AI: Still Early Days

Introduction

Headlines today abound with proclamations about an AI bubble. That is not surprising given how far the momentum factor had run earlier this year and especially the unprecedented amount of capital that has poured into AI data centers globally. In the short term, we think that access to and cost of capital are the biggest overhang on the group. However, our view is that the long-term opportunity for AI is intact and in some ways may be strengthened by the stricter access to capital we now see in the markets.

Corrections are commonplace in technology stocks, yet this one has been the worst in more than 25 years. The selloff has also been unusual in its speed. Goldman Sachs' U.S. High Beta Momentum basket fell about 37% in July, its worst month on record going back to 1999 (Hedgeye), and Bloomberg data put it as the worst month since November 2000 (Capitalbolsa). The Goldman Sachs High Beta Basket (GSPRHIMO) fell more than 46% from its high. More recently, we have seen healthcare AI-related companies also under significant share price pressure.

Despite this substantial volatility, we think that the long-term underpinnings of artificial intelligence revenue and profit pools are intact and growing. In this report, we : 1. showing how the industry today does not look like the types of tech busts we have seen in the past, 2. sizing what the revenue opportunity could be for AI both for consumer and enterprise applications, 3. comparing these revenue pools to the projected capital spending to see if the math pencils out and 4. identifying the key risks around financing for the AI build-out.

A look back at prior tech busts

Looking back in history, past tech busts announced themselves in significant earnings misses and in idle capacity.

- **PCs and memory, mid-to-late 1990s.** In this cycle, demand slowed while the leaders continued to add output. Worldwide PC shipments grew about 25% to just under 60 million in 1995 (Spokesman-Review/Dataquest), then only 16–18% to 68–71 million in 1996 (Tech Monitor/IDC, Dataquest). The larger vendors grew faster than the market in 1996, with IBM up about 27%, Compaq up 19–22% and HP up about 50%, while Dell's units rose 48% and then 55% in its fiscal 1996 and 1997 (Dell 10-K). That share came at the expense of weaker players: Packard Bell's shipments fell 3–5% and Apple's fell 23–

31% (Tech Monitor), while AST lost $99 million in fiscal 1995 before Samsung bought 46% of it and later the rest (Wikipedia). Ultimately, a Compaq-led price war drove IBM's PC unit to a $992 million pre-tax loss in 1998 (Tech Monitor). In short, too many players planned for a combined market share of more than 100%, which resulted in an overbuild.

- **Telecom, 2000–2002.** At the trough, only about 2–5% of installed fiber was lit (FDIC). Lucent and Nortel revenue fell 18% and 17%, respectively, in a single quarter of 2001, and Corning shut down fiber production entirely (EE Times).

The common thread in each case was that prices collapsed because supply far exceeded what customers were willing to use.

Chips and equipment, 1996–1998

The 1996 memory crash turned into a broader chip and equipment recession that did not bottom until 1998. Fabs ran short of work, and the toolmakers swung to losses.

Chipmaker	1996	Trough (1998)
TSMC	Utilization of 96% for the year, falling to 85% in Q4 (TSMC)	58% in Q3 1998 and 74% for the year; gross margin fell to 33.8% in Q4 (TSMC)
Micron	4Meg DRAM price fell 78% in fiscal 1996; net income dropped to $593M from $844M (SEC 10-K). Halted its $2.5B Lehi, Utah fab mid-build (Spokesman-Review)	Gross margin fell from 39.8% to 9.3% in fiscal 1998 after price drops of 75% and 60% (SEC 10-K); $48M net loss in its fiscal Q2 1998 (SEC 10-Q)
Intel	Kept building through the slump: announced Fort Worth Fab 16 in November 1996 (Intel)	Raised capex from $4.5B in 1997 to a planned $5.3B in 1998, while warning capacity could be underused (Intel 1997 annual report). Rival AMD's Fab 25 ran at about 80% and AMD lost money in 1996, 1997 and 1998 (AMD)

DRAM makers typically kept their fabs running and let prices absorb the glut, which is why Micron's pain points showed up in margins rather than in idle lines.

The equipment makers took the second hit:

Toolmaker	Before	Trough
Applied Materials	$498M net income in fiscal 1997	Fiscal 1998 net income fell to $231M. Q4 1998 sales dropped 47% and it lost $186.7M after restructuring charges; ongoing margin was 3.9% (Applied Materials). It never posted an annual loss
Lam Research	$38.6M net income in a single quarter of fiscal 1996 (EE Times)	Three straight annual losses: $30.7M in fiscal 1997, $144.6M in fiscal 1998 (SEC 8-K), and $54.5M in fiscal 1999 as revenue fell to $648M from $1.05B (Lam)

Today's memory market is, in our view, the mirror image: Micron says it can meet only half to two-thirds of DRAM demand (Tom's Hardware). We see new capacity coming to the market, but at a much more measured pace due to larger die sizes in next-generation DRAM. Orders for semiconductor equipment companies remain high, while spares and services reflect high utilization across the semiconductor manufacturers.

The revenue case: LLMs are already earning

LLM revenue is compounding faster than any forecast predicted, which is a key reason we view the spend as early rather than excessive. The two leading labs have gone from about $10 billion of combined run-rate at the end of 2025 to more than $100 billion today:

- **Anthropic:** run-rate topped $65 billion at the end of July, up from about $9 billion at the end of 2025, with Q2 revenue above $11.5 billion (Axios). It reported positive adjusted operating income (Sovereign).
- **OpenAI:** about $50 billion of run-rate at the end of September, targeting $70 billion or more by year end (Citybiz/Bloomberg).

Where the money comes from:

Use	What it earns today	Why it scales
Coding agents	Cursor at about $4B run-rate by mid-2026 (KuCoin/a16z); Claude Code above $2.5B by February (Pasquale Pillitteri)	Priced against developer salaries, not software seats
Enterprise API and agents	About 80% of Anthropic's revenue is enterprise; OpenAI is on track for more than half (Axios)	Gartner expects agentic workflows to reach 50% of GenAI model revenue by 2030, from 5% in 2025 (Software Strategies)
Consumer subscriptions	ChatGPT remains OpenAI's largest consumer line; the bigger prize is below	Ads, commerce and new tiers are still early
Cloud resale	Hyperscaler backlogs of about $2 trillion, above	Every lab's usage runs through rented compute

Coding: the first market at scale

Coding is where LLM revenue is already proven, in our view. Claude Code reached a $1 billion run-rate within six months of launch and more than $2.5 billion by February 2026 (VentureBeat). It is also the clearest example of pricing against labor rather than software seats, since an agent that does part of a developer's work can charge thousands of dollars a year rather than tens.

Pool	Basis	Size per year
U.S. developer wages	~1.7 million software developers at a $135,980 median wage (BLS)	~$230B
U.S. LLM TAM at 25%	25% of that wage bill	~$58B
Developer pay excluding China	~29.5 million professional developers (36.5 million per SlashData via Voxbooster, less ~7 million in China per Keyhole Software) at ~$40–60k blended pay (our assumption)	~$1.2–1.8T

Pool	Basis	Size per year
LLM TAM at 25%, excluding China	25% of that pay pool	~$295–440B

It is worth noting that these figures count only professional developers. We expect the pool to grow further as non-programmers build software with agents, which is another example of the applications we have not yet sized.

