"bubble" noun
bub-ble
1: a small globule that is typically hollow and light
2: something that lacks firmness, solidity, or reality
3: a state of booming economic activity that often ends in a sudden collapse [emphasis added]
4: a sound of or like that of bubbling or gurgling liquid
5: an enclosed or isolated sphere or experience or activity in which the like-minded members of a homogenous community support and reinforce their shared opinions
----Merrian-Webster Dictionary

Is AI a bubble?

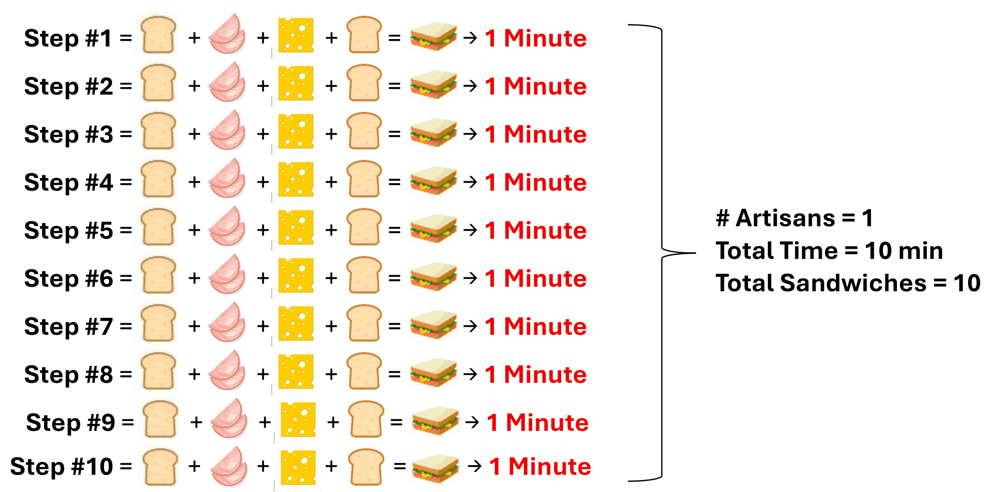
Does the sun rise in the East? Is the Pope Catholic? Is Notre Dame the best football team in the country? Okay, the last one may be a stretch.

Before we bore you with technical financial arguments about bubbles and valuations and all that stuff, let's address something mildly more interesting: What is AI?

AI starts with the hardware, or something called a graphics processing unit (GPU). The cousin to the GPU is the CPU, or central processing unit. Some of you may have heard of the CPU; they have been around for decades. GPUs are a comparatively newer technology.

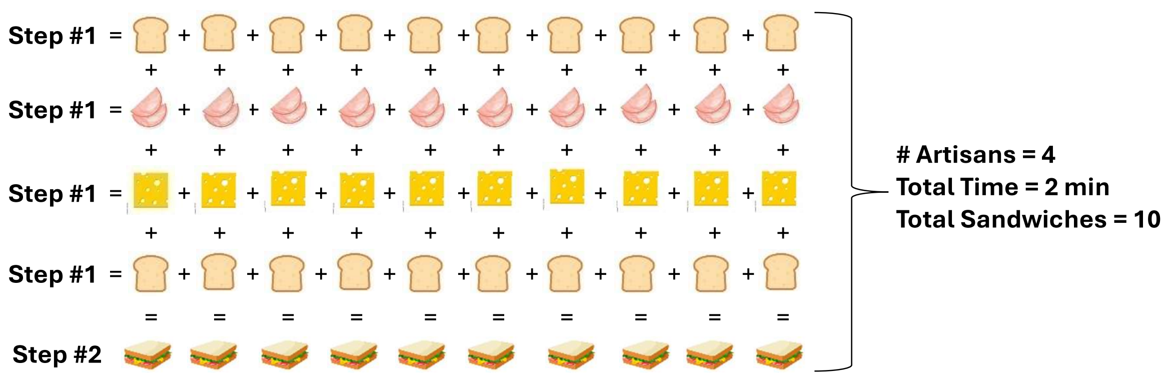
Conventional computing uses CPUs. AI uses GPUs. And here is the difference:

- The problem = You want to make 10 sandwiches[1].
- The solution:
  - In CPU world, the sandwich shop employs one artisan to make all the sandwiches. The artisan gets one slice of bread, adds a slice of meat, adds a slice of cheese, and finishes with another slice of bread. Each step is done sequentially. If it takes 1 minute to make one sandwich, it takes 10 minutes to make 10 sandwiches.



[1] This example was created using AI. In particular, Google Gemini was given the prompt "Give me an example of serial processing vs. parallel processing."

- In GPU world, the sandwich shop employs 4 artisans to make all the sandwiches. Each artisan is responsible for getting 10 of each ingredient: one artisan gets 10 slices of bread for the bottom; another gets 10 slices of meat; another gets 10 slices of cheese; and the last artisan gets 10 slices of bread for the top. Each artisan works simultaneously to get his ingredients. It takes each artisan 1 minute to get his 10 items. If they are all working simultaneously, it takes a total of 1 minute to gather all 40 ingredients. Once all the items are gathered, the four artisans work together to build all 10 sandwiches; that takes another minute. It takes them two minutes to make all 10 sandwiches.



To make a crude analogy, a CPU is like a 2nd grader doing simple addition and subtraction, while a GPU is a Ph.D. in mathematics doing complex number theory. Or it's like the old human-powered scooters we had as kids vs. the flying cars being invented now. A GPU is incredibly powerful, complex hardware compared to the conventional CPU hardware.

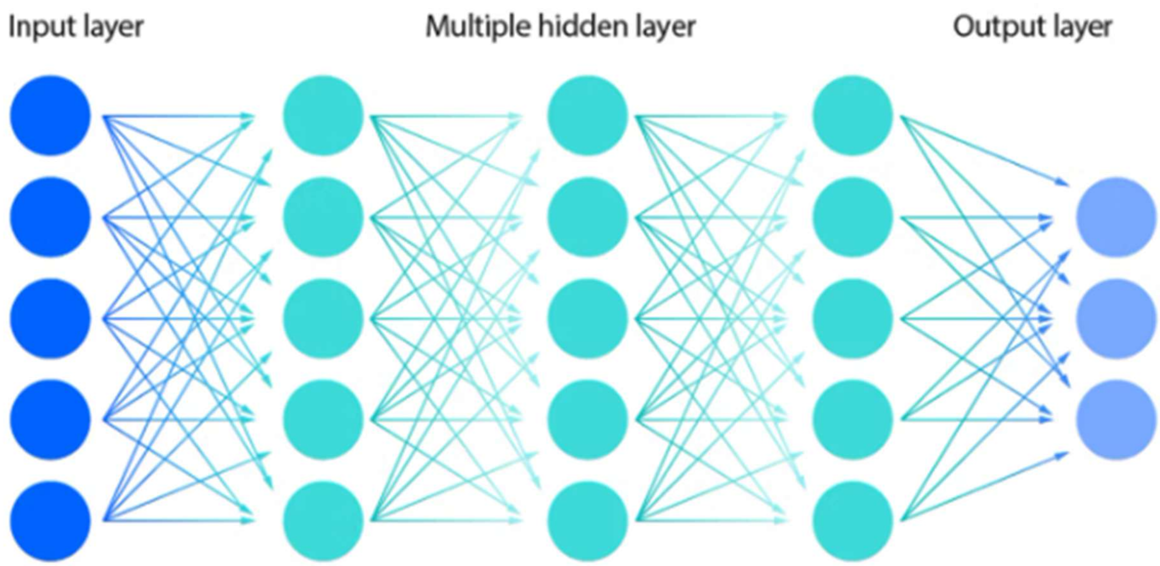
But a GPU is not AI. A GPU is only the hardware needed to run AI software.

AI is the collective term for the software "that can perform tasks needing human-like intelligence, such as learning, problem-solving, understanding language, and making decisions"[2].

In the old CPU, pre-AI world, you might go to Google and query "What are the symptoms of the common cold?", and you would get a range of results that directed you to webmd.com or mayoclinic.com. In new GPU, post-AI world, you might go to ChatGPT and query "I have a runny nose, a temperature of 99.8, and body fatigue. Please diagnose my ailment.", and in return, you'd get a comprehensive narrative of possible ailments; the symptoms and remedies of those ailments; and a range of detailed follow-up questions, that, if answered, would help the AI diagnose your exact ailment (much the same way a doctor would ask you probing questions to pinpoint your diagnosis).

