

*Why Artificial Intelligence Is a “House of Cards”**

These thoughts are drawn from an analysis by ZoomyZoomm (thanks to Carlo for the heads-up), which I find extremely interesting and absolutely worth sharing (if you can, read to the end).

*****Why artificial intelligence is a house of cards:*****

1. You pay \$200 per year for an AI app (like Cursor).
2. Cursor pays OpenAI \$500 for API tokens (of which \$300 is VC funding).
3. OpenAI pays AWS \$1,000 for processing (of which \$500 is VC funding).
4. AWS pays \$10,000 for Nvidia GPUs.

See the problem?

Unless you, as a user, are miraculously willing to pay \$1,000 for the AI app, the only thing sustaining artificial intelligence is venture capital (VC) funding.

*****No VC funding:*****

- The AI application layer isn't profitable.
- The LLM (Large Language Model) layer isn't profitable.
- The processing layer isn't profitable.
- The GPU layer isn't sustainable.

Bulls say that the cost of inference is dropping exponentially (and it is).

This would mean that the processing cost for LLMs like OpenAI (and by extension, AI apps like Cursor) would also decrease.

But this needs to happen **before** VC funding runs out.

And it depends on how much users are willing to pay.

If end users aren't able to pay the break-even price for the entire stack—

For example:

- Nvidia GPU costs amortized over their lifecycle
- Hyperscale cloud margins
- LLM inference margins
- AI wrapper margins

This could literally amount to \$1,000 per user.

Then the whole game collapses.

Currently, people pay \$20/month for mid-tier plans and \$200/month for professional plans, e.g., Cursor.

That's \$240/year or \$2,400/year.

I'm guessing heavy users likely burn through over \$10,000 in processing costs.

So, unless you can increase costs for these users by 10x, this isn't sustainable.

I see only three viable paths for these companies:

1. LLMs must drastically reduce model/inference costs (limiting usage for AI wrappers).
2. AI application companies need to start charging *much* more overall (hard to do now due to land-grabbing competition).
3. GPU prices drop substantially (e.g., a Nvidia competitor emerges).

This would alleviate *many* downstream costs, for example:

1. Hyperscalers can lower prices.
2. LLMs can lower API prices.
3. AI app companies can break even at current prices.

Ultimately, the "black swan" I see is that deflation is coming for every aspect of technology:

1. GPU costs falling (thanks to emerging competitors).
2. Cloud costs falling (thanks to neocloud and hyperscaler competition).
3. Inference costs falling (thanks to better hardware).

This means the **final** unit economics will improve.

But **when** this happens, and what margin profile each player has when everything settles, is still unclear.

If I had to guess:

****Winners****: GPU providers and hyperscalers.

****Losers****: LLMs and AI app companies.

This is why you see the entire HF community going long on:

NVDA, MSFT, GOOG, AMZN.

Because these four companies are poised to make a ton of money in the next decade, regardless of what happens in AI.

****BUT...****

Hating consensus, I look for ways to make the same trade at 50% of the valuation.

For me, that's China Tech:

NVDA = ***SMIC***

MSFT = ***Tencent***

GOOG = ***Baidu***

AMZN = ***Alibaba***

****PLUS****, I haven't mentioned other massive bottlenecks like:

- Rare earth minerals
- Energy production
- Embedded AI scalability

China leads in all three areas and has:

- All the rare earth minerals
- All the rare earth processing
- Unlimited clean energy (nuclear)
- 99% of robotic manufacturing capacity

What I'm saying is that it's inevitable China catches up on the LLM front (either through closed models like Bytedance or open-source like Qwen, DeepSeek, etc.).

So the next question becomes: what happens *after* models plateau?

It all comes down to atoms.