We recently listened to a Blackstone presentation where the company presented its research on Anthropic spend at its private portfolio companies. Blackstone surveyed approximately 1,400 of its portfolio companies from a variety of different industries and their Anthropic spend grew from $25M in September of 2025 to $525M in September of 2026 because "portfolio companies are paying for more AI because it pays them back." (Jon Gray, Blackstone)

The consumer prize: from tutors to idle cash

Today fewer than one in ten Americans pay for an AI subscription (Epoch AI). We anticipate that there is a long runway for growth here given what we know about increased utility of the service. An AI agent that acts on behalf of a household can take on paid services and, much more significantly, recover the money households lose through inertia. Below are some uses, ranked from smallest to largest in annual U.S. dollars; there are likely many more not included here:

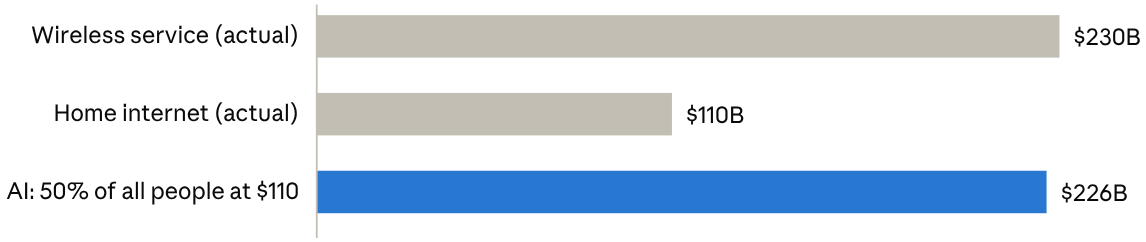
Pool	Size per year	What an agent does
Personal training	~$12B industry revenue (IBISWorld)	Plans, coaches and adjusts workouts
Interior design (residential)	~$11–17B; residential is about 41% of a $26.5–40.7B U.S. market (Intdesigners, citing IBISWorld and Credence Research)	Plans layouts, visualizes rooms and sources furniture
Tutoring and test prep	~$15–20B; $18.9B including driving schools (IBISWorld)	One-on-one teaching for students, on demand

Pool	Size per year	What an agent does
Home insurance overpayment	~$40–60B (our estimate: ~25% savings from switching on a ~$1,900 average premium, per Texas OPIC and Gerald)	Re-shops the policy at every renewal
Unused subscriptions	($252 per adult wasted, per CNET via WCPO)	Finds and cancels what goes unused
Auto insurance overpayment	~$80B (our estimate: ~$416 per driver not shopping, per NerdWallet, across ~200M drivers)	Compares quotes and switches carriers
Home care for the elderly	~$135–200B; services are about 52% of a $258B U.S. home healthcare market in 2026 (Precedence Research), and ~4.3 million home health and personal care aides earn a $35,800 median (BLS via Wagedex), or ~$200B with benefits	Coordinates care and schedules, monitors and reminds, keeps company, and supports aides
Forgone savings interest	~$50–400B (Ally's ~$50B per NGPF to Raisin's ~$396B per Yahoo Finance)	Moves idle cash to a high-yield account

Consumer LLM spending is best sized like a connectivity subscription, which avoids double counting across the use cases above. As benchmarks, Americans spend about $230 billion a year on cellular service (T-Mobile, Verizon, AT&T), roughly $50 per phone line a month (TelecomLead), and about $110 billion on home internet, roughly $80 per household a month (BroadbandNow). We assume 50% of the U.S. population, about 172 million people, pays $110 a month for AI. That is a U.S. consumer market of about $226 billion a year, nearly the size of the wireless business and well above home internet.

A U.S. consumer AI market of about $226B a year would rival wireless service

Annual U.S. consumer spend, $B



Wireless: 2025 carrier filings (T-Mobile, Verizon, AT&T) plus smaller carriers, our estimate · Home internet: BroadbandNow 2026 average × ~115M broadband homes, our estimate · AI: 50% of ~343M Americans at $110 a month, our assumption

As with wireless, usage can grow much faster than the bill: U.S. wireless data use rose 20% in 2025 while nominal prices fell more than 4% (CTIA via Telco Magazine).

The other key opportunity we see here is that these applications are sized on current users. If the LLMs are priced as we lay out, there will be new entrants that now can "afford" to use these services once out of reach, ranging from personal training to tutoring. We expect that the affordability factor will drive engagement and expand the total TAM of many of these markets.

LLMs providers also have the opportunity to sell advertising as engagement increases, which represents a sizable monetization pool in addition to paid subscription rates.

The small-business prize: 15% of four line items

The U.S. has 36.2 million small businesses, about 6.4 million of which have employees (SBA Advocacy, Big Ideas for Small Business). Very few have a marketing team, an inventory planner, a treasurer or a collections desk. If LLMs capture 15% of what these four functions cost as revenue, we estimate the TAM at roughly $75–170 billion a year:

Spend area	Annual U.S. SMB spend	LLM TAM at 15%	What an LLM does
Advertising	~$98B digital, for firms under 500 employees (TechBullion); up to ~$640B on Intuit's broader extrapolation (QuickBooks)	~$15–96B	Writes the ads, picks the channels, reallocates budget daily

Spend area	Annual U.S. SMB spend	LLM TAM at 15%	What an LLM does
Inventory management	~$290B carrying cost (our estimate: ~$670B economy-wide at a 25% carrying rate on $2.68T of inventory (Manufacturing Lead Gen), times small firms' 43.5% share of GDP)	~$44B	Forecasts demand, sets reorder points; 39% of small firms still track stock manually or not at all
Idle cash	~$40–155B of forgone interest (our estimate: $3.87T sits in noninterest-bearing deposits (FDIC via FRED); business checking pays ~0% vs. 3.5%+ in business savings (Beancount); low end assumes only a quarter is movable reserve cash)	~$6–23B	Sweeps excess cash to the best insured yield daily and moves it back before payroll
Billing and collections	~$58B (our estimate: ~$9,000 a year per firm chasing overdue invoices (SchedulingKit) across ~6.4M employer firms)	~$9B	Sends invoices, follows up, reconciles; small firms are owed ~$825B in unpaid invoices at any time

That 15% slice alone, before consumers or large enterprises, is more than the two leading labs' combined revenue today. We view idle cash as the cleanest case, since the gain is measurable to the dollar and a share-of-yield fee is easy to charge. The advertising figures are the most uncertain, as survey-based extrapolations run well above platform-measured spend.

These sizings only count how LLMs make existing work more efficient for businesses and consumers, and that alone can make the math work. We recognize they are limited by our imagination: the new applications nobody has thought of yet are what expand the profit pool, and they are not in these numbers.

Life sciences: drug discovery and clinical trials

We are seeing the positive impact of LLMs in healthcare today, whether with big drug companies or startups that may work without even having a wet lab. Drug development is a large and labor-intensive area of spend where LLMs can read the literature, design and screen molecules, write protocols, identify patients and draft regulatory filings. Global pharma R&D was $276 billion in 2021 (Chandra et al. via R&D World) and is heading past $300 billion (Statista). In 2024, almost 15% of it went to preclinical work and about 45% to clinical trials (Statista).