[2] This is the definition per Google Gemini.

AI relies on neural networks. In essence, It is is a software version of the human brain. A quick Google search reveals this image[3] for a neural network:



Let's use a real-world example to illustrate how the neural network works in practice:

- Suppose we want to create an AI model that can read handwritten numbers. This would be an easy task if we all had perfect penmanship and crafted our numbers identically. But, of course, we all write differently. The Catholic nuns in my grade school would roll over in their graves if they saw my penmanship today! How on Earth would AI be able to recognize my numbers?
- Enter the **<u>input layer</u>**. In this layer, we "feed" a large data set into the model.
    - For example, let's say we feed 100,000 handwritten math tests from students ranging in age from kindergarten to college and from geographies all across the world. The model now has 100,000 images (the math tests) to examine.
- Next comes the **<u>hidden layers</u>**. Think of each of these layers as a bundle of individual calculators. In each layer, not all calculators are turned on. Maybe there are three hidden layers, and each layer has 100 calculators. Maybe only 50 calculators are turned on in layer 1; 90 in layer 2; and 70 layer 3. The layers are sequential. Layer 1 calculators run first; the results from layer 1 calculators become the inputs for layer 2 calculators; etc.
    - In our example, the layer 1 calculators are identifying rounded vs. straight edges of numbers. For example, some numbers have rounded edges (like 0, 2, 3, 5, 6, 8, 9) while some numbers have straight edges (like 1, 2, 4, 5, 7). Some numbers feature both rounded edges and straight edges (like 2 and 5).
    - Maybe layer 2 calculators examine the variations in the way number edges are written. For example, two can be written as "2"; or can have a small loop where the curved portion meets the straight portion; or maybe the straight portion has a bit of an arc to its shape. Or think about the number one. Some folks simply write a vertical line; some write a vertical line with a small horizontal line on the bottom and a diagonal hash coming off the top; some include the diagonal hash but not the bottom horizontal line.

[3] Source: https://www.ibm.com/think/topics/neural-networks

    - Maybe layer 3 analyzes the lengths of each edge. Some people write an eight with symmetrical loops – the top loop and bottom loop are identical in size. Some people write eight with a smaller upper loop and a large bottom loop.
- Last comes the **<u>output layer</u>**. After the hidden layers do all their calculations, the model spits out a result.

Who decides how many hidden layers to have? The programmer writing the code for the model.

Who decides how many calculators to have in each hidden layer? The programmer writing the code.

Who decides which calculators turn on or off in each layer? Or who decides that a figure that looks like a zero is actually the uppercase letter "O"? Or who decides that a poorly drawn "5" is not really a letter "S", or that a well-drawn lowercase "l" is not really the number one? Well, that's the secret sauce of AI.

See, when the hidden layers are doing their thing, it is often difficult, if not impossible, for humans to understand why the model sees a "S" instead of a "5" (when it seems so apparently to the human eye). Or why did layer 2 only use 90 of its 100 calculators. Put another way, oftentimes, the computing being done by the hidden nodes is only partially understood or not understood at all by humans.

Before we keep going, let's quickly address AI **<u>training</u>**. Training is exactly what it sounds like: programmer load large data sets into a model (**input layer**), and the model is trained to interpret the data (via the **hidden layers**).

In our example, we fed 100,000 math tests into the model. The hidden layers examined size and shapes and variation of the edges of the numbers on all those tests. After seeing all sorts of different styles of the number seven, for example, the model should be able to identify a seven written by anyone. But what if it doesn't?

In that case, the programmer tweaks the hidden layers. For example, suppose the AI model did not identify an eight because its loops were more oval in nature than circular in nature. The programmer can tweak the model to accommodate more oval-like shapes for the loops in an eight. Or, the programmer can feed another 100,000 math tests into the model in the attempt to "show" the AI that an 8 has many variations.

No model is perfect. Every model will make some errors. But at some point, the programmers are sufficiently satisfied with the model accuracy. And that's when they let the model out of the cage. This is called **<u>inference</u>**.

Inference is when anyone – you, me, your mother-in-law, etc. – can plug any handwriting we want into the model and ask it to identify all numbers in the handwriting. Because of the extensive **training** done by the model, the model can now **infer** what numbers are contained on the specific sample of handwriting that we uploaded into the model.

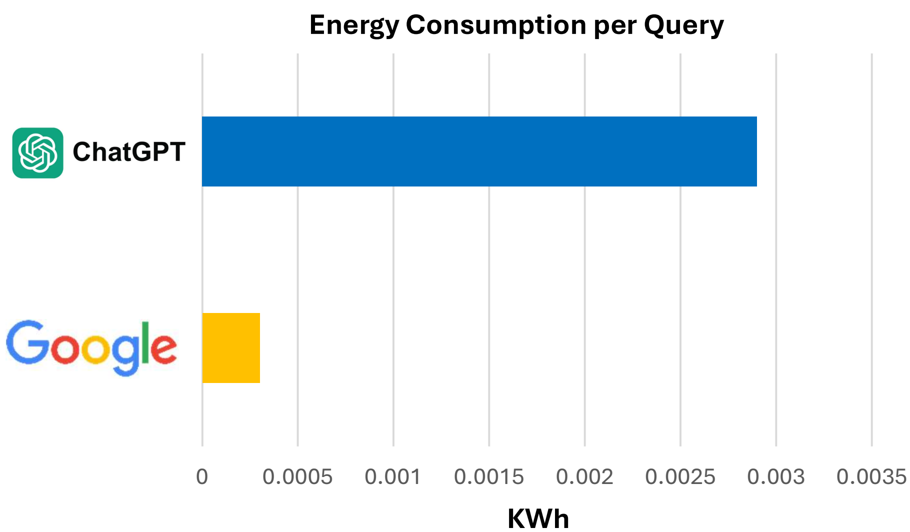
Now, let's tie this back to GPUs. The amount of calculations done by the calculators in each hidden layer is enormous. The calculations being done in hidden layer 1 are being done simultaneously.

For example, if layer 1 has 100 calculators and 50 are turned on, then 50 calculations are being done simultaneously. We know that GPUs are fantastic at doing multiple calculations at once, hence why they are necessary for building AI models.

Practical Limitations on AI

Recall in our sandwich example above that the sandwiches were made much faster in the "GPU" version of the sandwich shop then the "CPU" version of the shop. And that makes total sense when you consider that the GPU sandwich shop employed four artisans vs. the one artisan in the slower CPU shop. It is great when you can get your sandwich much faster, but are you willing to pay more that convenience?

See, GPUs can run circles around CPUs when it comes to computing power. But all the computing power comes at a price: it requires oodles more actual power to make GPUs run. Much like the GPU sandwich shop is faster but requires 4 workers, actual GPUs are way faster that CPUS but requires much more energy. How much more energy? We asked ChatGPT to help us with that[4].



ChatGPT uses 10x the amount of power consumption than Google. But so what? The average ChatGPT query only consumes 0.0029 KWh of power. That is an incredibly small amount. However, a lot of a little is still a lot. ChatGPT processes 2,500,000,000 queries per day[5]. That equates to 7,250,000 of kilowatt hours per day of power consumption.

[4] Chatgpt search query: "Compare energy consumption for chatgpt query vs google search query."
[5] We figured we would ask ChatGPT how hard it worked. The ChatGPT query was "How many queries does Chatgpt handle in one day?"

Just how much power is 7,250,000 kilowatt hours? Well, to take a page out of the Department of Energy's playbook, here is some perspective[6]. 7,250,000 kilowatt hours is equal to 7.25 gigawatts (GW). And 7.25 gigawatts is equal to:

13,680,000 Solar Panels

2,131 Onshore Wind Turbines

746 Offshore Wind Turbines

725,000,000 LED Bulbs

9,425,000 Horses

3.6 Hoover Dams

For perspective, the United States has about 75,000 wind turbines, including both onshore and offshore varieties[8]. Assuming all those turbines are operational all the time, the ChatGPT queries would consume 4% of all power output from those turbines on a daily basis!