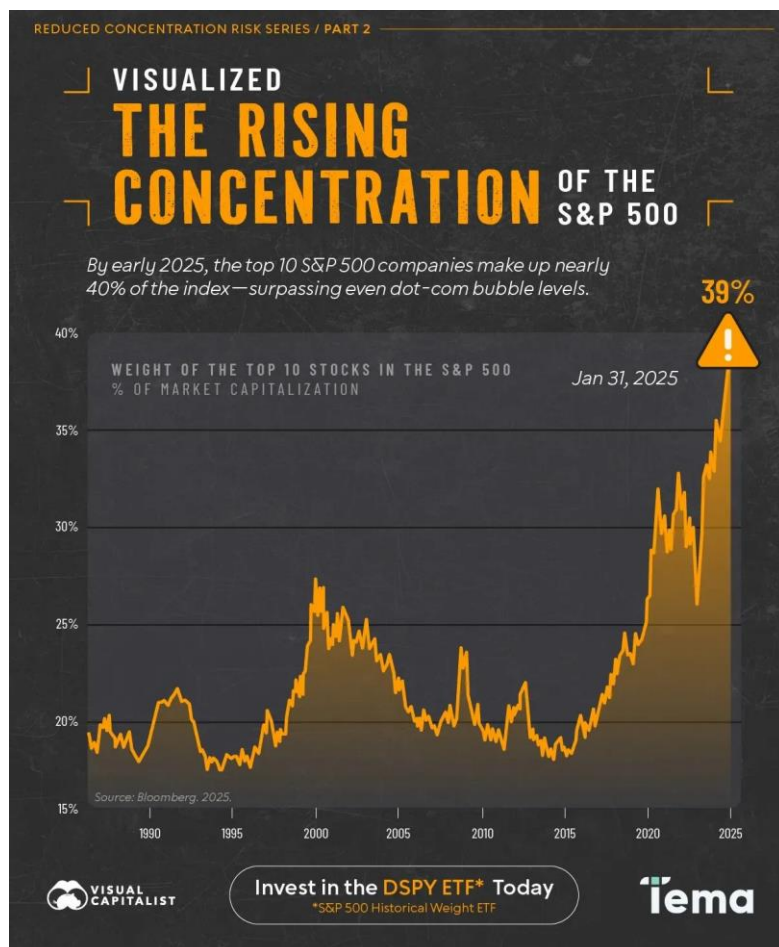
It all comes down to scale.

Here's what might be the first domino to fall in this "AI house of cards."

🕒 ****Why the Magnificent 7 are about to implode:****

1/ The wave of capital expenditure on AI by \$AMZN, \$MSFT, \$GOOG, \$META due to the rise in next-gen AI inference workloads has a significant unintended consequence the market isn't ready for.

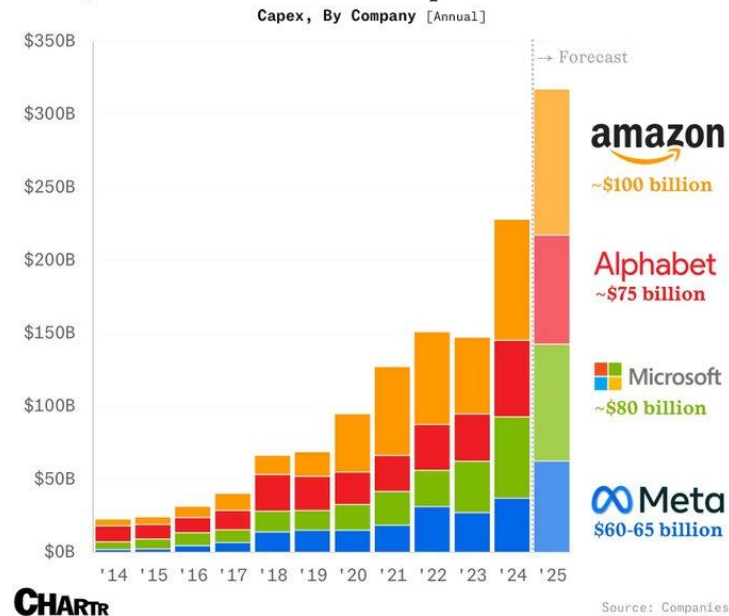
The hidden killer of AI is *value destruction* (caused by overuse and technological obsolescence).



2/ In 2025, AI capital expenditure will reach \$300 billion, funneled toward data centers and GPUs to support generative AI workloads.