Pool	Basis	Market size per year	LLM revenue at 10%	Excluding China
Drug discovery	15% of $300B R&D ($45B) to the $78.6B drug discovery market in 2026 (Towards Healthcare via BioSpace)	~$45–79B	~$4.5–8B	~$4–7B
Clinical trials	$91.7B clinical trials market in 2026 (Precedence Research via BioSpace) to 45% of $300B R&D ($135B)	~$92–135B	~$9–13.5B	~$8–12B
Total		**~$137–214B**	**~$14–21B**	**~$12–19B**

We assume China is about 10% of this spend (our placeholder; Asia-Pacific is about 15% of global R&D). These pools look small next to coding or customer support because they count only what is spent on discovery and trials. In our view, the bigger opportunity is the value of faster and cheaper approvals: a drug that reaches market a year earlier earns an additional year of patent-protected sales, which AI developers could capture through royalties, milestones or their own pipelines. That upside is not reflected in these numbers.

Logistics: 10% of a $2.1 trillion cost base

U.S. businesses spend about $2.4 trillion a year on logistics (CSCMP via SCDigest). Most of it is transportation, followed by the cost of carrying inventory and the forwarding, IT and services that coordinate it all. These are decision-heavy functions: dispatch, routing, load matching, carrier procurement and reorder timing. Agents are already moving in: FedEx aims to have AI agents in more than half of its core operational workflows by 2028 (PYMNTS/WSJ). Excluding the small-business inventory already counted above, the pool is about $2.1 trillion. If AI captures 10% of it, that is about $210 billion a year in the U.S.:

Spend area	Annual U.S. spend	AI TAM at 10%
Transportation: dispatch, routing, load matching, procurement	~$1.48T	~$148B
Inventory carrying, larger firms	$440B ($730B less ~$290B counted under small business)	~$44B
Coordination: forwarding, IT and services	~$187B	~$19B
Total	**~$2.1T**	**~$211B**

Transportation is the largest piece and the most aggressive assumption, since AI earns its share by cutting empty miles and pricing loads better rather than by moving goods itself.

Robotics: an incremental source of upside – this market is now inflecting

So far, artificial intelligence has mostly lived on a screen. We see an explosion of agentic usage continuing to drive demand - this has been especially important in 2026. The next big thing we see is physical AI with robotics leading the way. As robots become more prevalent in the coming years, we expect them to add another layer of demand for model use, and with it the need for more compute infrastructure and more compute at the edge. We are seeing this with autonomous vehicles and robotics in factories - we see this market gaining visible momentum now.

We view robots as an incremental source of LLM revenue that has yet to be appreciated. Today, the world builds roughly 1 million a year: more than 600,000 industrial robots installed in 2025 (IFR), almost 250,000 professional service robots (IFR), and humanoid shipments expected above 50,000 in 2026 (Counterpoint). We also think that the definition of "robotics" ought to include stationary analysis such as vision for everything from toll collecting to manufacturing monitoring. These applications also have a critical need for AI. In our bull case, about 70 million robots are in use by 2030, with each robot paying $300 a year for LLM functionality:

Year	Robots in use	Annual LLM revenue
2030	~70M	~$21B
2031	~108M	~$32B

Year	Robots in use	Annual LLM revenue
2032	~146M	~$44B
2035	~260M	~$78B
Steady state (7–10 year robot life)	~270–380M	~$80–114B

This assumes a ramp of about 4M, 9.5M and 19M units in 2027–2029, then about 38M a year thereafter. That is nearly 40 times today's output and about 40% of global car production, so we treat it as upside rather than a base case. There are two caveats. Chinese vendors shipped more than 97% of humanoids in the first half of 2026 (TechRepublic), so some of this pool may be out of reach for U.S. model vendors. In addition, some of the fee may go to on-device compute rather than cloud models, which would still support the edge component opportunity we describe later in this report. However, what this analysis does not include is the upside we are seeing today from things like sensing and vision enhancement, all of which benefit from AI. SpaceX, for example, has given great insight into what can be done from space.

Does the math close? The revenue pool against the build

To answer that question, we put every pool together, excluding China, and compared it with $5–6 trillion of AI capex we envision by 2030. We size each pool as it would stand in 2030, the window over which this build has to pay for itself. Our capture assumptions are 10% of the government, life sciences and logistics pools, 15% of small-business, legal, accounting and healthcare administration, and 35% of coding and customer support, on the view that by 2030 agents do much of this work rather than assist with it. Consumers are sized as subscriptions: 50% of the U.S. population at $110 a month. Beyond those named functions, we size the rest of white-collar work by occupation from BLS pay data: 35% of office and administrative support, which agents can largely run, and 15% of business and financial operations, management, sales and engineering and science roles. We also add advertising inside AI assistants and tutoring. We scale the U.S. consumer, small-business, enterprise knowledge-work and logistics pools to the world outside China by 2.0–3.2x (the U.S. share of ad spend at the low end and of GDP at the high end), then grow every pool to 2030 at 5% a year in nominal terms, roughly the pace of world nominal GDP, from its base year (2023 for the BLS pay data, today for the rest). The generative search ad and robot figures are already 2030 forecasts, so we leave them as they are.

Pool	Basis	Capture	LLM revenue, 2030 ($B/yr)
Consumer			**~550–881**
Subscriptions	50% of ~343M Americans at $110 a month (~$226B U.S.), scaled	50% paying	~550–881
Small and medium businesses			**~179–671**
Advertising	U.S. ~$98–640B, scaled	15%	~36–374
Inventory management	U.S. ~$290B carrying cost, scaled	15%	~105–170
Idle cash	U.S. ~$40–155B forgone interest, scaled	15%	~15–92
Billing and collections	U.S. ~$58B, scaled	15%	~23–35
Enterprise labor			**~913–1,236**
Coding	~$1.2–1.8T developer pay	35%	~502–754
Legal	$860B–1.05T global market (IMARC, Ken Research), less ~5% China	15%	~150–182
Accounting	$654B global market (IBISWorld), less ~10% China	15%	~108
Customer support	~$400–500B global call center market (Ringly), less ~10% China	35%	~153–192
Government administration			**~114**
U.S.	$170B executive and legislative (FRED), $103B state and local tax collection (FRED), $70B state and local courts (FRED)	10%	~41

Pool	Basis	Capture	LLM revenue, 2030 ($B/yr)
EU	3.1% of GDP on administration, fiscal affairs and general services, ~$600B (Eurostat)	10%	~73
Life sciences			**~15–24**
Drug discovery	~$45–79B, less ~10% China	10%	~5–9
Clinical trials	~$92–135B, less ~10% China	10%	~10–15
Other enterprise knowledge work			**~1,979–3,167**
Office and administrative support	U.S. ~$690B pay: 18.5M workers at ~$48K (BLS), less customer service reps and bookkeeping clerks; includes insurance and bank operations clerks; scaled	35%	~684–1,093
Business and financial operations	U.S. ~$780B pay (BLS), less accountants and auditors; includes claims adjusters, underwriters and loan officers; scaled	15%	~331–530
Management	U.S. ~$1.45T pay: 10.5M managers at ~$138K (BLS), scaled	15%	~610–977
Sales, excluding retail	U.S. ~$465B pay (BLS), less retail salespeople and cashiers, scaled	15%	~196–314
Engineering and science	U.S. ~$375B pay (BLS), scaled	15%	~158–253
Healthcare administration			**~66–91**
Billing and insurance administration	U.S. $361–496B a year (CAP via Harvard, CAP via Fiscal Times), U.S. only	15%	~66–91