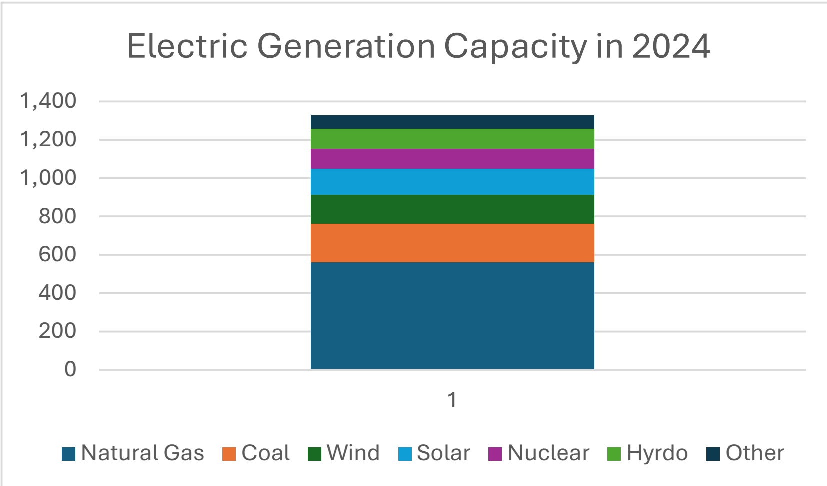
Or consider Hoover Dam, which provides power to 1,300,000 people[9]. 3.6 Hoover Dams would be enough to keep the lights on for 4,680,000 people. It would take an equal amount of energy to power all the queries from ChatGPT.

[6] Source: https://www.energy.gov/cmei/articles/how-much-power-1-gigawatt

[8] Source: https://energy.usgs.gov/uswtdb/#:~:text=Version:%20USWTDB_V8_2_20251210%20%2D%20Changelog%20%7C%20Detailed,about%20future%20updates%20and%20changes.

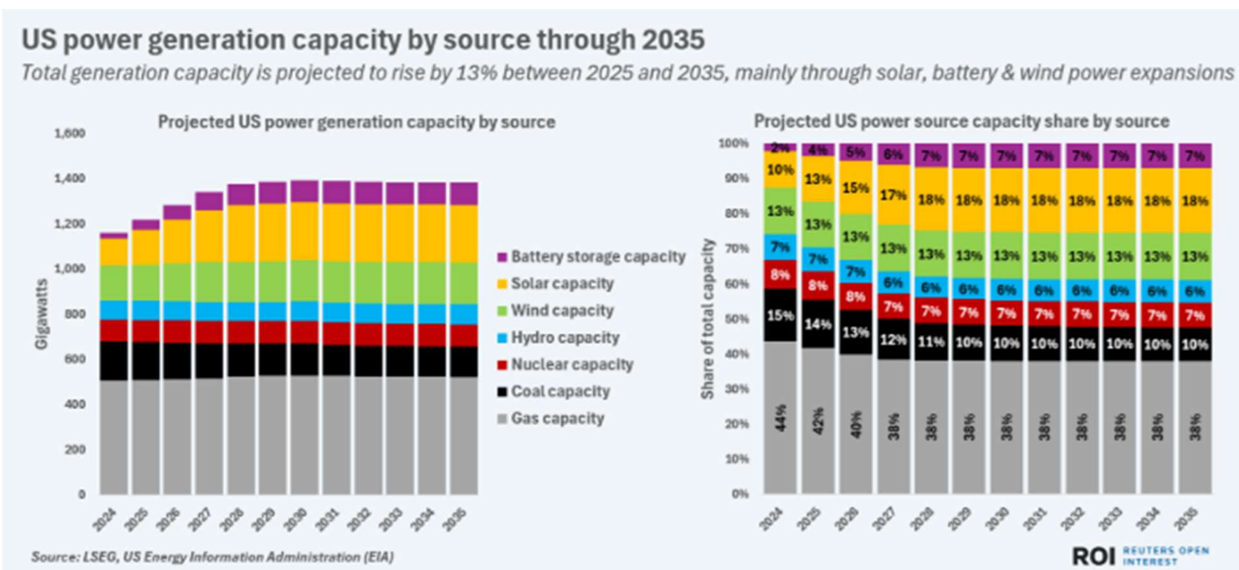
[9] Source: https://www.census.gov/about/history/stories/monthly/2025/september-2025.html#:~:text=In%20addition%20to%20flood%20control,more%20than%201.3%20million%20people.

In 2024, the United States had 1,326 GW of generating capacity[10]. Perhaps more importantly, more than 37 GW of generation capacity was added while 7.2 GW of capacity was retired. The net capacity added was 29.8 GW, or about 2.3% growth of total capacity generation.



If you have noticed your electricity costs increasing lately, you are about to find out why. The capacity of our current grid is bumping up against AI's insatiable appetite for power.

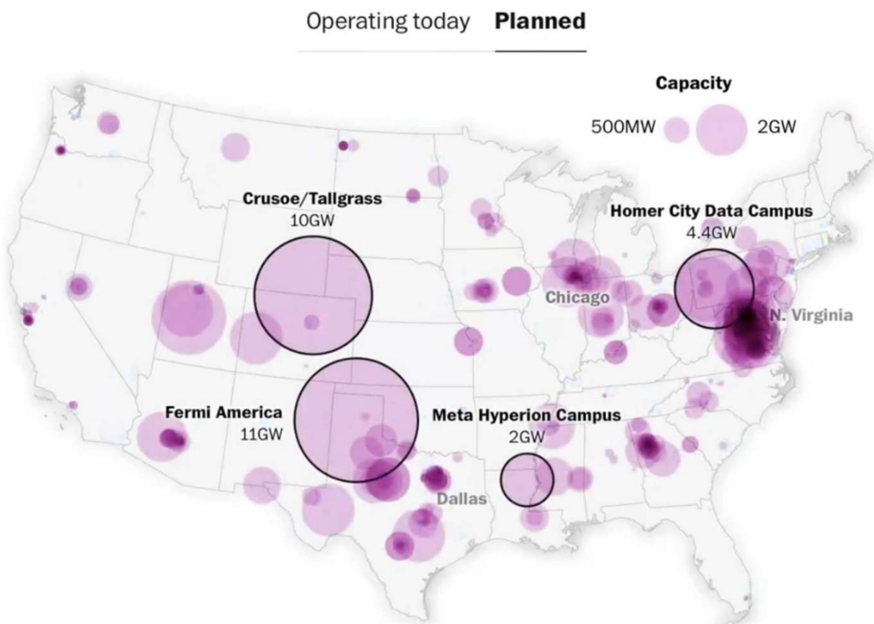
Here is Reuters' projection[11] for future U.S. power generation capacity through 2035. The takeaway from this chart is that capacity is expected to increase slowly over the next few years before tapering off into the 2030s. Put another away, the 2.3% growth rate in 2024 is probably the fastest our capacity will grow moving forward.



[10] Source: https://www.publicpower.org/system/files/documents/Americas-Electricity-Generation-Capacity-2025-Update.pdf#:~:text=The%20American%20Public%20Power%20Association's%20annual%20report,utility%2Dscale%20capacity%20by%20fuel%2C%20region%2C%20and%20ownership.

[11] Source: https://www.reuters.com/business/energy/charting-projected-us-power-capacity-mix-through-2035-2025-08-21/

Here is why that is a problem: AI will consume massive amounts of power, but the U.S. does not have the ability to generate that power. As of 2025, there are over 100 datacenters planned around the United States, all of varying size, but all requiring new sources of power generation. The chart below, published by the Washington Post[12],succinctly highlights all the planned datacenters.



Two problems become evident:

- First, the amount of new generation capacity needed is enormous.
- Second, the cost for new generation capacity is equally enormous.

As the datacenter map shows, the AI buildout will require a massive increase in generation capacity. In fact, a recent report from the National Center for Energy Analytics estimates that another 100 GWh of capacity will be needed by 2030[13]. Additionally, this does not include "the electricity demands for expanding the ancillary but directly related telecommunications networks, as well as that needed for reshored chip fabrication facilities that will manufacture the logic engines inside the data centers." If we conservatively assume the ancillary demand will be 10%, or 10 GWh, of the 100 GWh primary demand, that brings the total required capacity to 110 GWh.

On page 7, we noted that capacity is growing by about 2.3% per year for the next four years, or about 30 GWh per year. But, this growth is the natural growth of the existing system; it does not include AI-related growth. So, the U.S. needs another 110 GWh of capacity, or an increase of 8.3% in the next four years, to satisfy demand. That's a tall task.

