NexSky

CIO ANALYSIS

Revenue per Dollar of AI Capex: What the Number Leaves Out

Chip life decides what a dollar of capex earns. Falling token prices decide how long it keeps earning it.

Boutique Family Office

29 October 2026

1. The short version

Investors now quote one number for AI companies: how much revenue each dollar of capital spending earns. It is a useful sign of demand, but it says little about profit, because two things sit outside it.

> **Our view:** revenue per dollar of capex is a good sign of demand and a poor guide to profit. In our worked example, a chip written off over 3 years stops making a profit once the price of AI output falls about 35%. A chip written off over 6 years survives a fall of about 58%.

First, how long the chips last. A company chooses how many years to spread the cost of a chip over. The cash is the same either way, but the reported profit is not. In our example the profit margin is 35% on a 3 year life and 58% on a 6 year life. If the accounts say six years and the chips really last three, reported profit is about two thirds too high.

Second, how fast the price of AI output falls. Most costs are fixed once the chips are bought, so a price cut goes straight to profit. A 50% cut stretches the payback on our example from 2.3 years to 5.6 years, which is longer than a 3 year chip lasts.

Third, what the company sells. Newer models hold their price for now, while older models are turning into a commodity. The more of the revenue that sits in older models, the more exposed the company is.

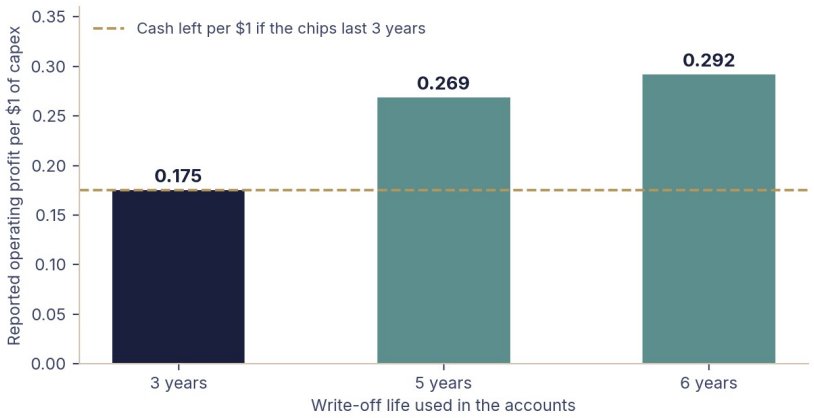
We would change our mind if three-year-old chips keep earning at high use in years four to six, and if volume keeps growing faster than prices fall. Section 4 sets out what to watch.

2. One dollar of capex, worked through

To see what the number leaves out, we follow one dollar of capital spending through a year. We use round numbers from public sources, not any one company's accounts, and show one year in cents under each chip life. The appendix gives the sources.

Input	What we assume
Where the dollar goes	70 cents on chips, servers and networking. 30 cents on the building, power and cooling, spread over 14 years
Revenue	50 cents a year
Running costs	7 cents a year: 3 for electricity and 4 for staff, maintenance and software
Chip life	3, 5 or 6 years. This is the one thing we change

Per $1 of capex, cents a year	3-year life	5-year life	6-year life
Revenue	50.0	50.0	50.0
Less running costs	7.0	7.0	7.0
Cash profit before replacing chips (EBITDA)	43.0	43.0	43.0
Less the yearly charge for the chips (depreciation)	25.5	16.1	13.8
Reported profit	17.5 (35%)	26.9 (54%)	29.2 (58%)



Same chips, same cash: the reported profit depends on the write-off life. The dashed line is the real profit if the chips last 3 years. Source: NexSky illustration.

The cash is 43 cents a year whichever chip life the accounts use, so the dollar pays back in 2.3 years. What changes is the yearly charge for the chips. Spread the cost over a longer life and the charge is smaller, so reported profit is higher. A measure that looks at cash profit before that charge, known as EBITDA, cannot see the difference at all.

