

In short, I believe the AI bubble is deeply unstable, built on vibes and blind faith, and when I say "the AI bubble," I mean the entirety of the AI trade.

And it's alarmingly simple, too.

But this isn't going to be saccharine, or whiny, or simply *worrisome*. I think at this point it's become a little ridiculous to not see that we're in a bubble. We're in a god damn bubble, it is so obvious we're in a bubble, it's *been* so obvious we're in a bubble, a bubble that seems strong but is actually very weak, with a central point of failure.

I may not be a contrarian, but I am a *hater*. I hate the waste, the loss, the destruction, the theft, the damage to our planet and the sheer *excitement* that some executives and writers have that workers may be replaced by AI — **and the bald-faced fucking lie that it's happening, and that generative AI is capable of doing so.**

And so I present to you — the Hater's Guide to the AI bubble, a comprehensive rundown of arguments I have against the current AI boom's existence. Send it to your friends, your loved ones, or print it out and eat it.

No, this isn't gonna be a traditional guide, but something you can look at and say "oh that's why the AI bubble is so bad." And at this point, I know I'm tired of being gaslit by guys in gingham shirts who desperately want to curry favour with other guys in gingham shirts but who also have PhDs. I'm tired of reading people talk about how we're "in the era of agents" that don't *fucking work* and *will never fucking work*. I'm tired of hearing about "powerful AI" that is actually crap, and I'm tired of being told the future is here while having the world's least-useful most-expensive cloud software shoved down my throat.

Look, the generative AI boom is a mirage, it hasn't got the revenue or the returns or the product efficacy for it to matter, everything you're seeing is ridiculous and wasteful, and when it all goes tits up I want you to remember that I wrote this and tried to say something.

## The Magnificent 7's Weakpoint: NVIDIA

As I write this, NVIDIA is currently sitting at \$170 a share — a dramatic reversal of fate after the pummelling it took from the DeepSeek situation in January, which sent it tumbling to a brief late-April trip below \$100 before things turned around.

The Magnificent 7 stocks — NVIDIA, Microsoft, Alphabet (Google), Apple, Meta, Tesla and Amazon — make up around 35% of the value of the US stock market, and of that, NVIDIA's market value makes up about 19% of the Magnificent 7. This dominance is also why ordinary people ought to be deeply concerned about the AI bubble. The Magnificent 7 is almost certainly a big part of their retirement plans, even if they're not directly invested.

Back in May, Yahoo Finance's Laura Bratton reported that Microsoft (18.9%), Amazon (7.5%), Meta (9.3%), Alphabet (5.6%), and Tesla (0.9%) **alone** make up **42.4% of NVIDIA's revenue**. The breakdown makes things worse. Meta spends 25% — and Microsoft an alarming 47% — of its capital expenditures on NVIDIA chips, and as Bratton notes, Microsoft also spends money renting servers from CoreWeave, which analyst Gil Luria of D.A.Davidson estimates accounted for \$8 billion (more than 6%) of NVIDIA's revenue in 2024. Luria also estimates that neocloud companies like CoreWeave and Crusoe — that exist only to prove AI compute services — account for as much as 10% of NVIDIA's revenue.

NVIDIA's climbing stock value comes from its continued revenue growth. In the last four quarters, NVIDIA has seen year-over-year growth of 101%, 94%, 78% and 69%, and, in the last quarter, a little statistic was carefully brushed under the rug: that NVIDIA *missed*, though narrowly, on data center revenue. This is exactly what it sounds like — GPUs that are used in servers, rather than gaming consoles and PCs (. Analysts estimated it would make \$39.4 billion from this category, and NVIDIA only (lol) brought in \$39.1 billion). Then again, it could be attributed to its problems in China, especially as the H20 ban has only just been lifted. In any case, it was a miss!

NVIDIA's quarter-over-quarter growth has also become aggressively normal — from 69%, to 59%, to 12%, to 12% again each quarter, which, again, isn't *bad* (it's pretty great!), but when 88% of your revenue is based on one particular line in your earnings, *it's a pretty big concern, at least for me*. Look, I'm not a stock analyst, nor am I pretending to be one, so I am keeping this simple:

- NVIDIA relies not only on selling **lots of GPUs each quarter**, but **it must always, always sell more GPUs the next quarter**.
- 42% of NVIDIA's revenue comes from Microsoft, Amazon, Meta, Alphabet and Tesla continuing to buy **more GPUs**.
- NVIDIA's value and continued growth is heavily reliant on **hyperscaler purchases** and **continued interest in generative AI**.
- The US stock market's continued health relies, on some level, on five or six companies (it's unclear how much Apple buys GPU-wise) spending billions of dollars on GPUs from NVIDIA.
  - An analysis from portfolio manager Danke Wang from January found that the Magnificent 7 stocks accounted for 47.87% of the Russell 1000 Index's returns in 2024 (an index fund of the 1000 highest-ranked stocks on FTSE Russell's index).

In simpler terms, **35% of the US stock market is held up by five or six companies buying GPUs**. If NVIDIA's growth story stumbles, it will reverberate through the rest of the Magnificent 7, making them rely on their own AI trade stories.

And, as you will shortly find out, ***there is no AI trade, because generative AI is not making anybody any money.***

## The Hollow "AI Trade"

I'm so tired of people telling me that companies are "making tons of money on AI." Nobody is making a profit on generative AI other than NVIDIA. No, really, I'm serious.

## The Magnificent 7's AI Story Is Flawed, With \$560 Billion of Capex between 2024 and 2025 Leading to \$35 billion of Revenue, And No Profit

If they keep their promises, by the end of 2025, Meta, Amazon, Microsoft, Google and Tesla will have spent over \$560 billion in capital expenditures on AI in the last two years, all to make around \$35 billion.

This is egregiously fucking stupid.

**Microsoft AI Revenue In 2025: \$13 billion, with \$10 billion from OpenAI, sold "at a heavily discounted rate that essentially only covers costs for operating the servers."**

**Capital Expenditures in 2025: \$80 billion**

As of January 2025, Microsoft's "annualized" — meaning [best month]x12 — revenue from artificial intelligence was around \$13 billion, a number that it chose not to update in its last earnings, likely because it's either flat or not growing, though it could in its upcoming late-July earnings. Yet the problem with this revenue is that \$10 billion of that revenue, according to The Information, comes from OpenAI's spend on Microsoft's Azure cloud, and Microsoft offers preferential pricing — "a heavily discounted rental rate that essentially only covers Microsoft's costs for operating the servers" according to The Information.

In simpler terms, 76.9% of Microsoft's AI revenue comes from OpenAI, and is sold at just above or at cost, making Microsoft's "real" AI revenue about \$3 billion, or around 3.75% of this year's capital expenditures, or 16.25% if you count OpenAI's revenue, which costs Microsoft more money than it earns.

The Information reports that Microsoft made \$4.7 billion in "AI revenue" in 2024, of which OpenAI accounted for \$2 billion, meaning that for the \$135.7 billion that Microsoft has spent in the last two years on AI infrastructure, it has made \$17.7 billion, of which OpenAI accounted for \$12.7 billion.

**Amazon AI Revenue In 2025: \$5 billion**

**Capital Expenditures in 2025: \$105 billion**

Things do not improve elsewhere. An analyst estimates that Amazon, which plans to spend \$105 billion in capital expenditures this year, will make \$5 billion on AI in 2025, rising, and I quote, "as much as 80%," suggesting that Amazon may have made a measly

\$2.77 billion in 2024 on AI in a year when it spent \$83 billion in capital expenditures.

[editor's note: this piece originally said "\$1 billion" instead of "\$2.77 billion" due to a math error, sorry!]

Last year, Amazon CEO Andy Jassy said that "AI represents for sure the biggest opportunity since cloud and probably the biggest technology shift and opportunity in business since the internet." I think he's full of shit.

## Google AI Revenue: \$7.7 Billion (at most)

## Capital Expenditures in 2025: \$75 Billion

Bank of America analyst Justin Post estimated a few weeks ago that Google's AI revenue would be in the region of \$7.7 billion, though his math is, if I'm honest, a little generous:

*Google's artificial intelligence model is set to drive \$4.2 billion in subscription revenue within its Google Cloud segment in 2025, according to an analysis from Bank of America last week.*

*That includes \$3.1 billion in revenue from subscribers to Google's AI plans with its Google One service, Bank of America's Justin Post estimates.*

*Post also expects that the integration of Google's Gemini AI features within its Workspace service will drive \$1.1 billion of the \$7.7 billion in revenue he projects for that segment in 2025.*

Google's "One" subscription includes increased cloud storage across Google Drive, Gmail and Google Photos, and added a \$20-a-month "premium" plan in February 2024 that included access to Google's various AI models. Google has claimed that the "premium AI tier accounts for millions" of the 150 million subscribers to the service, though how many millions is impossible to estimate — but that won't stop me trying!