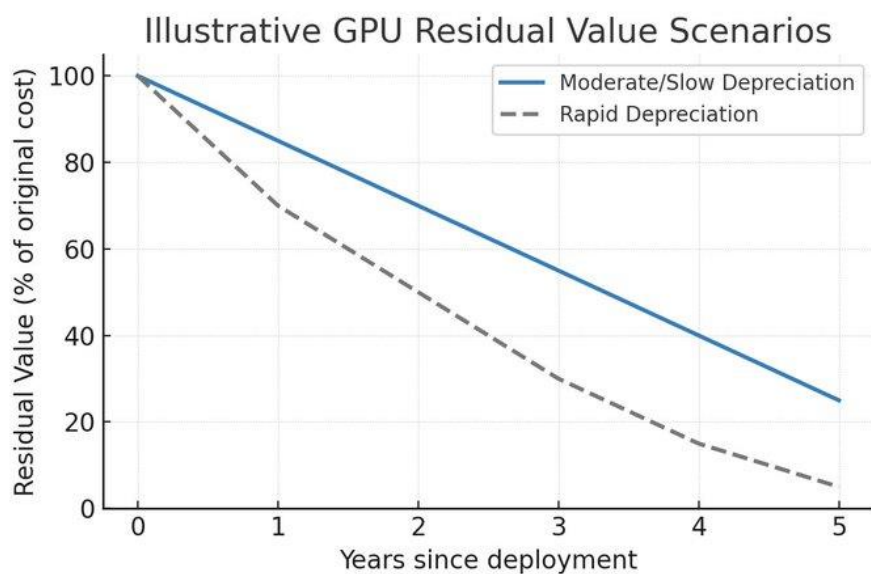
This is a rational decision given the demand for training and inference, but it overlooks a critical issue: hardware is being used at rates far beyond its original design specifications.

4 Big Tech Companies Plan On Spending \$315+ Billion On Capex This Year



3/ Traditional GPU depreciation schedules are five years, but the intense demands of next-gen AI (continuous high-load operations) cause physical wear much faster (within 2-3 years).

For example, Nvidia's shift to annual GPU architecture updates makes existing investments obsolete faster.



4/ Evidence of this discrepancy is already emerging.

\$AMZN has reduced the lifespan of AI servers from 6 years to 5 years, incurring costs of about \$920 million.

\$META's reduction in asset lifespan could increase depreciation expenses by over \$5 billion in 2026.

All of this directly impacts profits.

Company	Ticker	Nature of the change	Effect
Alphabet Inc.	GOOGL	Change in estimated useful life of servers from 4 to 6 years, and of networking equipment from 5 to 6 years	Decrease in depreciation expense of \$3.9 billion or increase in EPS of \$0.24 for the year ending December 31, 2023
AMAZON COM INC	AMZN	Change in estimated useful life of servers from 5 to 6 years, due to "continuous improvements in...hardware, software, and data center designs". Note that for Amazon this is the second change in 3 years - the company changed the depreciable life of servers from 4 to 5 years and of networking equipment from 5 to 6 years effective January 1, 2022	Decrease in depreciation expense of \$695 million or increase in EPS of \$0.07 for the three months ending on March 31, 2024. The 2022 change in estimate decreased depreciation expenses by \$3.6 billion or \$0.27 per share for the year ending December 31, 2022.
MICROSOFT CORP	MSFT	Due to increased efficiencies and advances in technology, the estimated useful lives of both server and network equipment increased from 4 to 6 years	An increase in operating income of \$3.7 billion and net income of \$3.0 billion, or \$0.40 per share for the year ending June 30, 2023
INTERNATIONAL BUSINESS MACHINES CORP	IBM	Increased the estimated useful life of servers and networking equipment from 5 to 6 years for new assets and from 3 to 4 years for used assets	Increase in income from continuing operations before income taxes of \$208 million or \$0.18 per share for the year ending December 31, 2023
Meta Platforms, Inc.	META	Increase in estimated useful lives of a majority of the servers and network assets from four years to 4.5 years, effective the second quarter of 2022, and further extension of the useful lives to 5 years effective the fourth quarter of 2022	Decrease in depreciation expenses of \$860 million or increase in EPS of \$0.26 for the year ending on December 31, 2022
ORACLE CORP	ORCL	Increased the estimated useful life of servers from 4 to 5 years	Decreased operating expenses by \$434 million for the year ending May 31, 2023

Source: Data sources - SEC filings & Audit Analytics CAE database; Analysis by Deep Quarry

5/ \$MSFT and \$META have extended depreciation periods to artificially boost short-term earnings.

But sooner or later, the chickens come home to roost.

You can't keep \$100 billion in "GPU assets" on your balance sheet that have melted and are useless. Eventually, you'll have to write them off.