Pool	Basis	Capture	LLM revenue, 2030 ($B/yr)
Advertising in AI assistants			**~100**
Generative search ads	Forecast to grow from $5.1B in 2026 to over $100B by 2030 (WPP Media via Carly)	Direct	~100
Education			**~30**
Tutoring	$131–134B global market (Voxbooster), less ~25% China	25%	~30
Logistics			**~512–820**
Coordination: forwarding, IT and services	U.S. ~$187B, 7.8% of $2.40T business logistics cost (CSCMP via SCDigest), scaled	10%	~45–73
Transportation: dispatch, routing, load matching, procurement	U.S. ~$1.48T transportation spend (CSCMP via SCDigest), scaled; fee on better decisions, not the cost of moving goods	10%	~360–576
Inventory carrying, larger firms	U.S. ~$440B: ~$730B carrying cost (CSCMP via SCDigest) less the ~$290B in the small-business row, scaled	10%	~107–171
Robots (bull case)	~70M in use by 2030, ~38M built a year after, $300 each, 7–10 year life		**~80–114**
Total			**~$4.54–7.25T**

Our view is that this forecast of $4.5–7.25T is likely to be low given how many applications will emerge that we cannot imagine at this point.

We think the right test is what the infrastructure has to earn, not just what it costs. Not all of the build wears out at the same pace. We assume about 35% of capex goes to the shell (land, power, cooling and buildings), with a 20-year life, and 65% to IT equipment (GPUs, servers and networking), with a 5.5-year life. That gives a blended life of about 7.4 years, so $2 trillion of capex depreciates by about $271 billion a year. For the build to earn its cost of capital, which we assume is 12%, return on invested capital has to equal that rate: $240 billion of after-tax operating profit. At a 17% cash tax rate, that is about $289 billion before tax, so the build needs about $561 billion of EBITDA, or roughly $1.12 trillion of revenue at a 50% EBITDA margin. The chip, cloud and data center owners, however, collect only part of what end customers pay for AI. We assume 30%, with the rest going to model developers, applications and services. End-market AI revenue therefore has to reach about $3.7 trillion a year for the infrastructure to earn its cost of capital, or about $4.7 trillion if investors demand a 600bp spread above it.

Against that bar, the pools above total $4.5–7.2 trillion by 2030. At a 30% share, infrastructure would collect $1.36–2.17 trillion. That covers the $3.7 trillion bar about 1.2 times at the low end and 1.9 times at the high end. Put as returns, the low end earns about 17% after tax on the capital, above a 12% cost of capital, and the high end about 34%. In other words, with agents doing more of the work and the pools growing with the economy, the build pays for itself by 2030 across the whole range, and that still excludes applications nobody has built yet.

How the cost of capital moves the bar

The infrastructure pencils out on an EBITDA basis, but EBITDA leaves out the capital it took to build it. The return the build has to earn depends on the cost of capital, and that is moving against it (see the credit section below). To test it, we solve for the revenue at which return on invested capital equals the cost of capital. Each percentage point on $2 trillion adds $20 billion of required after-tax profit. Grossed up for tax, that is about $24 billion more EBITDA, $48 billion more infrastructure revenue at a 50% margin and, at a 30% share, about $161 billion more end-market AI revenue a year. That is only about 4% of the requirement per point, because depreciation, about half of what the build has to earn, does not move with rates:

Cost of capital	Required after-tax profit	EBITDA needed	Infrastructure revenue	End-market AI revenue
8%	$160B	$464B	$0.93T	$3.09T

Cost of capital	Required after-tax profit	EBITDA needed	Infrastructure revenue	End-market AI revenue
9%	$180B	$488B	$0.98T	$3.25T
10%	$200B	$512B	$1.02T	$3.42T
11%	$220B	$536B	$1.07T	$3.58T
12% (base)	**$240B**	**$561B**	**$1.12T**	**$3.74T**
13%	$260B	$585B	$1.17T	$3.90T
14%	$280B	$609B	$1.22T	$4.06T
15% (base + 300bp)	$300B	$633B	$1.27T	$4.22T
16%	$320B	$657B	$1.31T	$4.38T
17%	$340B	$681B	$1.36T	$4.54T
18% (base + 600bp)	$360B	$705B	$1.41T	$4.70T

Assumes $2 trillion of capex, 35% of it shell depreciated over 20 years and 65% IT equipment over 5.5 years (a depreciation-weighted blended life of about 7.4 years, shorter and more conservative than the 10.6-year simple average, or ~$271 billion a year of depreciation), a 17% cash tax rate, a 50% EBITDA margin and a 30% infrastructure share of end-market revenue. Required profit is after tax, so return on invested capital equals the cost of capital.

In our experience, investment committees typically look for a spread of 300–600bp over the cost of capital before they commit, which puts the hurdle at 15–18%. At 15%, the bar is about $4.2 trillion of end-market revenue, which even our $4.5 trillion low case clears. At 18%, the bar rises to about $4.7 trillion: the low case falls short by about $0.2 trillion, while the midpoint of about $5.9 trillion and the high case of $7.2 trillion clear it comfortably. The low case breaks even at a cost of capital of about 17%, the midpoint at about 25% and the high case at about 34%, which leaves investors room to outperform in all but the most conservative scenario.

If only the debt-funded share reprices, the effect is smaller. Data center debt of roughly $400 billion against about $1.1 trillion of 2026 capex suggests about 35% of the build is borrowed. After the tax shield on interest, each point of higher borrowing cost lifts the

blended cost of capital by about 0.3 points, adding about $7 billion of required pre-tax profit, or roughly $47 billion of end-market revenue. The pressure falls hardest on the leveraged builders now paying 9–15%: moving the whole hurdle from 12% to 15% adds about $480 billion to the end-market revenue the build needs, and moving it to 18% adds about $960 billion.

What if the chips wear out faster? The sharper bear case is the life of the chips, not the cost of capital: critics argue GPUs last three to four years, not 5.5. If all IT equipment lasted four years, depreciation would rise from about $271 billion to $360 billion a year, lifting the bar about 16% to roughly $4.3 trillion of end-market revenue. Our low case would still clear it, earning about 13% after tax against a 12% cost of capital, and the high case about 30%.

So far the evidence runs the other way. Oracle said on its latest earnings call that it released GPUs coming off contract at about a 20% premium, most of them four or more years old, with utilization at 97.9% (call highlights). Amazon, the most cautious of the large clouds, depreciates a subset of its servers over five years, down from six (Hudson Labs).