But for the sake of argument, let us assume that the U.S. can build that much capacity. How is the capacity built? Here, the figures can get a bit murky. A quick ChatpGPT query for "What is the cost to build 1 GWh of electric capacity?" yields the following results:

[12] Source: https://www.washingtonpost.com/climate-environment/interactive/2025/giant-data-centers-energy-pollution/

[13] Source: https://energyanalytics.org/the-rise-of-ai-a-reality-check-on-energy-and-economic-impacts/

Technology	Cost per kW (USD)	Approx. Cost per 1 GW-nameplate
Natural Gas (combined cycle)	≈ $700–$1,200/kW	**$700 M – $1.2 B**
Onshore Wind	≈ $1,400–$1,800/kW	**$1.4 B – $1.8 B**
Solar PV	≈ $1,300–$2,700/kW	**$1.3 B – $2.7 B**
Utility Battery Storage (power capacity)	≈ $1,000–$4,500/kW	**$1.0 B – $4.5 B**
Nuclear	≈ $6,700–$7,500/kW	**$6.7 B – $7.5 B**

Off the bat, natural gas looks like a great option. It is the cheapest option to build. The U.S. as a bountiful supply to natural gas. And natural gas plants can typically be built faster than nuclear, for example.

However, there is a fly in the ointment. Natural gas, like some of the other technologies on this chart, have lower uptimes (the amount of time the plant is actually operational) compared to nuclear, for example. The table below came from another ChatGPT query: "What is the average uptime of these various technologies?

Technology	Approx. Capacity Factor (Uptime Equivalent)	Notes
Nuclear	~90–93 %	Very high utilization; designed for continuous baseload operation. The Department...
Natural Gas (Combined Cycle)	~55–62 %	Flexible operation; runs based on demand and fuel prices. The Department... +1
Coal	~40–60 %	Historically high but declining; varies by economic dispatch. Sustainability Di...
Hydroelectric	~30–45 %	Depends on water availability and seasonal flows. Sustainability Di...
Geothermal	~70–90 %	Relatively stable where resources are good. 1000whats
Biomass	~50–80 %	Variable by plant size and fuel supply. 1000whats

AI is always running. It does not sleep. It will consume power relentlessly. And therefore, it needs to be fed power relentlessly. Uptime matters tremendously.

Nuclear is the most expensive, but it has the most uptime. Natural gas is the cheapest, but it has much less uptime. In other words, one natural gas plant is cheaper to build than one nuclear plant, but we may need multiple natural gas plants to deliver the same amount of capacity as one nuclear plant <u>assuming power consumption is needed continuously 24/7</u>. The ratio of natural gas to nuclear uptime is 67%; in other words, natural gas plants only operates for 2/3 of the time that nuclear operates. So for every 1 nuclear plant built, we would essentially need 1.5 natural gas plants.

Thus, on the low-end, total cost to build 110 GWh of capacity from natural gas would be almost $200B [110 GWh needed * $1.2B per GWh * 1.5 ratio]. On the high-end, total cost to build the same capacity for nuclear would be $825B [110 GWh needed * $1.2B per GWh].

That's a big range in costs: $200B - $825B. But even on the low end of the range, where is $200B going to come from? And that doesn't include other costs like the actual fuel to make the electricity. Or the battery storage to store electricity.

This money must come from somewhere. And that somewhere is public and private investment. This is where part of the AI bubble comes into play. As we will see later, hoards of money is flowing into AI investments, both data centers, energy capacity, and ancillary technologies. Is that money chasing a goal that is too expensive? Or a goal that is on too ambitious of the timeline?

Shady Accounting

Forgive us, but now we have to talk about accounting! Accounting matters because Wall Street uses accounting to determine if an investment is a good investment or a bad investment. More profits equal good investment. Less profits equal bad investment. Accounting is the means by which we compute profits.

<u>Round-tripping</u>

Let's tee up a hypothetical scenario.

- Bob is a biologist who grows sun poms, the most delicious fruit in the world. In fact, the sun pom is a brand new hybrid fruit created by Bob. In a blind taste test, 100 tasters preferred the sun pom over all other fruits. Bob is determined to create a huge market for his sun poms, and since he holds the intellectual property protecting his hybrid strain, he essentially controls the market.
- Maria enjoys baking. She decides to open a bakery and bake sun pom pies. She speculates that many people will want a sun pom pie based on the blind taste test results. But of course, actual demand for pies is unknown. And she needs financing to get her bakery up and running.
- Meanwhile, other biologists are trying to reverse engineer the hybrid strain and create similar fruits to the sun pom.

Bob and Maria enter into an agreement as follows:

- Bob invests in Maria's business, owning a minority stake (less than 10%). Maria receives cash from the investment.
- Bob sells sun poms to Maria. Maria uses the sun pom inventory as collateral for a loan from outside lenders.
- With the cash from the loan and the cash from Bob's equity investment, Maria starts baking her pies.
- If Maria does not sell her pies, or only sells some of her pies, Bob agrees to buy all unsold pies from Maria.

There are a couple of problems here:

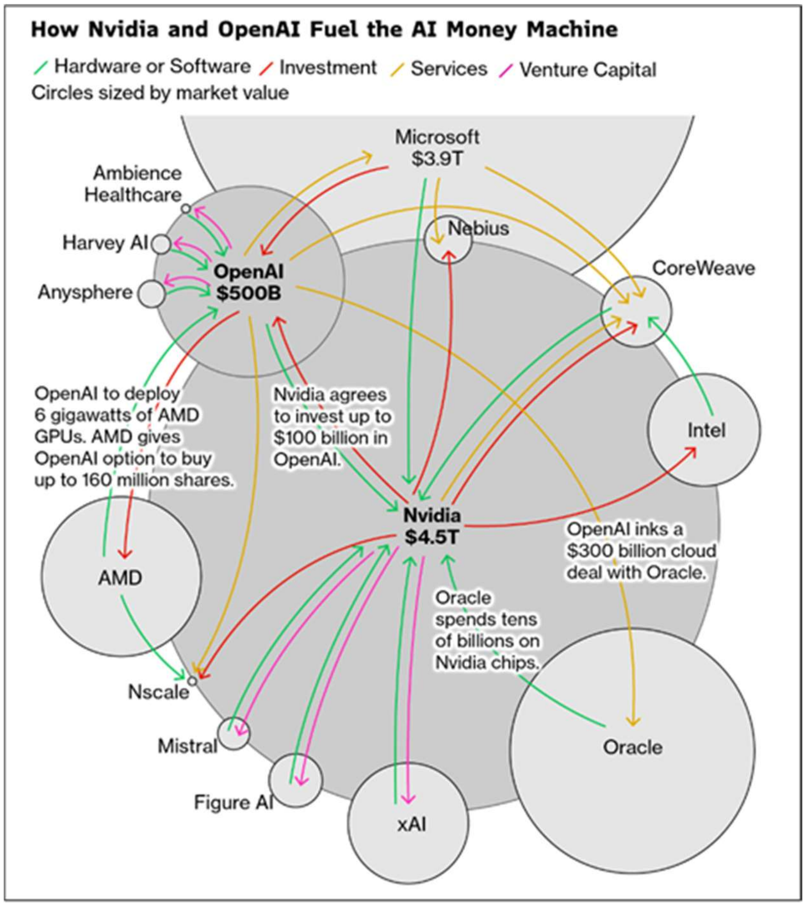
1. The only way Maria could get a loan to buy sun poms was to pledge those sun poms as collateral. So before she even acquired the sun poms, she needed a guarantee from Bob that the sun poms would be delivered. No one is going to lend Maria, a new baker, money unless she has access to the sun poms.
2. Bob is more than happy to guarantee the delivery of sun poms because he gets to book revenue for the sale of sun poms to Maria.
3. The sun poms are pledged as collateral for the loan. But sun poms are fruit, and fruit rots quickly. The fruit will rot within a week. If the fruit is converted into pies, the pies will spoil within a week. But that loan might not mature for a few months or even a few years. Thus, the collateral becomes worthless while the loan is still outstanding. If Maria cannot repay the loan, the lender gets wiped.
4. Bob must buy all of Maria's unsold pies. Why would he agree to this? Because he only cares about promoting his sun poms and maximizing his revenue. Revenue drives valuation, and his sun pom company becomes more valuable as he sells more sun poms. (Yes, having to spend money to buy unsold pies is an expense, but he can worry about that another day. His goal today is to maximize revenue).