Assuming that \$3.1 billion in 2025 revenue would work out to \$258 million a month, that would mean there were 12.9 million Google One subscribers also paying for the premium AI tier. This isn't out of the realm of possibility — after all, OpenAI has 15.5 million paying

subscribers — but Post is making a *generous* assumption here. Nevertheless, we'll accept the numbers as they are.

And the numbers fuckin' stink! Google's \$1.1 billion in workspace service revenue came from a forced price-hike on those who use Google services to run their businesses, meaning that this is likely not a number that can significantly increase without punishing them further.

\$7.7 billion of revenue — not profit! — on \$75 billion of capital expenditures. Nasty!

## Meta AI Revenue: \$2bn to \$3bn

### Capital Expenditures In 2025: \$72 Billion

Someone's gonna get mad at me for saying this, but I believe that Meta is simply *burning cash* on generative AI. There is no product that Meta sells that monetizes Large Language Models, but *every* Meta product now has them shoved into them, such as your Instagram DMs oinking at you to generate artwork based on your conversation.

Nevertheless, we do have *some sort of knowledge of what Meta is saying* due to the copyright infringement case Kadrey v. Meta. Unsealed judgment briefs revealed in April that Meta is claiming that "GenAI-driven revenue will be more than \$2 billion," with estimates as high as \$3 billion. The same document also claims that Meta expects to make **\$460 billion to \$1.4 trillion in total revenue through 2035**, the kind of thing that should get you fired in an iron ball into the sun.

Meta makes 99% of its revenue from advertising, and the unsealed documents state that it "[generates] revenue from [its] Llama models and will continue earning revenue from each iteration," and "share a percentage of the revenue that [it generates] from users of the Llama models...hosted by those companies," with the companies in question redacted. Max Zeff of TechCrunch adds that Meta lists host partners like AWS, NVIDIA, Databricks, Groq, Dell, Microsoft Azure, Google Cloud, and Snowflake, so it's *possible* that Meta makes money from licensing to those companies. Sadly, the exhibits further discussing these numbers are filed under seal.



Either way, we are now at \$332 billion of capital expenditures in 2025 for \$28.7 billion of revenue, of which \$10 billion is OpenAI's "at-cost or just above cost" revenue. Not great.

## Tesla Does Not Appear To Make Money From Generative AI

### Capital Expenditures In 2025: \$11 billion

Despite its prominence in the magnificent 7, Tesla is one of the least-exposed of the magnificent 7 to the AI trade, as Elon Musk has turned it into a meme stock company. That doesn't mean, of course, that Musk isn't touching AI. xAI, the company that develops racist Large Language Model "Grok" and owns what remains of Twitter, apparently burns \$1 billion a month, and The Information reports that it makes a whopping \$100 million in annualized revenue — so, about \$8.33 million a month. There is a shareholder vote for Tesla to potentially invest in xAI, which will probably happen, allowing Musk to continue to pull leverage from his Tesla stock until the company's decaying sales and brand eventually swallow him whole.

But we're not talking about Elon Musk today.

## Apple's AI Story Is Weird

### Capital Expenditures In 2025: around \$11 billion

Apple Intelligence radicalized millions of people against AI, mostly because it *fucking stank*. Apple clearly got into AI reluctantly, and now faces stories about how they "fell behind in the AI race," which mostly means that Apple aggressively introduced people to the features of generative AI by force, and it turns out that people don't really want to summarize documents, write emails, or make "custom emoji," and anyone who thinks they would **is a fucking alien**.

In any case, Apple hasn't bet the farm on AI, insomuch as it hasn't spent two hundred billion dollars on infrastructure for a product with a limited market that only loses money.

# The Fragile Five — Amazon, Google, Microsoft, Meta and Tesla — Are Holding Up The US Stock Market By Funding NVIDIA's Future Growth Story

To be clear, I am not saying that any of the Magnificent 7 are going to die — just that five companies' spend on NVIDIA GPUs largely dictate how stable the US stock market will be. If any of these companies (but especially NVIDIA) sneeze, your 401k or your kid's college fund will catch a cold.

I realize this sounds a little simplistic, but by my calculations, NVIDIA's value underpins around 8% of the value of the US stock market. At the time of writing, it accounts for roughly 7.5% of the S&P 500 — an index of the 500 largest US publicly-traded companies. A disturbing 88% of Nvidia's revenue comes from enterprise-scale GPUs primarily used for generative AI, of which five companies' spend makes up 42% of its revenue. **In the event that any one of these companies makes significant changes to their investments in NVIDIA chips, it will eventually have a direct and meaningful negative impact on the wider market.**

NVIDIA's earnings are, effectively, the US stock market's confidence, and everything rides on five companies — and if we're honest, really *four* companies — buying GPUs for generative AI services or to train generative AI models. Worse still, these services, while losing these companies massive amounts of money, ***don't produce much revenue, meaning that the AI trade is not driving any real, meaningful revenue growth.***

## ***But Ed, They Said Something About Points of Growth-***

Silence!

Any of these companies talking about "growth from AI" or "the jobs that AI will replace" or "how AI has changed their organization" are hand-waving to avoid telling you how much money these services are actually making them. If they were making good money and experiencing real growth as a result of these services, they wouldn't shut the fuck up about it! They'd be in your ear and up your ass hooting about how much cash they were rolling in!



And they're not, because they aren't rolling in cash, and are in fact blowing nearly a hundred billion dollars each to build massive, power-hungry, costly data centers for no real reason.

Don't watch the mouth — watch the hands. These companies **are going to say they're seeing growth from AI, but unless they actually show you the growth and enumerate it, they are hand-waving.**

## **Ed! Amazon Web Services Took Years To Become Profitable! People Said Amazon Would Fail!**

So, one of the most annoying and consistent responses to my work is to say that either Amazon or Amazon Web Services "ran at a loss," and that Amazon Web Services — the invention of modern cloud computing infrastructure — "lost money and then didn't."

The thing is, this statement is one of those things that people say because it *sounds rational*. Amazon *did* lose money, and Amazon Web Services *was* expensive, that's obvious, right?

The thing is, I've never really had anyone explain this point to me, so I am finally going to sit down and deal with this criticism, because every single person who mentions it thinks they just pulled Excalibur from the stone and can now decapitate me. They claim that because people in the past doubted Amazon — because, or in addition to the burn rate of Amazon Web Services as the company built out its infrastructure — that I too am wrong, because they were wrong about that.

This isn't Camelot, you rube! You are not King Arthur!

I will address both the argument itself and the "they" part of it too — because if the argument is that the people that got AWS wrong should not be trusted, then we should no longer trust them, the people actively propagandizing our supposed generative AI future.

***Right?***

So, I'm honestly not *sure* where this argument came from, because there is, to my knowledge, no story about Amazon Web Services where somebody suggested its burnrate would kill Amazon.

But let's go back in time to the May 31 1999 piece that some might be thinking of, called "Amazon.bomb," and how writer Jacqueline Doherty was mocked soundly for "being wrong" about Amazon, which has now become quite profitable.

I also want to be clear **that Amazon Web Services didn't launch until 2006, and Amazon itself would become reliably profitable in 2003**. Technically Amazon had opened up Amazon.com's web services for developers to incorporate its content into their applications in 2002, but what we consider AWS today — cloud storage and compute — launched in 2006.

But okay, what did she actually say?

*Unfortunately for Bezos, Amazon is now entering a stage in which investors will be less willing to rely on his charisma and more demanding of answers to tough questions like, when will this company actually turn a profit? And how will Amazon triumph over a slew of new competitors who have deep pockets and new technologies?*

*We tried to ask Bezos, but he declined to make himself or any other executives of the company available. He can ignore Barron's, but he can't ignore the questions.*

*Amazon last year posted a loss of \$125 million [\$242.6m in today's money] on revenues of \$610 million [\$1.183 billion in today's money]. And in this year's first quarter it got even worse, as the company posted a loss of \$61.7 million [\$119.75 million in today's money] on revenues of \$293.6 million [\$569.82 million in today's money].*

Her argument, for the most part, is that Amazon was burning cash, had a ton of competition from other people doing similar things, and that analysts backed her up:

*"The first mover does not always win. The importance of being first is a mantra in the Internet world, but it's wrong. The ones that are the most efficient will be successful," says one retail analyst. "In retailing, anyone can build a great-looking store. The hard part is building a great-looking store that makes money."*

Fair arguments for the time, though perhaps a little narrow-minded. The assumption wasn't that *what Amazon was building* was a bad idea, *but that Amazon wouldn't be the ones to build it*, with one saying:

*"Once Wal-Mart decides to go after Amazon, there's no contest," declares Kurt Barnard, president of Barnard's Retail Trend Report. "Wal-Mart has resources Amazon can't even dream about."*

In simpler terms: **Amazon's business model wasn't in question. People were buying shit online. In fact, this was just before the dot com bubble burst, and when optimism about the web was at a high point.** Yet the comparison stops there — people obviously liked buying shit online, it was the business models of many of these companies — like WebVan — that sucked.