Why 50% EBITDA margins are credible. Hyperscalers do not report margins on compute, but the pure-play neoclouds do, and we think they are a conservative proxy. CoreWeave reported a 59% adjusted EBITDA margin in Q2 2026 (MarketBeat), and Nebius's AI cloud segment 49.7% (Trending Topics). These are company-wide figures, so they already absorb ramp costs and central overhead, though CoreWeave's adds back stock-based compensation worth about 4–5 points (Daloopa).

The neocloud cost structure is leaner, with little R&D, sales or software spend and take-or-pay contracts that keep utilization high. Hyperscalers offset that in three ways. They charge more: on-demand H100 list prices at AWS, Azure and Google run about $6.90–12.30 a GPU-hour, against $4.00–4.50 at Lambda and Nebius (IntuitionLabs). They often source cheaper power. And they own most of their facilities, so that cost sits in depreciation, outside EBITDA, while a neocloud's colocation rent sits inside it.

The hyperscaler data points agree. Oracle targets 30–40% gross margins on its AI data centers over the life of a contract (TIKR), a figure struck after depreciation; adding back GPU depreciation of even 20–25% of revenue implies, by our estimate, an EBITDA-equivalent of about 50–65%. AWS earned about 57% on an EBITDA basis in Q1 2026: $14.2 billion of operating income plus $7.3 billion of depreciation on $37.6 billion of revenue (Hudson Labs). On that basis we think 50% is credible and conservative.

Why the spending continues. The capex forecast rests on new capacity still earning its keep, and today it does: prices are rising while the cost of each unit of compute falls. One-year H100 contract rates rose about 40% between October 2025 and March 2026, from $1.70 to $2.35 a GPU-hour (SemiAnalysis via Crypto Briefing). CoreWeave raised prices about 25% across SKUs in July and ended Q2 with a $104 billion backlog (MarketBeat), and Nebius lifted its on-demand H100 rate from $3.85 to $4.50 on October 1 (IntuitionLabs).

At the same time, each dollar buys more sellable compute. Blackwell delivers roughly 3–4 times Hopper's per-GPU inference throughput in MLPerf (NVIDIA), and Trainium3 is 30–40% more price-performant than Trainium2 (Amazon, via DEV). Microsoft raised inference throughput for its most-used Copilot models 40% in a single quarter through software and hardware work (Microsoft).

Adoption is broadening too. Meta's Muse agent, launched September 8, reached about 1.8 million daily users within a month (Citi via KuCoin) and more than 10.4 million downloads by October 16 (BNP Paribas via Let's Data Science), though monetization is still unproven.

Having said all this, the caveats matter. This is end-market revenue rather than profit, and we assume only 30% of it reaches the chip, cloud and data center owners. Capture rates of 35% assume agents do most of the work rather than simply assist with it, and the 2030 figures assume spending on these functions grows with the economy rather than shrinking as AI makes the work cheaper. The China shares for legal, accounting and support are our placeholders, and the government figures represent total spending in each function and exclude U.S. federal tax collection, federal courts and non-EU Europe. The enterprise rows use U.S. May 2023 occupational pay scaled abroad by the same 2.0–3.2x, which likely overstates pay outside the U.S.; they also overlap modestly with the government, healthcare and billing rows, since some clerical staff appear in both.

We think the key point here is that we have listed just a few of the applications we think are within reach , and we know well that we cannot foresee what new applications will emerge about as new problems are solved for the first time.

Today: sold-out capacity and rising prices for old chips

The other key point is: because the demand for intelligence is already so large, capacity utilization remains at very high levels. In our view, AI compute shows the opposite of a glut: prices are rising, even for prior-generation GPUs.

- **Older chips are getting more expensive.** One-year H100 contract pricing rose nearly 40%, from $1.70 to $2.35 per GPU-hour between October 2025 and March 2026, with on-demand capacity sold out across all GPU types (SemiAnalysis via Tiger).
- **Price hikes are still coming.** Nebius raised its on-demand H100 rate from $3.85 to $4.50 per GPU-hour on October 1, 2026 (Uvik GPU Price Index). CoreWeave lifted list prices about 25% across SKUs in July (Spendark). Tesla also cited rising rental prices for older-generation GPUs.
- **The H100 still has a job.** Rental rates are rising across Blackwell, Hopper and even A100, as older GPUs move to inference (UOB).

In an overbuild, the last-generation asset is typically the first to be dumped. Here, it is being bid up, which is not what we would expect to see if there were excess capacity.

Capacity is mostly fungible

We are also monitoring the risk of non-fungible capacity, a dynamic we have seen in prior cycles. Nvidia GPU capacity is highly fungible but custom chips may be less so. Amazon and Microsoft have made big commitments to the frontier labs to earmark capacity. We are not worried today about the ability of Anthropic or OpenAI to move forward on this capacity - but should open models gain share, this could create some pain points. However, we do note that even if a cluster built for training is not used, it could be used to serve many inference customers.

Bottom line: the facilities can generate cash flow because rental rates today justify the returns, so we expect capacity utilization to remain intact. We think the industry's health and growth depend on the cost of intelligence coming down - this has always been the case for any technological innovation - while the functionality continues to grow. The question is: what is the cost of capital for each one of these AI factories?

The real problem is credit, not capacity

The cost of capital for leveraged builders has risen sharply. While this does not affect business that is already locked in, it will determine whether the next project pencils out.

The contrast is clear within a single year:

Deal	Date	Cost of debt
TeraWulf $3.2B senior secured notes due 2030, after its Google-backed Fluidstack leases (announced Aug 14, 2025)	Oct 2025	7.75% coupon, priced at par (7.75% yield at issue)
Zenith Arc data center green bonds	Aug → Oct 2026	~9% → 11.3% yield
CoreWeave 9.625% notes due 2032, secondary market	Oct 14, 2026	~15% yield

The cleanest test, in our view, is to hold the tenant constant. Even the strongest names are now paying about 100–150bp more than a year ago:

Tenant	Then	Now	Change
Meta	Hyperion, Oct 2025: 6.58%	Sopaipilla, Jul 2026: 7.534%	~+95bp
Microsoft	QTS Fayetteville, Apr 2026: 5.70% (due 2036)	QTS Odyssey, Aug 2026: 7.228% (due 2031)	~+150bp, on a shorter bond

Meanwhile, builders whose leases rest on weaker tenants are now paying 9–10%:

Priced	Deal	Tenant	Coupon
Jul 23	Galaxy Helios II, $3.5B secured notes due 2031	CoreWeave	9.875%
Aug 13	Zenith Arc, $2.25B	Jane Street	8.875%
Sep 23	Digital Drive, $1.1B five-year notes	CoreWeave	9.25%

CoreWeave's credit-default swaps moved above 1,000bp on October 15 (ROIC). The base rate has moved as well, with the 10-year Treasury reaching about 5.17% in late September, its highest level since 2007 (CNBC via GoLocalProv).

Where debt still looks cheap, the cost has simply shifted to equity. Cipher priced $850 million of 0% convertibles this week, but its stock fell 12% on the dilution (TIKR).

