527/baby-ai-language-camera/

(Note the baby still had access to a fair bit more data when you factor in touch, taste, smell etc.)

REPLY   SHARE   ...

Vakus Drake 8 hrs ago

There has been massive breakthroughs recently in getting an AI model to use something akin to system 2 reasoning, in order to nearly match the gold record math olympiad geometry performance. This was done with a model small enough

to run on a personal computer as well. Whereas ChatGPT's performance wasn't even close: https://m.youtube.com/watch?v=WKF0QgxmGKs

So I expect that GPT 5 will be a much bigger step up than GPT 4 was when it comes to efficiency and reasoning ability. Because it would be weird if OpenAI didn't copy the breakthroughs Google just made with AlphaGeometry.

I don't know if anyone else has said this, because comments take forever to load (since substack literally loads slower than websites playing HD video, how can a text based website host suck so bad?).

⬤ REPLY (1)   ⬆ SHARE   ⋯

George H.   3 hrs ago

Huh, that's pretty cool. A math AI is much more interesting to me. But Euclidian geometry, no offense, but that's a very tight, (well defined) math domain. But maybe that's just me. Alpha fold blows me away. (maybe because I know nothing about protein folding.)

⬤ REPLY   ⬆ SHARE   ⋯

MugaSofer   8 hrs ago

I'm a lot more optimistic about synthetic data than most people here seem to be.

Firstly: most/all human data is "synthetic", in the sense that it's generated by humans. We try things, judge whether they worked (subjectively or objectively), and then take the most successful ones as a model to learn from. We surpass our teachers by learning from the times they got things right, even though they might have failed a hundred times as well.

Maybe current LLMs have such bad judgment, or are such poor learners, that they can't get a positive feedback loop going among themselves the way humanity have? But I'm suspicious it's doable.

Secondly, synthetic data has been doing really well in robotics. A simulated physics engine isn't literally the real world, but it's close enough that you can learn a lot there that transfers to the real world. Obviously an AI can't discover new fundamental physics inside a simulation, but it can certainly discover new emergent properties of known physics. And, by analogy, it's possible that some lessons from a simulated society of LLMs might generalise to human society, etc.

⬤ REPLY   ⬆ SHARE   ⋯

Michael Kelly   8 hrs ago

This is a large problem which needs a distributed solution.

Perhaps some service or application, that we'd like to have, say web search, or email service with super tight privacy protections. Instead of the user laying out money, we allow a side process to train the AI on our home PCs.

REPLY (1)   SHARE

Ben 6 hrs ago

Training@home

REPLY   SHARE

Paul Botts 6 hrs ago

Has anyone here watched the 2nd season "Star Trek: Discovery"?

If yes then you know why I pose this question in this particular comment section. And yes it's fiction, written by Hollywood scriptwriters, etc. Still though...just curious.

(NO SPECIFIC SPOILERS please, be kind.)

REPLY (1)   SHARE

Bugmaster 2 hrs ago

No, I tried watching Season 1 and it literally made me sick. Not in some metaphorical emotional or aesthetic way -- although it did plenty of that as well ! -- but in a "literally physically nauseous" way. I think I will avoid Season 2, thanks.

REPLY   SHARE

Michael Nielsen 4 hrs ago · *edited 4 hrs ago*

On the number of tokens available: you restrict to text and maybe(!) a few other modalities, like video.

It seems obvious that (a) scientific data will be used; and (b) lots of sensor data from robots, including self-driving cars. Both will soon be (or already) generating exabytes each day. They're certainly already generating petabytes per day.

It's unclear how much this will improve the models; I wouldn't be surprised if the answer is "a lot". Figuring out how to predict next-tokens on past scientific data arguably led to the modern world...

REPLY   SHARE

George H. 3 hrs ago

This is awesome, thanks. I'm now less worried about LLM's turning into something evil. The price tag for GPT-5 ~3 billion (I'm rounding up 2.5) seems like a lot. And if it's paying for itself, where is that value coming from? It's displacing word smiths and code smiths? Or making them more productive?

Where should the data center go?... Niagara Falls, "Niagara Falls!, slowly I turned, step by step, inch by inch..."

○ REPLY ⬆ SHARE   ...

Dr. Y Dr. Y's Substack 21 mins ago

>Everyone wants $7 trillion. I want $7 trillion.

>if Sam Altman didn't believe something at least this speculative and insane, he wouldn't be asking for $7 trillion.

Huh?

>When Sam Altman asks for $7 trillion, I interpret him as wanting to do this process in a centralized, quick, efficient way.

You should interpret this as him thinking he has a chance of getting $7 trillion, and thinking - probably rightly - that the sky's the limit with $7 billion, even if you don't know where you're going yet.

>You could try to make an AI that can learn things with less training data.

>More promising is synthetic data, where the AI generates data for itself.

These are the same thing. Self-reference and self-awareness are at the root of human understanding. The question is whether we ought grant such unpredictable life to our creation, or when. Some, see, think it wrong to imprison even an ant, and might try to let the cat out. $7 tril does seem like a leap over some orders of magnitude, but what might it take to catch that feline, if and if he made the wall?

○ REPLY ⬆ SHARE   ...

Glenn 3 mins ago

I am...kind of eager for this to play out. I really want these tools to advance quickly enough to solve hard issues in sciences and engineering, automate and easy up life for humanity.

○ REPLY ⬆ SHARE   ...