If this all sounds complicated and confusing, it is. If Maria never sells one pie, she doesn't care. Why? Because Bob guaranteed to buy all her pies? But the same money he'll use to buy her pies is money he received from Maria when she purchased all the sun poms from him. The money just moved in a circle. But where did that money come from? It came from the lender. And remember, the lender's collateral is worthless when the fruit is either converted to pies or rots. Weird, right?

But what if Maria sells pies to customers instead of to Bob? Excellent question. Now fresh money enters the picture, and real sales are generated. Money is no longer moving in a circle; customers' payments for pie inject fresh money into the system. This seems logical, perhaps even probable. Until you consider that other biologists are working to create their own unique strains of fruit to compete with Bob. And guess what? Those biologists are starting to see signs of success. And when they're successful, they'll encroach on Bob's monopoly.

Our example is hypothetical. But it is illustrative of what is happening in the AI space. For example, take Nvidia and Coreweave. In the real world, Nvidia = Bob and Coreweave = Maria. Nvidia sells GPUs to Coreweave. Corewave buys those GPUs with borrowed cash. Coreweave than builds data centers with the GPUs. Nvidia agrees to rent data center space from Coreweave. And money flows in a circle.

This is not just a Nvidia-Coreweave gimmick. It is an entire AI industry gimmick. Here is a great graphic from the folks over at Bloomberg.



Surely you recognize some of the names on this graphic. It's a Who's Who of Silicon Valley: OpenAI, Microsoft, Nvidia, Oracle, AMD.

In some way, shape, or form, these tech giants and many more are engaging in some form of the accounting practices noted above.

These accounting practices, as well as the ones we discuss below, are not illegal. In fact, they are permitted under normal accounting rules.

But they do beg the question of why such practices are needed if the demand for AI is indeed so pervasive that it jumps the barrier out of Silicon Valley and into the real economy.

Depreciation

Recall how the sun poms were either rotted or converted to pies before the loans funding their purchase matured. The collateral is worth less than the loans.

This same dynamic is occurring in the AI space.

GPUs are highly technical hardware. The individual components are fragile. When set up in giant arrays and pumped full of electrical current on a regular basis for years on end, these GPUs deteriorate. They break down. They need to be replaced. This makes sense. You cannot use something continuously all day every day and expect it to last forever. Herein lies the problem.

There is no black-and-white study on the useful life of a GPU. Over the last year, various news outlets have written on the subject. And is typical with new technology, there is a wide range of estimates for useful life: as low as one year and as high and seven to eight years. A Google AI architect is quoted in a Tom's Hardware article[14] saying that the functional GPU life is 1-3 years. A

[14] Source: https://www.tomshardware.com/pc-components/gpus/datacenter-gpu-service-life-can-be-surprisingly-short-only-one-to-three-years-is-expected-according-to-unnamed-google-architect

senior director at Lambda, a datacenter operator, said in a recent Business Insider article that GPUs can be utilized for seven to eight years[15]. Yet another article from Mass Compute laid out various scenarios, suggesting the GPUs depreciate rapidly in the first three years, and after those three years, can be repurposed for other uses that would extend useful life, albeit with lower performance[16]. And then there is The Economist magazine, who in a recent piece noted that Jensen Huang, CEO of Nvidia, the primary global manufacturer of GPUs, implied in 2024 that GPUs are basically obsolete each time a new version of GPU is released[17]. Nvidia has released a new GPU version each year, which would mean the previous year's version is obsolete after only 12 months.

In summary, the consensus appears to be that GPUs experience peak performance (and thus, rapid depreciation) in the first three years of usage. After that, they may remain functional, but perhaps not at a level that would be acceptable for running large-scale AI operations.

Why is this even relevant? Because it affects profits.

See, under accounting rules, companies are allowed to determine what depreciation schedules they wish to apply to GPUs. Here, most technology companies are choosing a depreciation schedule between four and six years.

And thus, a mismatch is created: actual functional lifespan is around three years (give or take a margin of error) while accounting lifespans are 4-6 years.

Perhaps a quick graphic would help illustrate the problem. Suppose XYZ Corp buys a GPU for $6,000. Let's assume the functional life is 3 years, but XYZ's accounting department chooses to depreciate it over 5 years.

	Depreciation over Functional Life	Depreciation over Book Accounting	
Depreciable Life	3 Years	5 Years	**A**
Cost	$6,000	$6,000	**B**
Annual Depreciation	$2,000	$1,200	**C=B/A**

Book accounting figures are the numbers used in financial reports that companies release to Wall Street. They are the figures that Wall Street uses to assess the success, or lack thereof, of a company. If accounting depreciation is less than actual depreciation, companies may be overstating profits. Remember, depreciation is an expense. So a smaller expense equals a larger profit.

---

[15] Source: https://www.businessinsider.com/ai-bubble-argument-wrong-gpus-nvidia-depreciation-data-centers-crusoe-2025-11#:~:text=Matt%20Rowe%2C%20senior%20director%20of,is%20possible%2C%22%20Rowe%20said.

[16] Source: https://massedcompute.com/faq-answers/?question=What%20is%20the%20typical%20lifespan%20of%20an%20NVIDIA%20data%20center%20GPU?#:~:text=Expected%20Lifespan%20in%20Data%20Center,GPUs%2C%20leading%20to%20longer%20lifespans.

[17] Source: https://www.economist.com/business/2025/09/18/the-4trn-accounting-puzzle-at-the-heart-of-the-ai-cloud

Michael Burry, the famous investor who made a fortune betting against the housing market in 2008 and was popularized in the movie The Big Short (an excellent watch, by the way), articulated this accounting argument in a recent Substack piece[18]. He estimates that depreciation will be understated (the same of profit being overstated) to the tune of $176 billion between 2026-2028.

And The Economist, in the same article cited previously, argues that reducing accounting depreciation to three years will reduce the collective value of Microsoft, Amazon, Google, Oracle, and Facebook by $780 billion. Depreciating GPUs over two years would reduce their value by $1.6 trillion.

We would argue that this debate over depreciation is just that: a debate. AI is in its infancy, and the true useful life of GPUs is simply TBD. But if it does turn out the useful lives are shorter than accounting lives, that fact could have a massive negative impact to the stock prices of AI companies.

Securitization

We won't spend much time here. This is the abridged version. A few pages back we discussed the loan that Maria got to start her sun pom pie business. And how the collateral for those loans may be worthless (or at least worth less than the loan value itself). So if you are the lender, you have a problem on your hands: You are owed money, and you might not be able to recover all the money you are owed. Insert securitization.

Imagine a whole bunch of these loans get issued. The lenders want to offload the risk of holding these loans. So the lenders bundle various loans together and sell the loans to investors. This is called an asset-backed security.

Basic example:

- 100 loans are issued.
- The lenders who issued these loans bundle them up into a new investment. We'll call the investment ABC.
- The lenders sell ABC to investors. The lenders make a commission from this sale. The lender also no longer owns the loans; they have offloaded the risk to the investors.
- The investors collect the interest owed on the loans from the borrowers. The investors collect the principal from the loans when the loans mature. And if the borrowers default, the investors absorb the losses.

Securitization is not an inherently bad thing. Home mortgages get packed up and sold as asset-backed securities. Car loans, too. All sorts of loans get packaged and resold this way.

Securitization goes haywire when the original loans go bad. Think 2008 housing bubble. One of the main causes of the housing bust was securitized home loans becoming worthless when homeowners default on their mortgages. The jury is out on AI loans. We already discussed how these loans may be risky. If that risk indeed materializes, investors could be left holding the bag.

[18] Source: https://substack.com/inbox/post/180598391?r=6p7b5o&utm_medium=ios&utm_source=post-publish&triedRedirect=true

Special purpose vehicles

Like securitization, this is another complex accounting tactic. An example can help drive home the idea:

- Suppose a school wants to build a new computer lab at a cost of $1,000,000
- The school does not want to finance the project itself. Why? Because the school either lacks the financial resources to afford it, or the school has the financial resources, but it does not want to spend those resources.
- The school creates the Computer Lab Club ("CLC").
- The CLC is a standalone legal entity. Parents and donors give money to CLC. The school may provide funds to CLC, too. However, the school will not provide more than 20% of the total funding. In our example, the parents and donors raise $800,000, and the school kicks in the other 20% at $200,000.
- The CLC builds the computer lab. The CLC rents time on the computers to the students at the school. The CLC uses the rental income to pay back the parents and donors who financed the lab. The school might backstop the deal by promising to pay rent to the CLC if it does not collect enough rent from the students.