We are not worried about current plans at the leveraged builders. Rather, we question whether new builds can earn a reasonable return at today's rates:

- **CoreWeave:** about $52 billion of debt and leases against $35–39 billion of 2026 capex (InsiderFinance).
- **Nebius:** about $10 billion of convertibles by March, then a $4.5 billion raise in August, its third in a year (SEC 6-K, Motley Fool).
- **Oracle:** about $130 billion of debt, negative $24 billion of free cash flow in fiscal 2026, and an S&P cut to BBB- in July (I amsterdam).
- **The OpenAI and xAI chains:** OpenAI's partners carry nearly $100 billion of debt tied to its sites (Irish Times/FT). An xAI chip vehicle raised about $12.5 billion of debt for Colossus 2 (Vested).

What tighter credit could mean for AI capex

Data center debt totaled roughly $400 billion over the trailing twelve months, and the expectation had been for at least as much in 2027. If neocloud and other non-hyperscale borrowing dries up, we estimate AI capex could come in about $200 billion below what investors expected a month ago for 2027, and about $300 billion below in 2028:

Year	Expected a month ago	Without non-hyperscale debt	Cut
2027	~$1.3T	~$1.1T	~$200B
2028	~$1.6T	~$1.3T	~$300B

We arrive at these figures by assuming no increase in capex from the builders with the highest cost of capital. In that case, debt issuance turns from a modest source of growth into a potentially material headwind to the year-over-year dollar change, and 2027 spending would be roughly flat with 2026's ~$1.1 trillion. These are rough estimates. In our view, this would slow the pace of the build rather than reverse it.

The core of the build is funded by cash engines

The bulk of AI spending, however, sits with companies whose operating cash flow covers it and whose credit the market still prices as safe.

Amazon, Microsoft, Alphabet and Meta plan about $725 billion of capex in 2026, up 77% from 2025 (AI Weekly/FT). That has squeezed free cash flow close to zero by choice, not through distress:

Company	2026 capex plan	Cash flow behind it
Microsoft	~$190B	Fiscal 2026 operating cash flow of $183B vs. $116B of capex (TIKR)
Amazon	~$200B	Trailing operating cash flow of $161.4B (TIKR)
Alphabet	$180–190B	Cloud revenue up 63% in Q1 (The Next Web)
Meta	$125–145B	Funds its own buildout plus off-balance-sheet JVs

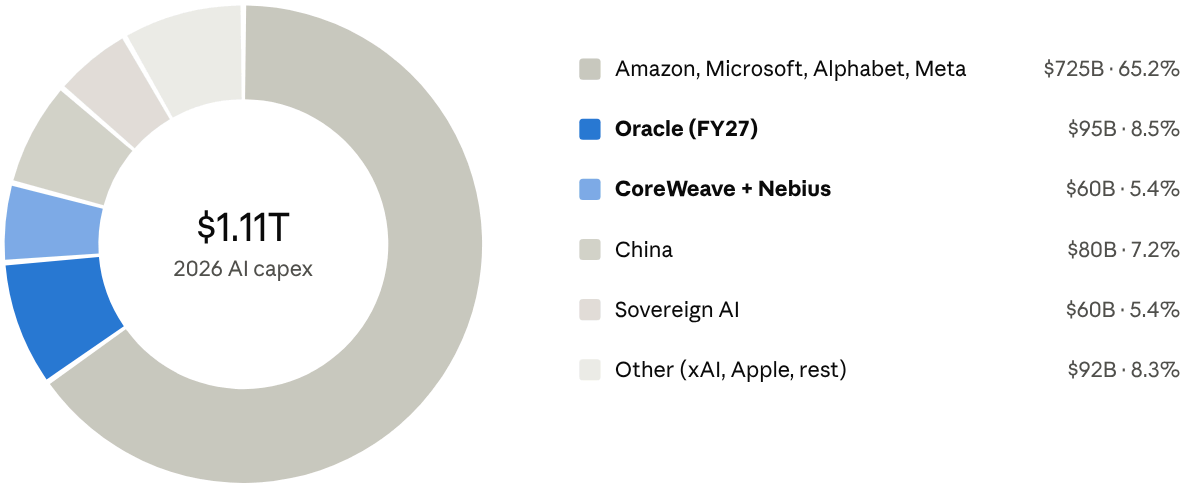
The credit market reflects this as of now. Since June, high-yield data center spreads have widened while investment-grade hyperscaler spreads have tightened modestly (Penn Mutual). In our view, that divergence is the opposite of a systemic bubble, where good and bad credits tend to sell off together.

Nvidia is the other anchor. Morgan Stanley estimates that its all-in credit exposure, mostly off-balance-sheet backstops, could approach $200 billion by end-2028 (Vested), yet the same model still shows about $350 billion of net cash by calendar 2028 (Tiger). Bank of America expects about $470 billion of free cash flow across 2026 and 2027 (Webull/Benzinga) and our model projects north of $300 billion of cash in FY28. If this is the case, then the backstops in the most extreme downside could all be covered by the free cash flow.

To put the leveraged builders in context, CoreWeave, Nebius and Oracle together account for roughly 14% of about $1.1 trillion of 2026 AI capex, while the four largest hyperscalers account for about two-thirds. The China, sovereign and other estimates are approximate, and some spend may be counted twice where Oracle and the neoclouds build for the same customers. Overall, we think the financial health of the spenders is solid. It will be important to carefully monitor earnings reports for any changes on this front. We expect margins on core hosting to rise despite a heavier depreciation burden.

CoreWeave, Nebius and Oracle are about 14% of 2026 AI capex

13.9% of about $1.11 trillion; the four largest hyperscalers are 65.2%



2026 capex, $B. Big four: FT via AI Weekly; Oracle FY27 guidance: Outlook Business; CoreWeave (InsiderFinance) and Nebius (TIKR) at range midpoints; xAI annualized pace: AI Weekly; China, sovereign AI, Apple and other: The Business Engineer AI capex map, May 2026

Early innings, with a credit filter

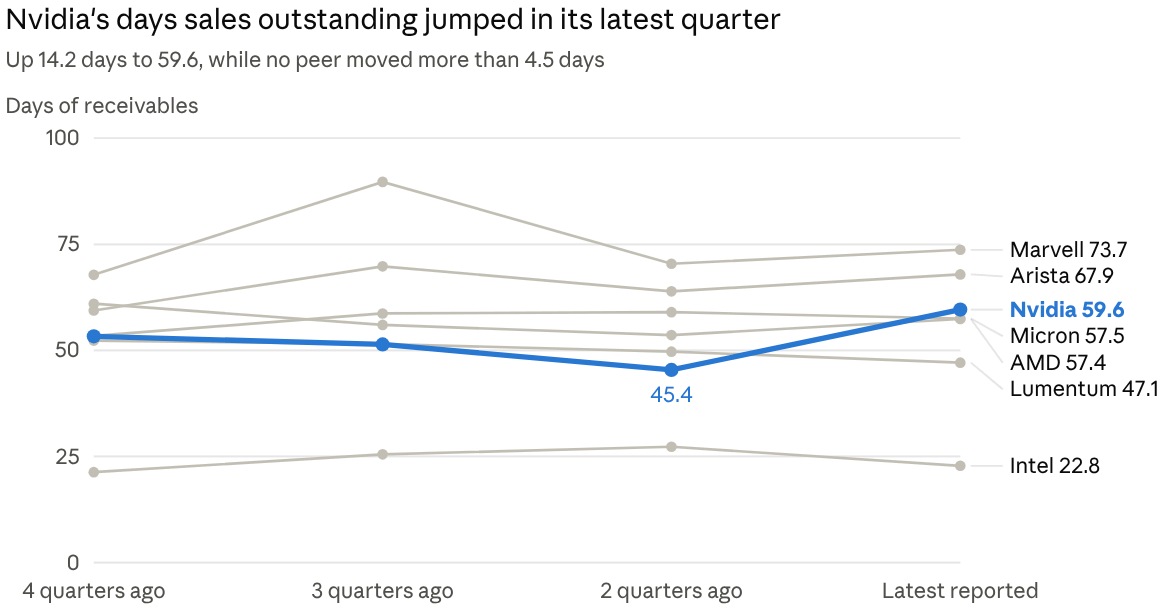
In our view, a large component of the volatility has been a repricing of who can afford to build rather than the end of the build. Demand is contracted rather than hoped for. Future revenue agreements at Amazon, Alphabet, Microsoft and Oracle reached about $2 trillion, up from $730 billion a year earlier (The Economist via The World This Wiki). Old chips are rising in price, and new ones are sold out. This comes back to our use cases: we think that both consumers and enterprises will pay for the intelligence because of its utility. Are token prices coming down? Yes, as they should. Are token costs going down? Yes: more efficient architectures and economies of scale.