The risk is a mismatch. If the accounts say six years and the chips last three, the company reports 29.2 cents of profit per dollar while the real figure, after replacing the chips, is 17.5 cents. If the chips do last six years, the accounts are right and the sceptics are wrong.

The example holds revenue per chip flat. In practice it may fall as chips age, and we leave out tax and the cost of capital. Each would make the picture harsher. No company discloses how revenue per chip fades with age, so we put no number on it.

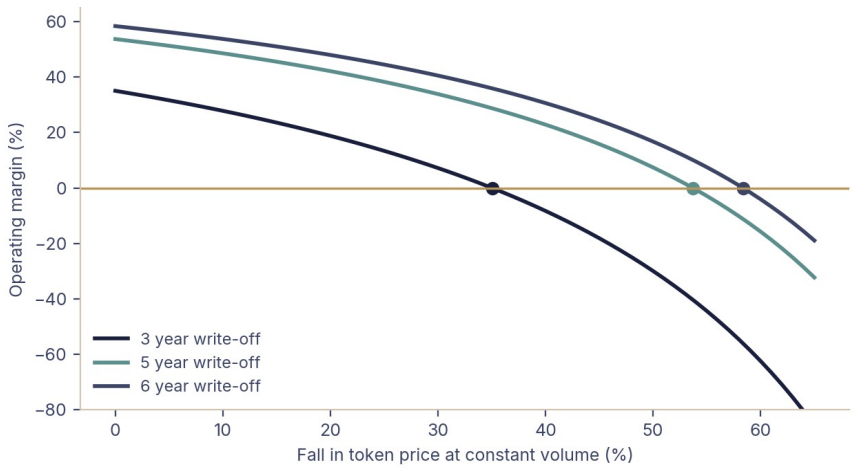
3. When the price of AI output falls

AI output is sold by the token, a small piece of a word. The price of a given level of capability is falling fast. Epoch AI, a research group, finds it has fallen between 9 and 900 times a year depending on the test, and cautions that the fastest falls may not last. The price pages of two major providers show the same pattern. Older models are much cheaper than a generation ago, while each new top model launches at a higher price. The market runs at two speeds: yesterday's capability is turning into a commodity, and the frontier holds its price until the next model arrives. The appendix has the evidence.

Most costs are fixed once the chips are bought, so a fall in price at constant volume comes straight off profit. The table keeps everything in section 2 and changes only the price.

Price falls by	Revenue per $1, cents	Margin, 3-year life	Margin, 6-year life	Payback
0%	50	35%	58%	2.3 years
20%	40	19%	48%	3.0 years
50%	25	Loss of 30%	17%	5.6 years

Profit reaches zero at a price fall of 35% for a 3 year life, 54% for 5 years and 58% for 6 years. At a 50% cut, the payback of 5.6 years is longer than the life of a 3 year chip, so the cash never comes back. A shorter chip life always breaks even at a smaller price cut.



A shorter chip life turns loss-making at a much smaller price cut. Dots mark break-even. Source: NexSky illustration.

More volume can offset a price cut, but only if the chips have spare capacity. A 50% cut needs twice the volume to hold revenue. Today compute is in short supply, so the cut falls straight through. There is a second effect over time. New chips serve more tokens per dollar, so a new fleet can sell at lower prices and still earn a decent return, while an old fleet cannot. That is how falling prices shorten the economic life of a chip.

The sales mix decides how much of this applies. If older models bring in 40% of revenue and their price halves while the newer ones hold, the average price fall is 20%, and a 3 year chip drops to a 19% margin. It also matters which chips serve the cheaper tier. Older models on old, largely written-off chips are a soft hit. Older models on the newest chips are a double hit. No company discloses this, so it is our reasoning and not a reported fact.