## But Let's Talk About Amazon Web Services

Amazon Web Services was an outgrowth of Amazon's own infrastructure, which had to expand rapidly to deal with the influx of web traffic for Amazon.com, which had become one of the world's most popular websites and was becoming increasingly more-complex as it sold things other than books. Other companies had their own infrastructure, but if a smaller company wanted to scale, they'd basically need to build their own thing.

It's actually pretty cool what Amazon did! Remember, this was the early 2000s, before Facebook, Twitter, and a lot of the modern internet we know that runs on services like Amazon Web Services, Microsoft Azure and Google Cloud. It invented the modern concept of compute!

But we're here to talk about Amazon Web Services being *dangerous for Amazon* and people *hating on it*.

A November 2006 story from Bloomberg talked about Jeff Bezos' Risky Bet to "run your business with the technology behind his web site," saying that "Wall Street [wanted] him to mind the store." Bezos, referred to as a "one-time internet poster boy" that became "a post-dot-com piñata." Nevertheless, this article has what my haters crave:

*But if techies are wowed by Bezos' grand plan, it's not likely to win many converts on Wall Street. To many observers, it conjures up the ghost of Amazon past. During the dot-com boom, Bezos spent hundreds of millions of dollars to build distribution centers and computer systems in the promise that they eventually would pay off with outsize returns. That helped set the stage for the world's biggest Web retail operation, with expected sales of \$10.5 billion this year.*

...

*All that has investors restless and many analysts throwing up their hands wondering if Bezos is merely flailing around for an alternative to his retail operation. Eleven of 27 analysts who follow the company have underperform or sell ratings on the stock--a stunning vote of no confidence. That number of sell recommendations is matched among large companies only by Qwest Communications International Inc. (Q ), according to investment consultant StarMine Corp. It's more than even the eight sell opinions on struggling Ford Motor Co. (F )*

Pretty bad, right? My goose is cooked? All those analysts seem pretty mad!

Except it's not, my goose is raw! Yours, however, has been in the oven for over a year!

Emphasis mine:

*By all accounts, Amazon's new businesses bring in a minuscule amount of revenue. **Although its direct cost of providing them appears relatively low because the hardware and software are in place**, Stifel Nicolaus & Co. (SF ) analyst Scott W. Devitt notes: "There's not going to be any economic return from any of these projects for the foreseeable future." Bezos himself admits as much.*

*But with several years of heavy spending already, he's making this a priority for the long haul. "We think it's going to be a very meaningful business for us one day," he says. "What we've historically seen is that the seeds we plant can take anywhere from three, five, seven years."*

That's right — the ongoing costs aren't the problem.

Hey wait a second, that's a name! I can look up a name! Scott W. Devitt now works at Wedbush as its managing director of equity research, and has said AI companies would enter a new stage in 2025...god, just read this:

*The second stage is "the application phase of the cycle, which should benefit software companies as well as the cloud providers. And then, phase three of this will ultimately be the consumer-facing companies figuring out how to use the technology in ways that actually can drive increased interactions with consumers."*

The analyst says the market will enter phase two in 2025, with software companies and cloud provider stocks expected to see gains. He adds that cybersecurity companies could also benefit as the technology evolves.

Dewitt specifically calls out Palantir, Snowflake, and Salesforce as those who would "gain." In none of these cases am I able to see the actual revenue from AI, but Salesforce itself said that it will see no revenue growth from AI this year. Palantir also, as discovered by the Autonomy Institute's recent study, recently added to the following to its public disclosures:

*There are significant risks involved in deploying AI and there can be no assurance that using AI in our platforms and products will enhance or be beneficial to our business, including our profitability.*

What I'm saying is that *analysts can be wrong! And they can be wrong at scale!* There is no analyst consensus that agrees with me. In fact, most analysts appear to be bullish on AI, despite the significantly-worse costs and total lack of growth!

Yet even in this Hater's Parade, the unnamed journalist makes a case *for* Amazon Web Services:

*Sooner than that, those initiatives may provide a boost for Amazon's retail side. For one, they potentially make a profit center out of idle computing capacity needed for that retail operation. Like most computer networks, Amazon's uses as little as 10% of its capacity at any one time just to leave room for occasional spikes. It's the same story in the company's distribution centers. Keeping them humming at higher capacity means they operate more efficiently, besides giving customers a much broader selection of products. And the more stuff Amazon ships, both its own inventory or others', the better deals it can cut with shippers.*

## But Amazon Web Services Cost Money Ed, Now You Shall Meet Your End!

Nice try, chuckles!

In 2015, the year that Amazon Web Services became profitable, Morgan Stanley analyst Katy Huberty believed that it was running at a "material loss," suggesting that \$5.5 billion of Amazon's "technology and content expenses" was actually AWS expenses, with a "negative contribution of \$1.3 billion."

Here is Katy Huberty, the analyst in question, declaring six months ago that "2025 [will] be the year of Agentic AI, robust enterprise adoption, and broadening AI winners."

So, yes, analysts really got AWS wrong. But putting that aside, there might actually be a comparison here! Amazon Web Services *absolutely* created a capital expenditures drain on Amazon. From Forbes's Chuck Jones:

*In 2014 Amazon had \$4.9 billion in capital expenditures, up 42% from 2013's \$3.4 billion. The company has a wide range of items that it buys to support and grow its business ranging from warehouses, robots and computer systems for its core retail business and AWS. While I don't expect Amazon to detail how much goes to AWS I*



*suspect it is a decent percentage, which means AWS needs to generate appropriate returns on the capital deployed.*

In today's money, this means that Amazon spent \$6.76 billion in capital expenditures on AWS in 2014. Assuming it was this much every year — it wasn't, but I want to make an example of every person claiming that this is a gotcha — it took \$67.6 billion and ten years (though one could argue it was nine) of pure capital expenditures to turn Amazon Web Services into a business that now makes billions of dollars a quarter in profit.

That's \$15.4 billion **less than Amazon's capital expenditures for 2024, and less than one-fifteenth its projected capex spend for 2025.** And to be clear, the actual capital expenditure numbers are likely much lower, but I want to make it clear that *even when factoring in inflation*, Amazon Web Services was A) a bargain and B) a fraction of the cost of what Amazon has spent in 2024 or 2025.

**A fun aside:** On March 30 2015, Kevin Roose wrote a piece for New York Magazine about the cloud compute wars, in which he claimed that, and I quote, "there's no reason to suspect that Amazon would ever need to raise prices on AWS, or turn the fabled "profit switch" that pundits have been speculating about for years." Less than a month later Amazon revealed Amazon Web Services was profitable. They don't call him "the most right man in tech journalism" for nothing!

## **Generative AI and Large Language Models Do Not Resemble Amazon Web Services or The Greater Cloud Compute Boom, As Generative AI Is *Not Infrastructure***

Some people compare Large Language Models and their associated services to Amazon Web Services, or services like Microsoft Azure or Google Cloud, and they are wrong to do so.

Amazon Web Services, when it launched, comprised of things like (and forgive how much I'm diluting this) Amazon's Elastic Compute Cloud (EC2), where you rent space on

Amazon's servers to run applications in the cloud, or Amazon Simple Storage (S3), which is enterprise-level storage for applications. In simpler terms, if you were providing a cloud-based service, you used Amazon to both store the stuff that the service needed *and* the actual cloud-based processing (compute, so like your computer loads and runs applications but delivered to thousands or millions of people).

This is a huge industry. Amazon Web Services alone brought in revenues of over \$100 billion in 2024, and while Microsoft and Google don't break out their cloud revenues, they're similarly large parts of their revenue, and Microsoft has used Azure in the past to patch over shoddy growth.

These services are also *selling infrastructure*. You aren't just paying for the compute, but the ability to access storage and deliver services with low latency — so users have a snappy experience — wherever they are in the world. The subtle magic of the internet is that it works at all, and a large part of that is the cloud compute infrastructure and oligopoly of the main providers having such vast data centers. This is much cheaper than doing it yourself, until a certain point. Dropbox moved away from Amazon Web Services as it scaled. It also allows someone else to take care of maintenance of the hardware *and* make sure it actually gets to your customers. You also don't have to worry about spikes in usage, because these things are usage-based, and you can always add more compute to meet demand.