What changes is the mix. We expect weaker, debt-funded builders to slow new projects or sell capacity to stronger hands. We expect that private equity or even governments will be at the ready to scoop up overleveraged assets given the useful life of the infrastructure. The cash-rich hyperscalers and Nvidia may also step in to consolidate the industry. That is consolidation inside a growing market, which in our view is the pattern of an early cycle rather than a late one.

What would change our view:

- **GPU rental prices.** H100 and other rental rates turning down

- **Hyperscaler spreads.** Investment-grade hyperscaler spreads widening alongside high-yield
- **Backlogs.** Backlog growth stalling or contracts being cancelled
- **Days sales outstanding.** At this point, most are stable or falling, but Nvidia's most recent jump did catch our attention, and we will be looking at that closely when the company reports in November.

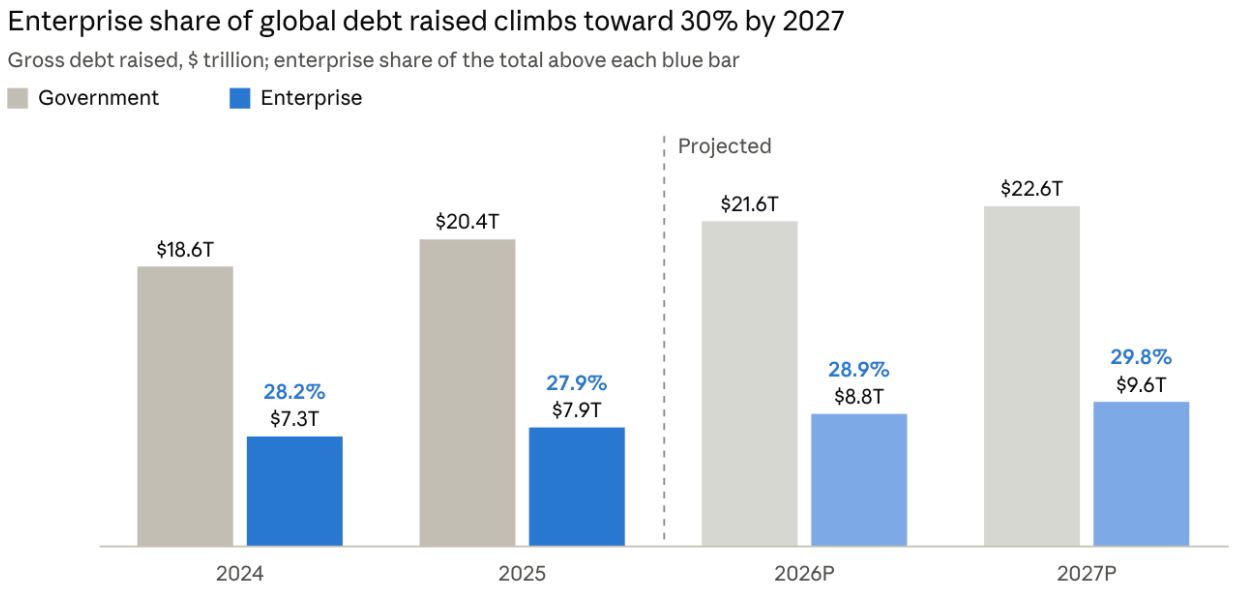


SEC filings (10-Q, 10-K, earnings releases) · each company's last four reported quarters · receivables ÷ quarterly revenue × days in quarter; Micron uses trade receivables

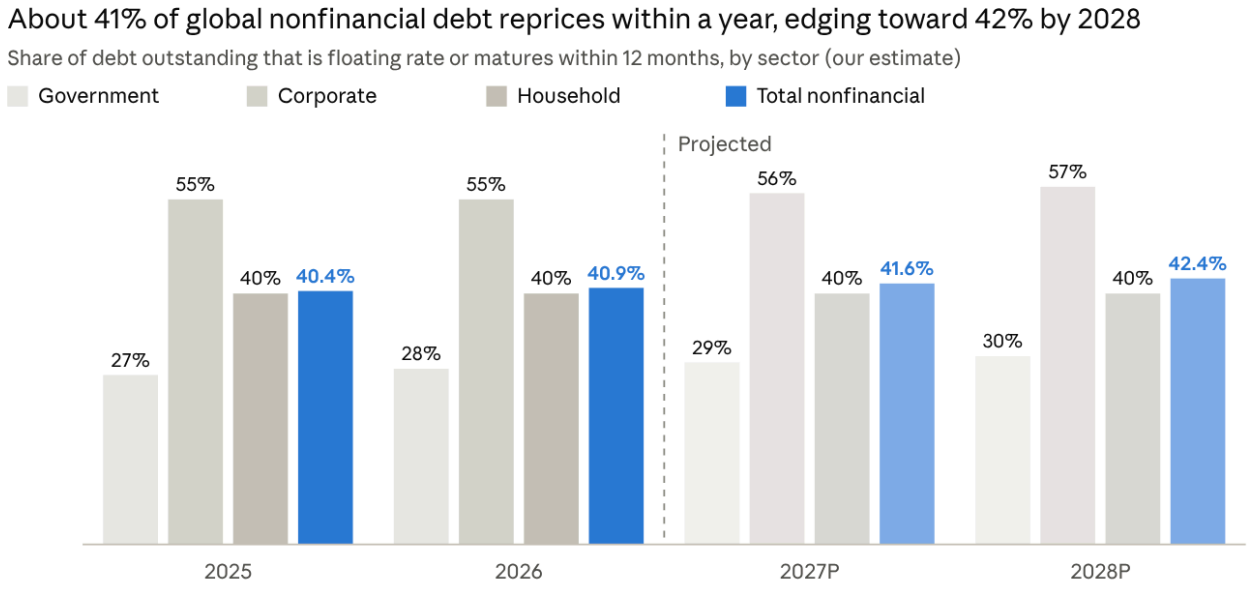
What's next: near peak fear, positioned for the long term

Short term, we are carefully monitoring the headwinds:

- **Macro:** continued conflict in Iran, higher inflation, rising rates in Europe and at leastE one more U.S. rate hike.
- **Credit:** if rates continue to rise, even the strong will face a higher hurdle to secure financing - and customers too will have to stretch, particularly small and mediumEbusinesses. We note that gross global bond issuance continues to grow and that theEshare of debt repricing within a year is climbing at a time when rates are meaningfullyEhigher.