4. What to watch

The bulls are right if	The bears are right if
Three-year-old GPUs still rent at high use in years 4 to 6	Old chips reprice to a fraction of new ones and fleets rotate in 3 to 4 years
Volume growth keeps outrunning the fall in token prices	Prices fall faster than volume grows once capacity catches up
Operating cash flow covers capex and lease payments without new equity	Free cash flow stays negative and growth depends on the next raise
Peers do not shorten chip lives in coming filings	More companies follow Amazon and depreciation steps up

We will watch the next round of cloud company filings for any change in the estimated life of servers. For a reader of any AI revenue multiple, five questions matter. What chip life sits behind the depreciation, and has it changed? How much of the revenue is earned at yesterday's capability? Which chips serve that cheaper tier? What does a three-year-old chip still earn in year four? And how much of the cost base is committed in advance, because fixed commitments signed at today's prices are what turn a price cut into a loss?

Appendix: the evidence behind the numbers

A. What the large companies assume about chip life

The published accounts show where the industry sits. Most large cloud companies use five to six years for servers. Of the companies below, Amazon is the only one to have moved the other way.

Company	Life of servers and network equipment	Source
Alphabet	Generally over six years	10-K FY2025, filed 5 Feb 2026
Microsoft	2 to 6 years	10-K FY2026, filed 29 Jul 2026
Amazon	Part of the fleet cut from six to five years from 1 Jan 2025, adding $1.4bn of depreciation and cutting net income by $1.0bn. Earlier moved from five to six from 1 Jan 2024	10-K FY2025, filed 6 Feb 2026, Note 1
Meta	5.5 years for most servers and network assets from 1 Jan 2025	10-K FY2025, filed 29 Jan 2026
Oracle	Raised from five to six years in fiscal 2025	10-Q for the quarter to 31 Aug 2024
CoreWeave	Six years for technology equipment, raised from five from 1 Jan 2023	10-K FY2025, filed 2 Mar 2026

Amazon cited the faster pace of technology development, particularly in artificial intelligence. Microsoft's note gives a range with six as the top, not a flat six. Companies that rent most of their compute carry the landlord's assumption inside the fees they pay, so the chip life that matters for them is the cloud provider's.

B. How fast prices are falling

Epoch AI tracks what it costs to get a fixed level of AI capability, for example the score a leading model reached on a standard test. In a study published in March 2025, with data to February 2025, it found that the price of reaching a given level fell by 9 to 900 times a year, depending on the test. A fall of 9 times means the price is one ninth of what it was a year earlier, which is 89% lower. A fall of 900 times means 99.9% lower. Epoch cautions that the fastest falls came in the most recent year and that it is less clear they will persist.

Two major providers' own price pages, read on 29 October 2026, show the same pattern across model generations. Each row compares a tier of model with the generation before it. We pair the models by name, because the providers do not state the pairings.

Tier	Provider A	Provider B
Small	List price down 90%	Down 50% to 58%
Mid	Down 33%	Down 50%
Large	Down 20% to 73%, by earlier model	Not compared
New top tier at launch	2.5 times the tier below	Up 67% to 100% on the previous top model

One provider notes that its newer models use a tokenizer that produces about 30% more tokens for the same text, so the fall in price per piece of work is smaller than the fall in list price.

C. The sales mix

The price fall that reaches profit is the average of the newer and older tiers, weighted by revenue. The table halves the price of the older tier only and holds the newer tier.

Older models' share of revenue	Average price fall	Margin, 3-year life	Margin, 6-year life	Payback
0%	0%	35%	58%	2.3 years
20%	10%	28%	54%	2.6 years
40%	20%	19%	48%	3.0 years
60%	30%	7%	41%	3.6 years

A 20% price cut in section 3 is the same as starting from 40 cents of revenue per dollar, the low end of our range, so that table also shows the result if revenue is lower than we assume. At the high end, 60 cents, the margins would be 46% and 65% and the payback 1.9 years. The number that matters is revenue per dollar of capex for each generation of chip, not the average. Companies rarely split revenue by model generation, so outsiders can bracket the mix but not measure it.