There is, of course, nuance — security-specific features, content-specific delivery services, database services — behind these clouds. You are buying into the infrastructure of the infrastructure provider, and the reason these products are so profitable is, in part, because you are handing off the problems and responsibility to somebody else. And based on that idea, there are multiple product categories you can build on top of it, because **ultimately cloud services are about Amazon, Microsoft and Google running your infrastructure for you**.

Large Language Models and their associated services are completely different, despite these companies attempting to prove otherwise, and it starts with a very simple problem: why did any of these companies build these giant data centers and fill them full of GPUs?

Amazon Web Services was created out of necessity — Amazon's infrastructure needs were so great that it effectively had to build both the software and hardware necessary to deliver a store that sold theoretically everything to theoretically anywhere, handling both the traffic from customers, delivering the software that runs Amazon.com quickly and reliably, and, well, making sure things ran in a stable way. It didn't need to come up with a reason for people to run web applications — they were already doing so themselves, but in ways that cost a lot, were inflexible, and required specialist skills. AWS took something that people already did, and what there was a proven demand for, and made it better. Eventually, Google and Microsoft would join the fray.

And that appears to be the only similarity with generative AI — that due to the ridiculous costs of both the data centers and GPUs necessary to provide these services, it's largely impossible for others to even enter the market.

Yet after that, generative AI feels more like a *feature* of cloud infrastructure rather than infrastructure itself. AWS and similar megaclouds are versatile, flexible and multifaceted. Generative AI does what generative AI does, and that's about it.

You can *run* lots of different things on AWS. What are the different things you can run using Large Language Models? What are the different use cases, and, indeed, user requirements that make this the supposed "next big thing"?

Perhaps the argument is that generative AI is the next AWS or similar cloud service because you can build the next great companies on the infrastructure of others — the models of, say, OpenAI and Anthropic, and the servers of Microsoft.

So, okay, let's humour this point too. You can build the next great AI startup, and you have to build it on one of the megclouds because they're the only ones that can afford to build the infrastructure.

One small problem.

## **Companies Built On Top Of Large Language Models Don't Make Much Money (In Fact, They're**

# Likely All Deeply Unprofitable)

Let's start by establishing a few facts:

- Outside of one exception — Midjourney, which claimed it was profitable in 2022, which may not still be the case, I've reached out to ask— every single Large Language Model company is unprofitable, often wildly so.
- Outside of OpenAI, Anthropic and Anysphere (which makes AI coding app Cursor), there are no Large Language Model companies — either building models or services on top of others' models — that make more than \$500 million in annualized revenue (meaning month x 12), and outside of Midjourney (\$200m ARR) and Ironclad (\$150m ARR), according to The Information's Generative AI database, and Perplexity (which just announced it's at \$150m ARR), there are **only twelve generative AI-powered companies making \$100 million annualized (or \$8.3 million a month) in revenue**. Though the database doesn't have Replit (which recently announced it hit \$100 million in annualized revenue), I've included it in my calculations for the sake of fairness.
  - Of these companies, two have been acquired — Moveworks (acquired by ServiceNow in March 2025) and Windsurf (acquired by Cognition in July 2025).
  - For the sake of simplicity, I've left out companies like Surge, Scale, Turing and Together, all of whom run consultancies selling services for training models.
- Otherwise, there are seven companies that make \$50 million or more ARR (\$4.16 million a month).

None of this is to say that one hundred million dollars isn't a lot of money to you and me, but in the world of Software-as-a-Service or enterprise software, *this is chump change*. Hubspot had revenues of \$2.63 billion in its 2024 financial year.

We're three years in, and generative AI's highest-grossing companies — outside OpenAI (\$10 billion annualized as of early June) and Anthropic (\$4 billion annualized as of July), and both lose billions a year after revenue — have three major problems:

- Businesses powered by generative AI do not seem to be popular.
- Those businesses that are remotely popular are deeply unprofitable...

- ...and even the less-popular generative AI-powered businesses are deeply unprofitable.

But let's start with Anysphere and Cursor, its AI-powered coding app, and its \$500 million of annualized revenue. Pretty great, right? It hit \$200 million in annualized revenue in March, then hit \$500 million annualized revenue in June after raising \$900 million. That's amazing!

Sadly, it's a mirage. Cursor's growth was a result of an unsustainable business model that it's now had to replace with opaque terms of service, dramatically restricted access to models, and rate limits that effectively stop its users using the product at the price point they were used to.

It's also horribly unprofitable, and a sign of things to come for generative AI.

## **Cursor's \$500 Million "Annualized Revenue" Was Earned With A Product It No Longer Offers, And Anthropic/OpenAI Just Raised Their Prices, Increasing Cursor's Costs Dramatically**

A couple of weeks ago, I wrote up the dramatic changes that Cursor made to its service in the middle of June on my premium newsletter, and discovered that they timed precisely with Anthropic (and OpenAI to a lesser extent) adding "service tiers" and "priority processing," which is tech language for "pay us extra if you have a lot of customers or face rate limits or service delays." These price shifts have also led to companies like Replit having to make significant changes to its pricing model that disfavor users.

I will now plagiarise myself:

- In or around May 5, 2025 — Cursor closes a \$500 million funding round.
- May 22 2025 — Anthropic launches Claude 4 Opus and Sonnet, and on May 30, 2025 adds Service Tiers, including priority pricing specifically focused on cache-heavy products like Cursor.

- May 30, 2025 — Reuters reports that Anthropic's "annualized revenue hit \$3 billion," with a "key driver" being "code generation." This translates to around \$250 million in monthly revenue.
- June 9 2025 — CNBC reports OpenAI has hit \$10 billion in annualized revenue. They say "annual recurring revenue," but they mean annualized.
  - The very same day, OpenAI cuts the price of its o3 model by 80%, which competes directly with Claude 4 Opus.
    - This is a direct and aggressive attempt to force Anthropic to raise prices, or try and muscle in on its terrain.
- On or around June 16 2025 — Cursor changes its pricing, adding a new \$200-a-month "Ultra" tier that, in its own words, is "made possible by multi-year partnerships with OpenAI, Anthropic, Google and xAI," which translates to "multi-year commitments to spend, which can be amortized as monthly amounts."
- A day later, Cursor dramatically changed its offering to a "usage-based" one where users got "at least" the value of their subscription — \$20-a-month provided more than \$20 of API calls — in compute, along with arbitrary rate limits and "unlimited" access to Cursor's own slow model that its users hate.
- June 18 — Replit announces its "effort-based pricing" increases.
- July 1 2025 — The Information reports Anthropic has hit "\$4 billion annual pace," meaning that it is making \$333 million a month, or an increase of \$83 million a month, **or an increase of just under 25% in the space of a month.**

In simpler terms, Cursor raised \$900 million and very likely had to hand large amounts of that money over to OpenAI and Anthropic to keep doing business with them, *and then immediately changed its terms of service to make them worse.* As I said at the time:

*While some may believe that both OpenAI and Anthropic hitting "annualized revenue" milestones is good news, you have to consider **how** these milestones were hit. Based on my reporting, I believe that both companies are effectively doing steroids, forcing massive infrastructural costs onto big customers as a means of covering the increasing costs of their own models.*

*There is simply no other way to read this situation. By making these changes, Anthropic is intentionally making it harder for its largest customer to do business,*



*creating extra revenue by making Cursor's product worse by proxy. What's sickening about this particular situation is that it doesn't really matter if Cursor's customers are happy or sad — they, like OpenAI's enterprise Priority Access API, require a long-term commitment which involves a minimum throughput of tokens for second as part of their Tiered Access program.*

*If Cursor's customers drop off, both OpenAI and Anthropic still get their cut, and if Cursor's customers somehow outspend even those commitments, they'll either still get rate limited or Anysphere will incur more costs.*

Cursor is the largest and most-successful generative AI company, and these aggressive and desperate changes to its product suggest A) that its product is deeply unprofitable and B) that its current growth was a result of offering a product that was not the one it would sell in the long term. **Cursor misled its customers, and its current revenue is, as a result, highly unlikely to stay at this level.**

Worse still, the two Anthropic engineers who left to join Cursor two weeks ago just returned to Anthropic. This heavily suggests that whatever they saw at Cursor wasn't compelling enough to make them stay.