In the real world, it looks something like this:

- Facebooks wants access to AI datacenters, but it does not want to build the datacenter itself
- Private lenders and investments bank, like JP Morgan, form a special purposes vehicle (SPV) to build the datacenter. The lenders and JP Morgan loan money to the SPV. Facebook may also take a small equity stake in the SPV.
- The SPV builds the datacenter and leases space to Facebook.
- Facebook may enter into a contract with the SPV to guarantee some payment of rent to the SPV, thus guaranteeing a revenue stream to the SPV investors.
- With little to no actual capital outlay, Facebook gets datacenter access.

In both the hypothetical school computer lab example and the real-world Facebook example, there is a strong economic reason why the CLC / SPV is needed. The school / Facebook benefits because it gets access to the resources it wanted. Neither the school nor Facebook needs to borrow money or spend its own cash to build the resources. Yes, both the school and Facebook pay rent to use the resources, but paying rent is substantially cheaper than building the resources outright. The investors in the CLC / SPV get paid back with the rental income received by the CLC / SPV.

But here is the problem. We noted earlier that GPUs may deteriorate faster than expected. What happens when the investors / lenders to the SPV extend a loan to the SPV for five years, but the GPUs purchased and owned by the SPV deteriorate over two years? Or three years? All of sudden, one of a few realities sets in:

- New GPUs are needed to replace the old ones. The SPV must raise more debt, and the original lenders may be subordinated to the new lenders.
- The SPV does not buy new GPUs. The company renting from the SPV decides to stop renting since the old GPUs are effectively useless. The SPV sees a decrease in rental income, and thus, the lenders see a decrease in loan repayments.
- The SPV does not buy new GPUs. The company renting from the SPV has a long-term lease obligation to the SPV, so it must continue paying rent even if the GPUs are worthless. The

SPV realizes rental income and passes that income to the lenders in the form of loan repayments. But the company renting from the SPV has an expense with no corresponding benefit.

In any of the scenarios, there is a loser. Maybe even multiple losers. Multiply these loses by many SPVs, and it is not a stretch to witness contagion in the debt markets[19] [20].

One final observation: SPVs are neither inherently bad nor inherently good. They just are. But they can be abused when used incorrectly. Enron, the darling of Wall Street in the late 1990s, filed for bankruptcy due, in large part, to using SPVs for project financing and seeing those SPVs collapse.

Accounting Summary

Yes, the accounting is dense. The allure of AI is enough to convince some folks to gloss right over the accounting. But, at the end of the day, a company must show strong revenue and profit numbers to attract investors, and revenues and profits are both accounting metrics.

For those gluttons for punishment, might we recommend this great read from Edward Zitron @ Where's Your Ed At?[21]. It is a long but digestible examination of the accounting practices at Nvidia, the posterchild for the AI bubble, and how those practices resemble similar practices used in past high-profile scandals such as Enron and Worldcom.

Valuation

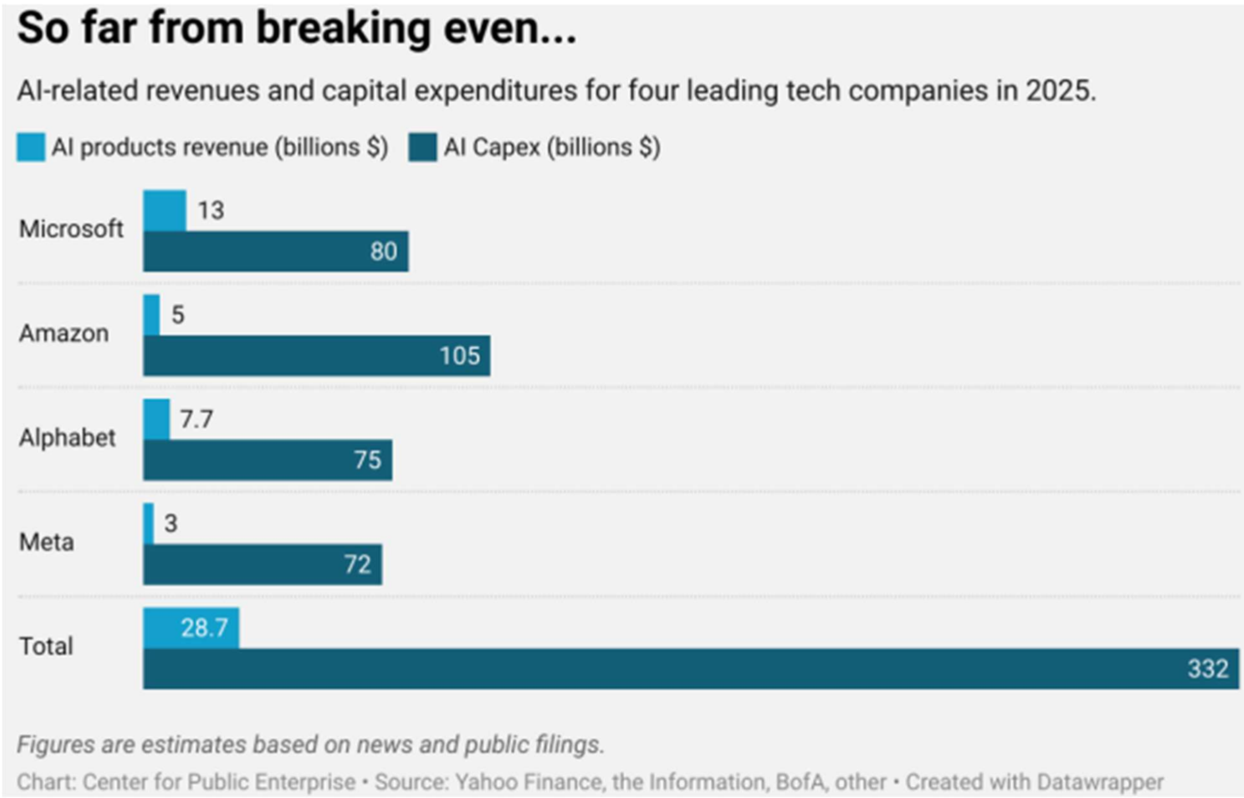
In our examination of the AI bubble thus far, we have discussed energy constraints and accounting issues. The final bubble topic concerns valuations in the AI sector. Those valuations are outright outrageous.

What is the point of making an investment? This is not a trick question. Anyone making any investment wants to make money on that investment. When you invest money into your 401(k), you hope you make money. When you buy a rental property, you hope you get rental income. We make investments with the expectation that we will make money on those investments.

This logic is no different for AI companies. For a company, investment is called CAPEX, short for capital expenditures. CAPEX is a simple concept: A company spends loads of cash on new projects. The companies hope those projects produce revenue for the company. What happens when a company spends money on projects, and those projects don't produce revenue? We may soon find out.

[19] Source: https://rondodson.substack.com/p/the-new-shadow-system-ai-finance
[20] Source: https://www.linkedin.com/pulse/ai-financing-through-spvs-quietly-shifts-risk-debt-reinout-te-brake-fxjie/
[21] Source: https://www.wheresyoured.at/nvidia-isnt-enron-so-what-is-it/

The chart below shows total AI CAPEX vs. AI revenue. Remember, if you're spending money on projects (CAPEX), you want revenue from those projects. By that metric, AI is a terrible investment. Total AI CAPEX for Microsoft, Amazon, Alphabet (aka Google), and Meta (aka Facebook) is $332 billion. But AI revenue from that CAPEX is only $28.7 billion[22]. For those doing the math, that works out to an 8.6% return [28.7 revenue / 332 CAPEX].



But we cannot stop here. Revenue is just the first step in figuring profits. Remember, profit is revenue less expenses. One way to back into profit using revenue figures is to determine the gross margin of a company. Gross margin is simply how much each $1.00 of revenue makes it to the bottom line. For example, a gross margin of 70% means that, for every $1.00 of revenue, the profit is $0.70[23]. Gross margins in AI seem to cluster around 60%[24].