Government: OECD Global Debt Report 2025–26 · Enterprise: OECD corporate bonds, AIMA (private credit), Simpson Thacher/Lexology (project finance), L&G, RBC and Crane Frontier (data center ABS/CMBS), with our estimates · 2026 OECD projection, 2027 our projection



Our estimate · Government: Joint Economic Committee (U.S. debt maturing in 12 months), OECD GlobalE Debt Report 2026 (bills, maturities) · Corporate and household shares: our assumptions · Weighted by IIFE Global Debt Monitor, Sept 2026 · Excludes financial-sector debt

- **Deglobalization:** this has concerned us for many years. In our view, it creates inflation and slows innovation.
- **Politics:** the upcoming U.S. elections.

- **Supply from the SpaceX IPO:** up to about 1.3 billion more shares unlock after Q3 results, and the full lockup ends December 8 (Everything PR).
- **Anthropic IPO pricing:** after the October 14 investor day, some investors discussed a valuation of about $1.5 trillion versus the roughly $2 trillion expected (FinanceFeeds/ Reuters). Some have argued that the IPO should price closer to its last raise in May, at about $1 trillion ($965 billion post-money), given the sequential slowdown and market correction. If pricing doesn't adjust, one of the two most important IPOs of the last five years risks breaking issue. A successful and stable IPO will be critical in our view. There has also been discussion of pushing back the Anthropic IPO given market conditions - we think this could also contribute to a delayed stabilization of the market because two of the most important participants, OpenAI and Anthropic, are not public, which makes it harder to fully understand the puts and takes of their profitability.
- **Leverage coming out:** at least one large, highly levered AI-focused hedge fund was forced to sell its public holdings this summer, and more large funds are rumored to be in trouble now.

In our view, we are near or at peak fear. We also think it is impossible to time the bottom. But given our view that we are in the early days of artificial intelligence adoption, it makes sense to revisit the sector now and look out to the long-term opportunities.

Who wins?

We also see possible upside from the shakeout itself. If builders with weaker balance sheets are removed from further expansion, the capacity that does get built rests on a stronger base going forward.

Our broader thinking on where value accrues:

- **Model layer:** we expect consolidation around a handful of players, with rapid product cycles as newer models reach the market. That leapfrogging will likely cap operating margins at scale for model vendors, but we still see upside from current levels. We have seen how quickly a model can rise and fall in popularity within months.. This does suggest more volatility ahead at this layer. We also see more and more frontier models with reasonably strong performance, which may pressure pricing. Ultimately, we think this will drive further adoption, as the falling cost of technology has always gone hand in hand with industry growth and advancement. Yet we are monitoring carefully to see how this plays out vis-à-vis long-term margins at the model layer. The cutthroat competition at the model layer may ultimately broaden out adoption due to falling token prices.

- **Compute:** we see significant demand for both training and inference. A consolidated customer list gives those customers more pricing power should those with higher cost of capital get priced out. But the stability and financial health of the surviving strong players make for a better long-term backdrop.
- **Components:** we see continued opportunity in faster optical interconnect, more advanced timing, new networking architectures and denser memory, whether in the data center or at the edge. We see increasing test complexity from semiconductors to networking.
- **Security:** we see this as table stakes. While security from chip to model is a must, there is also a very large opportunity for external security providers to independently monitor security - we see a sea change now as customers are actively reaching out to security providers and consultants to ensure safe agentic and AI use.
- **Supercharged software:** for those with moats of data and/or industry expertise, we see the use of LLMs as deepening engagement and revenue growth opportunities.

Who loses?

We see two kinds of losers. The first two groups below lose on credit now, which is cyclical; the rest lose as adoption grows, which is structural. That split fits our view that today's problem is credit, not demand.

- **Leveraged builders with weak tenants:** developers and neoclouds now borrowing at 9–15%, often against leases to less creditworthy tenants, may find that new builds no longer earn their cost of capital. We expect them to slow new projects or sell capacity to stronger hands, and their shareholders face dilution where cheaper debt is replaced by equity.
- **Lenders to that segment:** high-yield data center bonds, private credit and securitizations that rest on a single weak tenant. Spreads have already split from investment-grade hyperscaler debt, and a rising share of debt repricing within a year means more of it refinances at higher rates.
- **Businesses that profit from customer inertia:** the consumer and small-business pools we size are, in part, someone else's revenue. If agents re-shop insurance, cancel unused subscriptions and sweep idle cash, the losers are insurers that rely on customers not shopping around, banks earning a spread on non-interest-bearing deposits and subscription businesses that depend on forgotten sign-ups.
- **Labor-based services priced per hour or per head:** if AI captures a third of coding and customer support and a meaningful share of legal, accounting and back-office work, outsourced call centers, IT services, staffing firms, freight brokerage and tutoring face

pressure on volume and pricing.

- **Software priced per seat without data or industry moats:** the mirror image of our supercharged software winners. If agents are priced against labor rather than seats, customers need fewer seats, and software without proprietary data or industry expertise has little to defend.
- **Sub-scale model vendors:** we expect consolidation around a handful of players, so second-tier labs lose as leapfrogging and a growing number of capable frontier and open models pressure pricing. Custom-chip capacity tied to a single lab is also less fungible than Nvidia GPU capacity and could suffer if open models gain share.

Does AI solve our inflation problem?

If we are correct that we are dealing with a credit issue rather than a demand or overbuild issue, we wonder then if, ironically, artificial intelligence is also the solution. Historically, technology has been a great deflationary force and a strong productivity tool. If AI can increase its power as a tool for identifying cost inefficiencies, presenting lower-cost alternatives and executing those solutions on a continuous basis, then we wonder if the inflation problem might be resolved more quickly than the market currently expects.

Disclaimer

The information contained herein is based solely on the opinions of Erika Klauer, and nothing should be construed as investment advice. This material is provided for informational purposes, and it is not, and may not be relied on in any manner as, legal, tax, or investment advice or as an offer to sell or a solicitation of an offer to buy an interest in any fund or investment vehicle managed by Science & Technology Partners or any other Science & Technology Partners entity.

The information and data are as of the publication date unless otherwise noted. Content obtained from third-party sources, although believed to be reliable, has not been independently verified as to its accuracy or completeness and cannot be guaranteed. Science & Technology Partners has no obligation to update, modify, or amend the content of this post nor notify its readers in the event that any information, opinion, projection, forecast, or estimate included changes or subsequently becomes inaccurate.

The information above may contain projections or other forward-looking statements regarding future events or expectations. Predictions, opinions and other information discussed in this document are subject to change continually and without notice of any kind and may no longer be true after the date indicated. Science & Technology Partners assumes no duty to and does not undertake to update forward-looking statements.