As I also said:

*While Cursor may have raised \$900 million, it was really OpenAI, Anthropic, xAI and Google that got that money.*

*At this point, there are no profitable enterprise AI startups, and it is highly unlikely that the new pricing models by both Cursor and Replit are going to help.*

*These are now the new terms of doing business with these companies — a shakedown, where you pay up for priority access or "tiers" or face indeterminate delays or rate limits. Any startup scaling into an "enterprise" integration of generative AI which means, in this case, anything that requires a certain level of service uptime) has to commit to both a minimum amount of months and a throughput of tokens, which means that the price of starting an AI startup that gets any kind of real market traction just dramatically increased.*

*While one could say "oh perhaps you don't need priority access," the "need" here is something that will be entirely judged by Anthropic and OpenAI in an utterly opaque manner. They can — and will! — throttle companies that are too demanding on their system, as proven by the fact that they've done this to Cursor multiple times.*

## Why Does Cursor Matter? Simple: Generative AI Has No Business Model If It Can't Do Software As A Service

I realize it's likely a little boring hearing about software as a service, **but this is the only place where generative AI can really make money. Companies buying hundreds or thousands of seats are how industries that rely upon compute grow, and without that growth, they're going nowhere.**

To give you some context, Netflix makes about \$39 billion a year in subscription revenue, and Spotify about \$18 billion. These are the single-most-popular consumer software subscriptions in the world — and OpenAI's 15.5 million subscribers suggest that it can't rely on them for the kind of growth that would actually make the company worth \$300 billion (or more).

Cursor is, as it stands, the *one* example of a company thriving using generative AI, and it appears its rapid growth was a result of selling a product at a massive loss. As it stands today, Cursor's product is significantly worse, and its Reddit is full of people furious at the company for the changes.

In simpler terms, Cursor was the company that people mentioned to prove that startups could make money by building products on top of OpenAI and Anthropic's models, yet the truth is that the only way to do so and grow is to burn tons of money. While the tempting argument is to say that Cursor's "customers are addicted," this is clearly not the case, nor is it a real business model.

This story also showed that Anthropic and OpenAI are the biggest threats to their customers, and will actively rent-seek and punish their success stories, looking to loot as much as they can from them.

**To put it bluntly: Cursor's growth story was a lie. It reached \$500 million in annualized revenue selling a product it can no longer afford to sell, suggesting material weakness in its own business and any and all coding startups.**

It is also remarkable — and a shocking failure of journalism — that this isn't in every single article about Anysphere.

## No, Really, Where Are The Consumer AI Startups?

I'm serious! Perplexity? Perplexity only has \$150 million in annualized revenue! It spent 167% of its revenue in 2024 (\$57m, its revenue was \$34m) on compute services from Anthropic, OpenAI, and Amazon! It **lost \$68 million!**

And worse still, it has no path to profitability, and ***it's not even anything new! It's a search engine!*** Professional gasbag Alex Heath just did a flummoxing interview with Perplexity CEO Aravind Srivas, who, when asked how it'd become profitable, appeared to experience a stroke:

*Maybe let me give you another example. You want to put an ad on Meta, Instagram, and you want to look at ads done by similar brands, pull that, study that, or look at the AdWords pricing of a hundred different keywords and figure out how to price your thing competitively. These are tasks that could definitely save you hours and hours and maybe even give you an arbitrage over what you could do yourself, because AI is able to do a lot more. And at scale, if it helps you to make a few million bucks, does it not make sense to spend \$2,000 for that prompt? It does, right? So I think we're going to be able to monetize in many more interesting ways than chatbots for the browser.*

Aravind, do you smell toast?

And don't talk to me about "AI browsers," I'm sorry, it's not a business model. How are people going to make revenue on this, hm? What do these products actually do? Oh they can poorly automate accepting LinkedIn invites? It's like God himself has personally blessed my computer. Big deal!

In any case, it doesn't seem like you can really build a consumer AI startup that makes anything approaching a real company. Other than ChatGPT, I guess?

## The Generative AI Software As A Service Market Is Small, With Little Room For Growth And No Profits To Be Seen

Arguably the biggest sign that things are troubling in the generative AI space is that we use "annualized revenue" at all, which, as I've mentioned repeatedly, means multiplying a month by 12 and saying "that's our annualized!"

The problem with this number is that, well, people cancel things. While your June might be great, if 10% of your subscribers churn in a bad month (due to a change in your terms of service), that's a chunk of your *annualized* revenue gone.

But the worst sign is that ***nobody is saying the monthly figures, mostly because the monthly figures kinda suck!*** \$100 million of annualized revenue is \$8.33 million a month. To give you some scale, Amazon Web Services hit \$189 million (\$15.75 million a month) in revenue in 2008, two years after founding, and while it took until 2015 to hit profitability, it actually hit break-even in 2009, though it invested cash in growth for a few years after.

Right now, *not a single generative AI software company is profitable, and none of them are showing the signs of the kind of hypergrowth that previous "big" software companies had.* While Cursor is technically "the fastest growing SaaS of all time," it did so using what amounts to fake pricing. You can dress this up as "growth stage" or "enshittification (it isn't by the way, generally price changes make things profitable, which this did not)," but Cursor lied. It lied to the public about what its product would do long-term. It isn't even obvious whether its current pricing is sustainable.

Outside of Cursor, what other software startups are there?

***Glean?***

Everyone loves to talk about enterprise search company Glean — a company that uses AI to search and generate answers from your company's files and documents.

In December 2024, Glean raised \$260 million, proudly stating that it had over \$550 million of cash in hand with "best-in-class ARR growth." A few months later in February 2025, Glean announced it'd "achieved \$100 million in annual recurring revenue in fourth quarter FY25, cementing its position as one of the fastest-growing SaaS startups and reflecting a surging demand for AI-powered workplace intelligence." In this case, ARR could literally mean anything, as it appears to be based on quarters — meaning it could be an average of the last three months of the year, I guess?

Anywho, in June 2025, Glean announced it had raised *another funding round*, this time raising \$150 million, and, troublingly, added that since its last round, it had "...surpassed \$100M in ARR."

***Five months into the fucking year and your monthly revenue is the same? That isn't good! That isn't good at all!***

Also, what happened to that \$550 million in cash? Why did Glean need more? *Hey wait a second, Glean announced its raise on June 18 2025, two days after Cursor's pricing increase and the same day that Replit announced a similar hike!*

It's almost as if its pricing dramatically increased due to the introduction of Anthropic's Service Tiers and OpenAI's Priority Processing.

I'm guessing, but isn't it kind of weird that all of these companies raised money about the same time?

Hey, that reminds me.

## **There Are No Unique Generative AI Companies — And Building A Moat Based On Technology Is Near-Impossible**

If you look at what generative AI companies do (note that the following is not a quality barometer), it's probably doing one of the following things:

- A chatbot, either one you ask questions or "talk to"
  - This includes customer service bots
- Searching, summarizing or comparing documents, with increasing amounts of complexity of documents or quantity of documents to be compared
  - This includes being able to "ask questions" of documents
- Web Search
- "Deep Research" — meaning long-form web search that generates a document
- Generating text, images, voice, or in some rare cases video
- Using generative AI to write, edit or "maintain" code
- Transcription
- Translation
- Photo and video editing

Every single generative AI company that isn't OpenAI or Anthropic does one or a few of these things, and I mean *every* one of them, and it's because every single generative AI company uses Large Language Models, which have inherent limits on what they can do. LLMs can generate, they can search, they can edit (kind of!), they can transcribe (sometimes accurately!) and they can translate (often less accurately).

As a result, it's very, very difficult for a company to build something **unique**. Though Cursor is successful, it is ultimately a series of system prompts, a custom model that its users hate, a user interface and connections to models by OpenAI and Anthropic, both of whom have competing products *and* make money from Cursor and its competitors. Within weeks of Cursor's changes to its services, Amazon and ByteDance released competitors that, for the most part, do the same thing. *Sure* there's a few differences in how they're designed, but ***design is not a moat, especially in a high-cost, negative-profit business, where your only way of growing is to offer a product you can't afford to sustain.***

The only other moat you can build..is the services you provide, which, when your services are dependent on a Large Language Model, are dependent on the model



developer, ***who, in the case of OpenAI and Anthropic, could simply clone your startup, because the only valuable intellectual property is theirs.***

You may say "well, nobody else has any ideas either," to which I'll say that I fully agree. My Rot-Com Bubble thesis suggests we're out of hypergrowth ideas, and yeah, I think we're out of ideas related to Large Language Models too.

At this point, I think it's fair to ask — are there **any** good companies you can build on top of Large Language Models? I don't mean *add features related to*, I mean **an AI company that actually sells a product that people buy at scale that isn't called ChatGPT.**

## Established Large Language Models Are A Crutch

In previous tech booms, companies would make their *own* "models" — their own infrastructure, or the things that make them distinct from other companies — but the generative AI boom effectively changes that by making everybody build stuff on top of *somebody else's models*, because training your own models is both extremely expensive and requires vast amounts of infrastructure.

As a result, much of this "boom" is about a few companies — really two, if we're honest — getting other companies to try and build functional software for them.

## OpenAI And Anthropic Are Their Customers' Weak Point

I wanted to add one note — that, ultimately, OpenAI and Anthropic are bad for their customers. Their models are popular (by which I mean their customers' customers will expect access to them) meaning that OpenAI and Anthropic can (as they did with Cursor) arbitrarily change pricing, service availability or functionality based on how they feel that day. Don't believe me? Anthropic cut off access to AI coding platform Windsurf because it *looked like* it might get acquired by OpenAI.