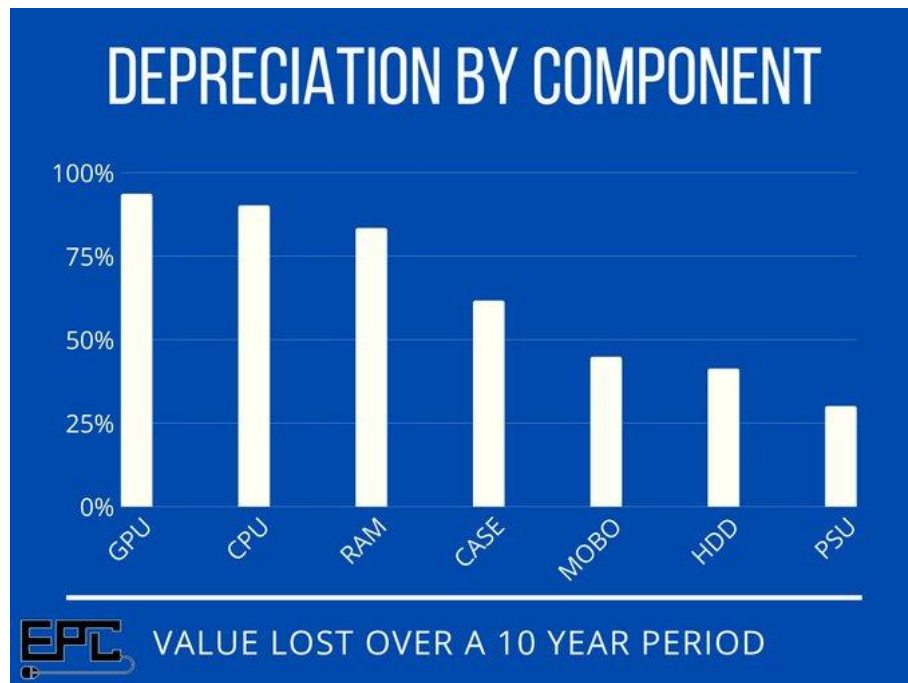
META PLATFORMS, INC.
CONDENSED CONSOLIDATED BALANCE SHEETS
(In millions)
(Unaudited)

	June 30, 2025	December 31, 2024
Assets		
Current assets:		
Cash and cash equivalents	\$ 12,005	\$ 43,889
Marketable securities	35,066	33,926
Accounts receivable, net	16,561	16,994
Prepaid expenses and other current assets	9,981	5,236
Total current assets	73,613	100,045
Non-marketable equity investments	21,988	6,070
Property and equipment, net	147,039	121,346
Operating lease right-of-use assets	15,662	14,922
Goodwill	20,654	20,654
Other assets	15,788	13,017
Total assets	\$ 294,744	\$ 276,054

6/ You can't have it both ways. Running GPUs at >100% utilization isn't without drawbacks.

Generative AI is resource-intensive.

You end up stressing hardware beyond standard operational limits, leading to faster resource depletion, lower resale value, and higher write-offs.



7/ Write-off charges would appear as a one-time hit to profitability.

It's not the write-off itself that worries me. It's the signal it will send to the market.

Faster depreciation = lower utilization = lower revenues = lower ROI.

And if they overspent on GPUs? Is the market prepared?

8/ This doesn't mean AI is a massive bubble and these stocks will crash by 80%.

They're cash cows with negative net debt and clear use cases in their verticals.

AI is real.

BUT the market will soon wake up to reality. Nothing goes up forever.

9/ In short:

- \$META, \$AMZN, \$MSFT, \$GOOG have gone all-in on AI investments.
- For now, the market is okay with this level of spending.
- While strategically sound, accelerated hardware depreciation could force \$10B+ write-offs.
- If that happens, expect a -20% sell-off.

****CONCLUSION****

Artificial Intelligence is here and will become increasingly central to the global financial landscape, but the names to invest in are not in the U.S. but in *China*.

What we're seeing is the behavior of traders unaccustomed to sustained uptrends.

That's why they sell as soon as they see a +5%.

But when they realize this rally still has legs for +20/30/50/100%, they'll shift from profit-taking to *greed* and buy increasing volumes, holding for longer periods.

Software is at risk because of AI.

E-commerce is *not* at risk because of AI.

Fintech is *not* at risk because of AI.

The U.S. is in a valuation bubble.

Latin America is *not* in a valuation bubble.

China is *not* in a valuation bubble.

Southeast Asia is *not* in a valuation bubble.

Investing is so easy, friend.

****BUY CHINA TECH****