If we apply a gross margin of 60% to the $28.7 billion in AI revenue from the prior page, we arrive at profits of $17.2 billion [$28.7 billion * 60%]. Using this profit figure of $17.2 billion, the actual return on CAPEX is 5.2% [$17.2 billion profit / $332 billion CAPEX].

---

[22] Source: https://publicenterprise.org/wp-content/uploads/Bubble-or-Nothing.pdf

[23] Strictly defined, gross margin is revenue less the cost of goods sold. From gross margin, operating expenses - like payroll, R&D, admin, and taxes – must be subtracted to arrive at profit. Our definition of gross margin here is simplified for purposes of making the argument easier to understand.

[24] Sources: https://archive.ph/JCxiL and https://www.tanayj.com/p/the-gross-margin-debate-in-ai

The funds for CAPEX must come from somewhere. That somewhere is from three sources:

(1) Existing assets (i.e. existing cash), and/or
(2) Issuance of new stock (i.e. selling stock to raise cash), and/or
(3) Borrowing money (i.e. issuing debt to raise cash)

All three options have some sort of "cost".

For example, if existing assets are used, the cost is the foregone alternate uses of those assets for other projects. Suppose cash could be used to invest in project ABC that would generate 5% returns. If that cash is instead used to finance project XYZ, then the company better be sure that XYZ returns more than 5%.

If new stock is issued, existing shareholders are diluted. Existing shareholders are going to require some level of returns to compensate them for the dilution.

And debt carries an interest rate. If that debt is financing project ABC and the debt interest rate is 6%, the project ABC better return more than 6%.

The weight average cost of capital (WACC) is essentially the "cost" of the three sources of CAPEX. A simple example: Suppose Company RST has a WACC of 6.5%. If RST invests $50 million of CAPEX into a new project, that project better return more than 6.5% to RST. If RST only earns 4.5% on a project with a WACC of 6.5%, than the company is losing value. Put another way, the company is "paying" 6.5% (the WACC) to earn a return of 4.5%.

We asked ChatGPT "What is the WACC for Microsoft, Amazon, Alphabet, and Meta?"

📊 Estimated WACC for Big Tech (2025)

Company	Estimated WACC
Microsoft (MSFT)	~10.3% GuruFocus
Amazon (AMZN)	~11.7% HL
Alphabet (Google, GOOGL)	~10.3% HL
Meta Platforms (META)	~11.4% (or ~11.6%) GuruFocus +1

So let's tie this all together. For the CAPEX investment to benefit the company – at least from a monetary perspective – the return on the CAPEX investment must be larger than the cost of the CAPEX investment. With this in mind, examine the table at the top of the following page . For each of the companies below, each company's CAPEX return < WACC. In other words, **the cost for each company to develop AI (WACC) is more than the benefits derived from that AI (CAPEX Return).**

	A	B	C=A*B	D	E=C/D	
	Revenue ($B)	**Gross Margin**	**Profit ($B)**	**CAPEX ($B)**	**CAPEX Return**	**WACC**
Microsoft	13	60%	7.8	80	9.8%	10.3%
amazon	5	60%	3	105	2.9%	11.7%
Alphabet Google	7.7	60%	4.62	75	6.2%	10.3%
Meta	3	60%	1.8	72	2.5%	11.4%
TOTAL	28.7		17.22	332	5.2%	11.0%

To analogize this relationship, it is akin to taking a loan from the bank at 5% interest and investing the borrowed money into an investment paying 4%. It does not make any sense. It is an economically bad idea. It is lighting money on fire.

Of course, this analysis has been limited to just four companies. But these four companies are arguably the Four Horseman of the AI bubble. For anecdotal evidence that AI investment costs (WACC) may far surpass actual benefits of AI (CAPEX returns), we would like to highlight a few excerpts from a recent The New Yorker article[25] about corporate AI investment:

> *According to a recent survey carried out by economists at Stanford, Clemson, and the World Bank, in June and July of this year* [2025]*, almost half of all workers—45.6 per cent, to be precise—were using A.I. tools. And yet, a new study, from a team of researchers associated with M.I.T.'s Media Lab, reports, "Despite $30 - $40 billion in enterprise investment into GenAI, this report uncovers a surprising result in that 95% of organizations are getting zero return."*
>
> *The study's authors examined more than three hundred public A.I. initiatives and announcements, and interviewed more than fifty company executives. They defined a successful A.I. investment as one that had been deployed beyond the pilot phase and had generated some measurable financial return or marked gain in productivity after six months. "Just 5% integrated AI pilots are extracting millions in value, while the vast majority remain stuck with no measurable P&L"—profit-and-loss—"impact," they wrote.*
>
> *But the idea that many companies are struggling to achieve substantial returns jibed with another recent survey, by Akkodis, a multinational consulting firm. After contacting more than two thousand business executives, the firm found that the percentage of C.E.O.s who are "very confident" in their firm's A.I.-implementation strategies has fallen from eighty-two per cent in 2024 to forty-nine per cent this year.*

Then there is this beauty from the vaulted private equity firm Bain & Company: By 2030, Bain estimates that $2 trillion of AI-related revenue will be needed to support the planned build-out in AI over the next four years.

Per Barchart, the total revenue of the Computers and Technology Sector of the S&P 500 – basically, all the tech stocks in the largest stock index in the world – is $2.66 trillion. If revenue needs to

[25] Source: https://archive.ph/BGWrY

increase by $2 trillion over the next four years, revenue growth of 75% is needed [2.00 / 2.66]. This is not impossible, but it sure is heck will not be easy. Over the last four years, total revenue growth for this sector is 59%[26].

To be sure, AI is very important. We would argue it is critically important. It is becoming integrated into our lives, much the same way that the internet was integrated into our lives in the late 1990s. But we cannot and should not conflate importance with value. Today, we cannot live without the internet; it is essential to how everything functions. This importance was imaged by the stock market in the late 1990s when every tech stock went up and up and up and up…until they didn't. When the tech bubble burst, the NASDAQ – the tech-heavy stock index – fell over 80%. The internet was important then, and the potential for its expansion was unimaginable. But that did not mean that tech stocks should trade at nose-bleed levels. When the "market" finally realized that values had become detached from reality, stock investors – especially tech stock investors – got wiped out. AI is following the same pattern. Will its fate be any different?

If you managed to survive this far, well done! That was dense. And intense. If the bubble is real – and we very much think it is – what does that mean for the future of the stock market? We're going to borrow a page from past commentaries and review three fundamental metrics to help inform the direction and magnitude of stock returns ahead:

- Price to Sales
- Buffet Indicator
- Discounted Cash Flows

These measurements were discussed in detail in our June 30, 2024 commentary. We refer you to that commentary for a refresher on the mechanics and meaning of these metrics.

Price to Sale Ratio

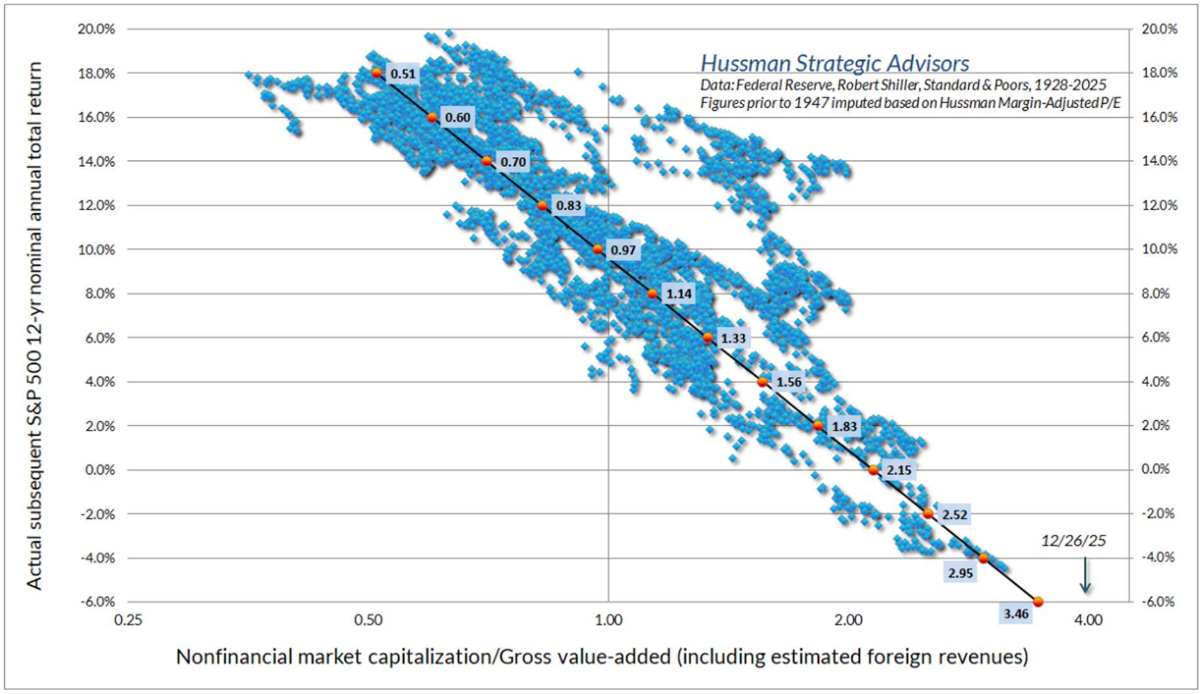
On the next page, we present our proxy for the price-to-sale ratio of the S&P 500. This chart is courtesy of Hussman Strategic Advisors[27]. [As an aside, we highly recommend the free monthly commentaries that Hussman makes available on its website (https://www.hussmanfunds.com/category/comment/).] The current ratio is literally off the chart, and that is a bad thing.