Even by big tech standards this fucking sucks. And these companies will do it again!

# The Limited Use Cases Are Because Large Language Models Are All Really Similar

Because all Large Language Models require more data than anyone has ever needed, they all basically have to use the same data, either taken from the internet or bought from one of a few companies (Scale, Surge, Turing, Together, etc.). While they can get customized data or do customized training/reinforcement learning, these models are all transformer-based, and they all function similarly, and the only way to make them different is by training them, which doesn't make them *much different*, just *better at things they already do*.

## Generative AI Is Simply Too Expensive To Build A Sustainable Business On Top Of It

I already mentioned OpenAI and Anthropic's costs, as well as Perplexity's \$50 million+ bill to Anthropic, Amazon and OpenAI off of a measly \$34 million in revenue. These companies cost too much to run, and their functionality doesn't make enough money to make them make sense.

The problem isn't just the pricing, but how unpredictable it is. As Matt Ashare wrote for CIO Dive last year, generative AI makes a lot of companies' lives difficult through the massive spikes in costs that come from power users, with few ways to mitigate their costs. One of the ways that a company manages their cloud bills is by having some degree of predictability — which is difficult to do with the constant slew of new models and demands for new products to go with them, especially when said models can (and do) cost more with subsequent iterations.

As a result, it's hard for AI companies to *actually budget*.

## Companies Are Using The Term "Agent" To Deceive Customers and Investors

"But Ed!" you cry, "What about AGENTS?"

## Let me tell you about agents.

The term "agent" is one of the most egregious acts of fraud I've seen in my entire career writing about this crap, and that includes the metaverse.

When you hear the word "agent," you are meant to think of an autonomous AI that can go and do stuff **without oversight**, replacing somebody's job in the process, and companies have been pushing the boundaries of good taste and financial crimes in pursuit of them.

Most egregious of them is Salesforce's "Agentforce," which lets you "deploy AI agents at scale" and "brings digital labor to every employee, department and business process."

**This is a blatant fucking lie. Agentforce is a god damn chatbot platform, it's for launching chatbots, they can sometimes plug into APIs that allow them to access other information, but they are neither autonomous nor "agents" by any reasonable definition.**

Not only does Salesforce not actually sell "agents," its own research shows that agents only achieve around a 58% success rate on **single-step tasks**, meaning, to quote The Register, "tasks that can be completed in a single step without needing follow-up actions or more information." On multi-step tasks — so, you know, *most tasks* — they succeed a depressing 35% of the time.

Last week, OpenAI announced its own "ChatGPT agent" that can allegedly go "do tasks" on a "virtual computer." In its own demo, the agent took 21 or so minutes to spit out a plan for a wedding with destinations, a vague calendar and some suit options, and then showed a pre-prepared demo of the "agent" preparing an itinerary of how to visit every major league ballpark. In this example's case, "agent" took 23 minutes, and produced arguably the most confusing-looking map I've seen in my life.



It also missed out every single major league ballpark on the East Coast — including Yankee Stadium and Fenway Park — and added a random stadium in the middle of the Gulf of Mexico. What team is that, eh Sam? The Deepwater Horizon Devils? Is there a baseball team in North Dakota?

I should also be clear **this was the pre-prepared example**. As with every Large Language Model-based product — and yes, that's what this is, even if OpenAI won't talk about what model — results are extremely variable.

Agents are difficult, because tasks are difficult, even if they can be completed by a human being that a CEO thinks is stupid. What OpenAI appears to be doing is using a virtual machine to run scripts that its models trigger. Regardless of how well it works (it works very very poorly and inconsistently), it's also likely very expensive.

In any case, **every single company you see using the word agent is trying to mislead you**. Glean's "AI agents" are chatbots with if-this-then-that functions that trigger events using APIs (the connectors between different software services), **not taking actual actions, because that is not what LLMs can do**.

ServiceNow's AI agents that allegedly "act autonomously and proactively on your behalf" are, despite claiming they "go beyond 'better chatbots,'" still ultimately chatbots that use APIs to trigger different events using if-this-then-that functions. Sometimes these chatbots can also answer questions that people might have, or trigger an event somewhere. Oh, right, that's the same thing.

The closest we have to an "agent" of any kind is a coding agent, which can make a list of things that you might do on a software project and then go and generate the code and push stuff to Github when you ask them to, and they can do so "autonomously," in the sense that you can let them just run whatever task seems right. When I say "ask them to" or "go and" I mean that these agents are not remotely intelligent, and when let run rampant fuck up everything and create a bunch of extra work. Also, a study found that AI coding tools made engineers 19% slower.

Nevertheless, **none of these products are autonomous agents, and anybody using the term agent likely means "chatbot."**

*And it's working because the media keeps repeating everything these companies say.*

## But Really Though, Everybody Is Losing Money On Generative AI, And Nobody's Making A Profit

I realize we've taken kind of a scenic route here, but I needed to lay the groundwork here, because I am well and truly alarmed.

According to a UBS report from the 26th of June, the public companies running AI services are making absolutely pathetic amounts of money from AI:

Company	AI Apps Revenue
Microsoft	\$2.1b annualized revs
ServiceNow	> \$250m ACV
Adobe	\$125m ARR
Salesforce	> \$100m ARR
<b>Total</b>	<b>&gt; \$2.5b ARR</b>

ServiceNow's use of "\$250 million ACV" — so annual contract value — may be one of the more honest explanations of revenue I've seen, putting them in the upper echelons of AI revenue unless, of course, you think for two seconds, whether these are AI-specific contracts. Or, perhaps, are they contracts *including AI*? Eh, who cares. It's also year-long agreements that could churn, and according to Gartner, over 40% of "agentic AI" projects will be canceled by end of 2027.

And really, ya gotta laugh at Adobe and Salesforce, both of whom have talked so god damn much about generative AI and yet have only made around \$100 million in

annualized revenue from it. Pathetic! These aren't futuristic numbers! They're barely product categories! And none of this seems to include costs.

*Oh well.*

## OpenAI and Anthropic *Are* The Generative AI Industry, Are Deeply Unstable and Unsustainable, and Are Critical To The AI Trade Continuing

I haven't really spent time on my favourite subject — OpenAI being a systemic risk to the tech industry.

To recap:

- OpenAI and Anthropic both lose billions of dollars a year **after revenue**, and their stories **do not mirror any other startup in history, not Uber, not Amazon Web Services, nothing**. I address the Uber point in this article.
- SoftBank is putting itself in dire straits *simply to fund OpenAI once*. This deal threatens its credit rating, with SoftBank having to take on what will be multiple loans **to fund the remaining \$30 billion of OpenAI's \$40 billion round, which has yet to close and OpenAI is, in fact, still raising**.
  - This is before you consider the other \$19 billion that SoftBank has agreed to contribute to the Stargate data center project, money that it does not currently have available.
- OpenAI has promised \$19 billion to the Stargate data center project, money it **does not have** and **cannot get without SoftBank's funds**.
  - **Again, neither SoftBank nor OpenAI has the money for Stargate right now.**
- OpenAI must convert to a for-profit by the end of 2025, or it loses \$20 billion of the remaining \$30 billion of funding. If it does not convert by October 2026, its current funding converts to debt. It is demanding remarkable, unreasonable concessions from Microsoft, which is refusing to budge and is willing to walk away from the negotiations necessary to convert.
- OpenAI does not have a path to profitability, and its future, like Anthropic's, is dependent on a continual flow of capital from venture capitalists and big tech, who



must also continue to expand infrastructure.

Anthropic is in a similar, but slightly better position — it is set to lose \$3 billion this year on \$4 billion of revenue. It also has no path to profitability, recently jacked up prices on Cursor, its largest customer, and had to put restraints on Claude Code after allowing users to burn 100% to 10,000% of their revenue. These are the actions of a desperate company.

Nevertheless, OpenAI and Anthropic's revenues amount to, by my estimates, more than half of the entire revenue of the generative AI industry, **including the hyperscalers**.

**To be abundantly clear: the two companies that amount to around half of *all generative artificial intelligence revenue* are *ONLY LOSING MONEY*.**

I've said a lot of this before, which is why I'm not harping on about it, **but the most important company in the entire AI industry needs to convert by the end of the year or it's effectively dead, and even if it does, it burns billions and billions of dollars a year and will die without continual funding. It has no path to profitability, and anyone telling you otherwise is a liar or a fantasist.**

Worse still, outside of OpenAI...what is there, really?