D. Where the inputs come from

Power. A current-generation rack of 72 GPUs draws about 120 to 132 kW, according to the vendors' datasheets. At the US industrial electricity price of 9.64 cents per kWh (EIA, August 2026, preliminary) and a 30% uplift for cooling and losses, running flat out all year, that is roughly $145,000 a year per rack, or 4.1 cents per dollar of capex on a $3.5m rack. Epoch

AI's cost model assumes a lower cooling overhead and 71% utilisation, and gives about $78,000 a year, or 2.2 cents. We use 3 cents, the midpoint, with a range of 2 to 4 cents depending on use and cooling.

The cost of a rack. Nvidia does not publish a rack price, so every anchor is an analyst estimate. Epoch AI's cost model, built on SemiAnalysis inputs and published in May 2026, puts servers at $21.2m per megawatt of IT load and networking at $4.9m. At 0.132 MW per rack that is about $3.4m per rack. Analyst estimates reported in the trade press run from about $3.0m to $3.9m for a rack of this class, the top end including networking and storage. We use $3.5m, with a range of $3.1m to $3.9m. The rack price only enters the power line, so the headline results do not depend on it.

Building, power plant and cooling. The split of the dollar is read from Epoch AI's cost model for a US hyperscale site with this class of rack. It puts chips and networking at 69% of up-front cost and the building, land and utility works at 31%. We round that to 70 and 30. The same model uses a 14 year life for the building and 5 years for the IT, so we write the building off over 14 years. As a cross-check, Cushman and Wakefield's 2026 development cost guide puts the most modern new facilities at $17.6m per megawatt all in. No company states its own building, power and cooling cost per megawatt in its filings.

Revenue per dollar of capex. CoreWeave, the closest pure play on GPU capacity, reported revenue of $5.1bn on capex of $10.3bn in 2025, which is 50 cents per dollar. That flatters the picture, because much of the capex was not yet earning. On cumulative capex for 2023 to 2025 the figure is 23 cents. The large cloud companies sit higher, with Google Cloud at 64 cents and AWS at 98 cents of company-wide capex in 2025, but their revenue includes software and their capex serves every use, so they are not like for like. We use 50 cents, with 40 to 60 cents as the range and the CoreWeave cumulative figure as the downside. These ratios are our calculation from reported revenue and capex. The companies do not state them.

E. Sources and open points

Company filings: Alphabet 10-K FY2025, Amazon 10-K FY2025, Meta 10-K FY2025, Microsoft 10-K FY2026, CoreWeave 10-K FY2025, Oracle 10-Q Q1 FY2025. Power: US Energy Information Administration, Electric Power Monthly Table 5.6.A, August 2026 (preliminary, released 23 October 2026). Rack power draw: vendor datasheets from Supermicro, Gigabyte and HPE. Price trends: Epoch AI, LLM inference price trends, 12 March 2025 (epoch.ai/data-insights/llm-inference-price-trends). Rack, building and power cost: Epoch AI, AI data centre cost breakdown, 14 May 2026 (epoch.ai/data-insights/ai-datacenter-cost-breakdown), and Cushman and Wakefield, 2026 Data Center Development Cost Guide. Token prices: the two providers' own pricing pages, 29 October 2026.

Open points. The rack price, the building cost and the Cushman and Wakefield figure rest on analyst and consultancy estimates, several reported second-hand. Revenue per dollar of capex is our calculation from reported revenue and capex, and the large cloud segment figures are read from filing text, not the full statements. The 30% cooling uplift and the 4 cents of staff, maintenance and software are our own assumptions and are small. No filing discloses revenue per chip as chips age, and that is the input we would most like to see. Point-in-time analysis as of 29 October 2026.

NexSky analysis, point-in-time as of 29 October 2026. Illustrative figures are NexSky calculations from the cited sources. Company figures are from published annual and quarterly reports. Not investment advice or a solicitation.