The current ratio is implying that the stock market may experience annual returns over the next decade of -8% per year. This implies a total drawdown (e.g. loss) of 57% in the next decade. Wowser!

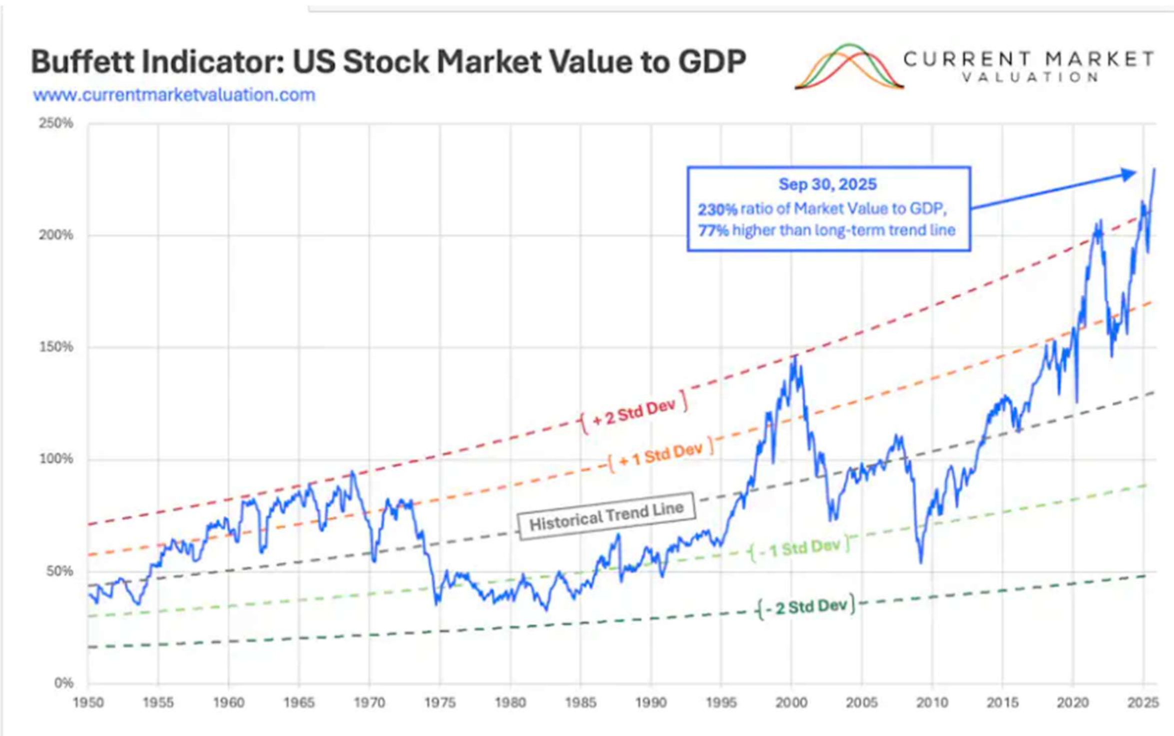
---

[26] Data derived from Barchart.com. NOTE: This calculation excludes revenue growth from NVDA. NVDA contributed roughly 66% of all revenue growth in the Computers and Technology sector over the last 4 years. NVDA revenue growth has been slowly substantially over time. Its 4-yr average growth is 170% per year, but its1-yr growth is 114%. While the 1-yr growth rate is still very high, it is a rapid deceleration from the 170% per year pace. Given NVDA's size in the S&P 500, its revenue growth exerted a long-term unrealistic effect on the overall index. Put another way. It is unreasonable to assume NVDA will continue to growth revenues at the long-run average (or even the short-run average).

[27] Source: https://www.hussmanfunds.com/comment/mc260104/



Buffet Indicator
Warren Buffett, the Oracle of Omaha, has a favorite indicator: the ratio of market cap to GDP. Per Warren, this one indicator is great way to anticipate stock returns <u>over the long-run</u>. Below is the chart of the Buffett Indicator over time[28].



[28] Source: https://www.currentmarketvaluation.com/models/buffett-indicator.php

Like the price to sale ratio, the Buffett Indicator is in unprecedented territory. At its current level of 230%, the indicator is implying that the stock market may experience annual returns over the next decade of -7% per year. This implies a total drawdown (e.g. loss) of 52% in the next decade. Wowser!

Discounted Cash Flows

The discounted cash flows (DCF) idea is a relatively simply concept. Suppose you want to earn 10% on your money. One year from now, you know your investment will pay you $100. How much would you pay for the investment today? After a little middle school algebra, you figure that you would pay $90.90 today. After all, $90.90 * 10% = $9.10. And $90.90 investment + $9.10 return = $100 payout.

Suffice it so say, if we know future cash flows (i.e $100 in our case above) and we know the actual value today (i.e. $90.90 in our example), we can solve for the rate of return.

Well, guess what? We can easily estimate future cash flows for the S&P 500. And we know the actual value of the S&P 500. So we can solve for the expected rate of return of the S&P 500.

The S&P 500 has historically averaged a 10.3% annual rate of return.

Using DCF modeling, we can estimate the following outcomes:

- At the 12/31/2025 value for the S&P 500 (e.g. 6,845), the expected future return is 6.1%. This is well below the historical average of 10.3%.
- For the S&P 500 to actually return 10.3% in the future, the stock market would need to fall by 78%. This may be a counter-intuitive concept. Why must the stock market go down to make future returns go up? If the starting value is lower (e.g. if the stock market decreases in value), it must gain more to hit the same ending goal. For example, recall our example of getting $100 one year from now. If we want a 5% return, we would pay $95.24 today. But if we want a 10% return, we would only pay $90.90 today. The starting value must go down for the return to go up.

Summary of Valuation Metrics

All three valuation metrics are widely-used and well-respected. And all metrics point to the same conclusion: If and when the AI bubble pops, stocks could be in for a wild ride. Our theses is that AI is causing a massive bubble, and we've laid out the case for why. Now the math backs up that argument. But remember, buy low and sell higher. When the stock market sells off, it's time to go bargain shopping for stocks!

	@ Dec 31, 2025	
	Implied Max Drawdown	Implied 10-Year Avg Return
Price to Sales	-57%	-8%
Buffet Indicator	-52%	-7%
Discount Cash Flow	-78%	-14%

The Future Looks Bright, but First…

Like most technological innovations, AI will reshape how the world functions. If used as a force for good, that innovation will make the world a better place. The internet served a similar function in the late 1990s. The internet has its fair share of negative qualities, but its impact on the world has been profoundly positive. The internet's growth, however, came with a price: a large, protracted recession from 2000-2002, that, once experience, cleared the way to that positive growth. Will AI experience the same trajectory? "History doesn't repeat itself, but it tends to rhyme."

…Ending on a High Note

We always like to end with a little levity. In honor of those of us suffering through one of the coldest winters in memory, here are a few jokes to warm the soul.