## There Is No Real AI Adoption, Nor Is There Any Significant Revenue

As I wrote earlier in the year, there is really no significant adoption of generative AI services or products. ChatGPT has 500 million weekly users, and otherwise, it seems that other services struggle to get 15 million of them. And while the 500 million weekly users sounds — and, in fairness, is — impressive, there's a world of difference between someone using a product as part of their job, and someone dicking around with an image generator, or a college student trying to cheat on their homework.

**Sidebar:** Google cheated by combining Google Gemini with Google Assistant to claim that it has 350 million users. Don't care, sorry.

This is worrying on so many levels, chief of which is that everybody has been talking about AI for three god damn years, everybody has said "AI" in every earnings and media appearance and exhausting blog post, and we still can't scrape together the bits needed to make a functional industry.

I know some of you will probably read this and point to ChatGPT's users, and I quote myself here:

It has, allegedly, 500 million weekly active users — and, by the last count, only 15.5 million paying subscribers, an absolutely putrid conversion rate even before you realize that the actual conversion rate would be monthly active subscribers. That's how any real software company actually defines its metrics, by the fucking way.

Why is this impressive? Because it grew fast? It literally had more PR and more marketing and more attention and more opportunities to sell to more people than any company has ever had in the history of anything. Every single industry has been told to think about AI for three years, and they've been told to do so because of a company called OpenAI. There isn't a single god damn product since Google or Facebook that has had this level of media pressure, and both of those companies launched without the massive amount of media (and social media) that we have today.

ChatGPT is a very successful growth product and an absolutely horrifying business. OpenAI is a banana republic that cannot function on its own, it does not resemble Uber, Amazon Web Services, or any other business in the past other than WeWork, the other company that SoftBank spent way too much money on.

And outside of ChatGPT, there really isn't anything else.

## **Yes, Generative AI "Does Something," But AI Is Predominantly Marketed Based On Lies**

Before I wrap up — I'm tired, and I imagine you are too — I want to address something.

Yes, generative AI has functionality. There are coding products and search products that people like and pay for. As I have discussed above, **none of these companies are profitable, and until one of them is profitable, generative AI-based companies are not real businesses.**

In any case, the problem isn't so much that LLMs "don't do anything," but that people talk about them doing things they can't do.

- The use of the word "agent" is a deliberate attempt to suggest that LLMs are autonomous.
- Any and all stories about AI replacing jobs are intentionally manipulative attempts to boost stock valuations and suggest that models are capable of replacing human workers at scale. Allison Morrow of CNN has an excellent piece about this. As I discussed in this piece, this is one of the more egregious failures of the tech media I've ever seen, willingly publishing Dario Amodei outright making stuff up.
- The discussion of the term "AGI" is an attempt to suggest that Large Language Models can create conscious intelligence, a fictional concept that Meta's chief AI scientist says won't come from scaling up LLMs.
  - Members of the media: every time you talk about the "really smart engineers they're paying," know that you are doing marketing for these companies, when what's really happening is people are giving tens of millions of dollars to guys who will work on teams that are pursuing a totally-unproven concept.
- The use of the word "singularity" is similarly manipulative.
- The use of stories about models "lying, cheating and stealing to reach goals" or "stop themselves being turned off" are intentionally deceptive, as these models can (and clearly are) being prompted to take these actions.
  - To be abundantly clear, the manipulative suggestion here is that these models are autonomous or conscious in some way, which they are not.

I believe that the generative AI market is a \$50 billion revenue industry masquerading as a \$1 trillion one, and the media is helping.

## The AI Trade Is Entirely About GPUs, And Is Incredibly Brittle As A Result

As I've explained at length, the AI trade is not one based on revenue, user growth, the efficacy of tools or significance of any technological breakthrough. Stocks are not moving based on whether they are making money on AI, because if they were, they'd be moving downward. However, due to the vibes-based nature of the AI trade, companies are benefiting from the press inexplicably crediting growth to AI with no proof that that's the case.

OpenAI is a terrible business, and the only businesses worse than OpenAI are the companies built on top of it. Large Language Models are too expensive to run, and have limited abilities beyond the ones I've named previously, and because everybody is running models that all, on some level, do the same thing, it's very hard for people to build really innovative products on top of them.

And, ultimately, *this entire trade hinges on GPUs.*

CoreWeave was initially funded by NVIDIA, its IPO funded partially by NVIDIA, NVIDIA is one of its customers, and CoreWeave raises debt on the GPUs it buys from NVIDIA to build more data centers, while also using the money to buy GPUs from NVIDIA. This isn't me being polemic or hysterical — this is quite literally what is happening, and how CoreWeave operates. If you aren't alarmed by that, I'm not sure what to tell you.

Elsewhere, Oracle is buying \$40 billion in GPUs for the still-unformed Stargate data center project, and Meta is building a Manhattan-sized data center to fill with NVIDIA GPUs.

OpenAI is Microsoft's largest Azure client — an insanely risky proposition on multiple levels, not simply in the fact that it's serving the revenue at-cost but that Microsoft executives believed OpenAI would fail in the long term when they invested in 2023 — and Microsoft is NVIDIA's largest client for GPUs, meaning that any changes to Microsoft's future interest in OpenAI, such as reducing its data center expansion, would eventually hit NVIDIA's revenue.

Why do you think DeepSeek shocked the market? It wasn't because of any clunky story around training techniques. It was because it said to the market that NVIDIA might not sell more GPUs every single quarter in perpetuity.

Microsoft, Meta, Google, Apple, Amazon and Tesla aren't *making much money* from AI — in fact, they're losing billions of dollars on whatever revenues they do make from it. Their stock growth is not coming from actual revenue, but the vibes around "being an AI company," which means absolutely jack shit when you don't have the users, finances, or products to back them up.

So, really, everything comes down to NVIDIA's ability to sell GPUs, and this industry, if we're really honest, at this point only exists to do so. Generative AI products do not provide significant revenue growth, its products are not useful in the way that unlocks significant business value, and the products that have some adoption run at such a grotesque loss.

## I'm Alarmed!

I realize I've thrown a lot at you, and, for the second time this year, written the longest thing I've ever written.

But I needed to write this, because I'm really worried.

We're in a bubble. If you do not think we're in a bubble, you are not looking outside.

[Apollo Global Chief Economist Torsten Slok said it last week.](#) Well, okay, what he said was much worse:

*"The difference between the IT bubble in the 1990s and the AI bubble today is that the top 10 companies in the S&P 500 today are more overvalued than they were in the 1990s," Slok wrote in a recent research note that was widely shared across social media and financial circles.*

We are in a bubble. Generative AI does not do the things that it's being sold as doing, and the things it can actually do aren't the kind of things that create business returns, automate labor, or really do much more than one extension of a cloud software platform. The money isn't there, the users aren't there, every company seems to lose money and some companies lose so much money that it's impossible to tell how they'll survive.

Worse still, this bubble is entirely symbolic. The bailouts of the Great Financial Crisis were focused on banks and funds that had failed because they ran out of money, and the TARP initiative existed to plug the holes with low-interest loans.

There are few holes to plug here, because even if OpenAI and Anthropic somehow became eternal money-burners, *the AI trade exists based on the continued and continually-increasing sale and use of GPUs*. There are limited amounts of capital, but also limited amounts of data centers to actually put GPUs, and on top of that, at some point growth will slow at one of the Magnificent 7, *at which point costs will have to come down from things that lose them tons of money, such as generative AI*.

Before you ask-

## But, Isn't The Cost Of Inference Going Down?

You do not have proof for this statement! The cost of tokens going down is not the same thing as the cost of inference goes down! Everyone saying this is saying it because a guy once said it to them! You don't have proof! I have more proof for what I am saying!

While *it theoretically might be*, all evidence points to larger models costing more money, especially reasoning-heavy ones like Claude Opus 4. Inference is not the only thing happening, and if this is your one response, you are a big bozo and doofus and should go back to making squeaky noises when you see tech executives or hear my name.

## But Ed, What About ASICs?

Okay, so one argument is that these companies will use ASICs — customized chips for specific operations — to reduce the amount they're spending.

A few thoughts:

- When? Say OpenAI and Broadcom actually build their ASIC in 2026 (they won't) — how many of them will they build? Do they have contracts with companies that can actually produce high-performance silicon, of which there are only three (Samsung, TSMC, and arguably SMIC, which is currently sanctioned), and these companies



typically have their capacity booked well in advance. Even starting a production run of a semiconductor product can take weeks. Do they have the server architecture prepared? Have they tested it? Does it work? Is the performance actually good? Microsoft has failed to create a workable, reliable ASIC. What makes OpenAI special?

- It takes a lot of money to build these chips and they are yet to prove they're better than NVIDIA GPUs for AI compute, and even if they do, are they going to retrofit every data center? Can they build enough?
- If this actually happens, *it still fucks up the AI trade. NVIDIA STILL NEEDS TO SELL GPUs!*

I am worried because despite all of these obvious, brutal and near-unfixable problems, everybody is walking around acting like things are going great with AI. The New York Times claims everybody is using AI for everything — a blatant lie, one that exists to prop up an industry that has categorically failed to deliver the innovations or returns that it promised, yet still receives glowing press from a tech and business media that refuses to look outside and see that the sky is red and frogs are landing everywhere.

Other than the frog thing, I'm not even being dramatic. Everywhere you look in the AI trade, things get worse — no revenue, billions being burned, no moat, no infrastructure play, no comparables in history other than the dot com bubble and WeWork, and a series of flagrant lies spouted by the powerful and members of the press that are afraid of moving against market consensus.

Worse still, despite NVIDIA's strength, *NVIDIA is the market's weakness*, through no fault of its own, really. Jensen Huang sells GPUs, people want to buy GPUs, and now the rest of the market is leaning aggressively on one company, feeding it billions of dollars in the hopes that the things they're buying start making them a profit.

And that really is the most ridiculous thing. At the center of the AI trade sits GPUs that, on installation, immediately start losing the company in question money. Large Language

Models burn cash for negative returns to build products that all kind of work the same way.

If you're going to say I'm wrong, sit and think carefully about why. Is it because you don't want me to be right? Is it because you think "these companies will work it out"? This isn't anything like Uber, AWS, or any other situation. It is its own monstrosity, a creature of hubris and ignorance caused by a tech industry that's run out of ideas, built on top of *one company*.

You can plead with me all you want about how there are actual people using AI. You've probably read the "My AI Skeptic Friends Are All Nuts" blog, and if you're gonna send it to me, read the response from Nik Suresh first. If you're going to say that I "don't speak to people who actually use these products," you are categorically wrong and in denial.

I am only writing with this aggressive tone because, for the best part of two years, I have been made to repeatedly explain myself in a way that no AI "optimist" is made, and I admit I resent it. I have written hundreds of thousands of words with hundreds of citations, and *still, to this day*, there are people who claim I am somehow flawed in my analysis, that I'm missing something, that I am somehow failing to make my case.

The only people failing to make their case are the AI optimists still claiming that these companies are making "powerful AI." And once this bubble pops, I will be asking for an apology.

## I Don't Like What's Happening

I love ending pieces with personal thoughts about stuff because I am an emotional and overly honest person, and I enjoy writing a lot.

I do not, however, enjoy telling you at length how brittle everything is. An ideal tech industry would be one built on innovation, revenue, real growth based on actual business returns that helped humans be better, not *outright lie* about replacing them. All that generative AI has done is show how much lust there is in both the markets *and the media* for replacing human labor — and yes, it is in the media too. I truly believe there are multiple reporters who feel genuine excitement when they write scary stories about how

Dario Amodei says white collar workers will be fired in the next few years in favour of "agents" that will never exist.

Everything I'm discussing is the result of the Rot Economy thesis I wrote back in 2023 — the growth-at-all-costs mindset that has driven every tech company to focus on increasing quarterly revenue numbers, even if the products suck, or are deeply unprofitable, or, in the case of generative AI, **both**.

Nowhere has there been a more pungent version of the Rot Economy than in Large Language Models, or more specifically *GPUs*. By making everything about growth, you inevitably reach a point where the only thing you know how to do is spend money, and both LLMs and GPUs allowed big tech to *do the thing that worked before* — building a bunch of data centers and buying a bunch of chips — without making sure they'd done the crucial work of "making sure this would create products people like." As a result, we're now sitting on top of one of the most brittle situations in economic history — our markets held up by whether four or five companies will continue to buy chips that start losing them money the second they're installed.

I am disgusted by how many people are unwilling or unable to engage with the truth, favouring instead a scornful, contemptuous tone toward anybody who doesn't believe that generative AI is the future. If you are a writer that writes about AI smarmily insulting people who "don't understand AI," you are a shitty fucking writer, because either AI isn't that good or you're not good at explaining why it's good. Perhaps it's both.

If you want to know my true agenda, it's that I see something in generative AI and its boosters something I truly dislike. Large Language Models authoritatively state things that are incorrect because they have no concept of right or wrong. I believe that the writers, managers and executives that find it exciting do so because it gives them the ability to pretend to be intelligent without actually learning anything, to do everything they can to avoid actual work or responsibility for themselves or others.

There is an overwhelming condescension that comes from fans of generative AI — the sense that they know something you don't, something they double down on. We are being forced to use it by bosses, or services we like that now insist it's part of our

documents or our search engines, not because it does something, but because those pushing it need us to use it to prove that they know what's going on.

To quote my editor Matt Hughes: "...generative AI...is an expression of contempt towards people, one that considers them to be a commodity at best, and a rapidly-depreciating asset at worst."

I haven't quite cracked why, but generative AI also brings out the worst in some people. By giving the illusion of labor, it excites those who are desperate to replace or commoditize it. By giving the illusion of education, it excites those who are too idle to actually learn things by convincing them that in a few minutes they can learn quantum physics. By giving the illusion of activity, it allows the gluttony of Business Idiots that control everything to pretend that they do something. By giving the illusion of futurity, it gives reporters that have long-since disconnected from actual software and hardware the ability to pretend that they know what's happening in the tech industry.

And, fundamentally, its biggest illusion is economic activity, because despite being questionably-useful and burning billions of dollars, its need to do so creates a justification for spending billions of dollars on GPUs and data center sprawl, which allows big tech to sink money into something and give the illusion of growth.

I love writing, but I don't love writing this. I think I'm right, and it's not something I'm necessarily happy about. If I'm wrong, I'll explain how I'm wrong in great detail, and not shy away from taking accountability, but I really do not think I am, and that's why I'm so alarmed.

What I am describing is a bubble, and one with an obvious weakness: one company's ability to sell hardware to four or five other companies, all to run services that lose billions of dollars.

At some point the momentum behind NVIDIA slows. Maybe it won't even be sales slowing — maybe it'll just be the suggestion that one of its largest customers won't be buying as many GPUs. Perception matters just as much as actual numbers, and sometimes more, and a shift in sentiment could start a chain of events that knocks down the entire house of cards.

I don't know when, I don't know how, but I really, really don't know how I'm wrong.

I hate that so many people will see their retirements wrecked, and that so many people intentionally or accidentally helped steer the market in this reckless, needless and wasteful direction, all because big tech didn't have a new way to show quarterly growth. I hate that so many people have lost their jobs because companies are spending the equivalent of the entire GDP of some European countries on data centers and GPUs that won't actually deliver any value.

But my purpose here is to explain to you, no matter your background or interests or creed or whatever way you found my work, why it happened. As you watch this collapse, I want you to tell your friends about why — the people responsible and the decisions they made — and make sure it's clear that there *are* people responsible.

Sam Altman, Dario Amodei, Satya Nadella, Sundar Pichai, Tim Cook, Elon Musk, Mark Zuckerberg and Andy Jassy have overseen a needless, wasteful and destructive economic force that will harm our markets (and by a larger extension our economy) and the tech industry writ large, and when this is over, they must be held accountable.

And remember that you, as a regular person, can understand all of this. These people want you to believe this is black magic, that you are wrong to worry about the billions wasted or question the usefulness of these tools. You are smarter than they reckon and stronger than they know, and a better future is one where you recognize this, and realize that power and money doesn't make a man righteous, right, or smart.

I started writing this newsletter with 300 subscribers, and I now have 67,000 and a growing premium subscriber base. I am grateful for the time you've given me, and really hope that I continue to help you see the tech industry for what it currently is — captured almost entirely by people that have no interest in building the future.

**Subscribe to Ed Zitron's Where's Your Ed At**

Get the free version of my newsletter. No spam ever, unsubscribe anytime.

Your email address

Subscribe

#### ◉ SHARE



#### ◉ ABOUT THE AUTHOR

**Edward Zitron**

View all

#### MORE LIKE THIS



**Bubble Trouble**



**Godot Isn't Making it**





## Anthropic and OpenAI Have Begun The Subprime AI Crisis



## Burst Damage



## There Is No AI Revolution



## Deep Impact

### EXPLORE



## The Era Of The Business Idiot



## The Upcoming Remote Work Company Culture War



## The Rot Economy



## The Myth of Mentorship, And How We've Failed Entry-Level Workers



## The Man Who Killed Google Search



## Sam Altman Is Full Of Shit

○ COMMENTS



# Welcome to Where's Your Ed At!

Subscribe today. It's free. Please.

Your email address

SUBSCRIBE



[Home](#) [About](#) [Better Offline](#)

[Sign up](#)

©2025 Ed Zitron's Where's Your Ed At. Published with Ghost & Tuuli.

Ivory

⌵