

Who Actually Profits from AI?



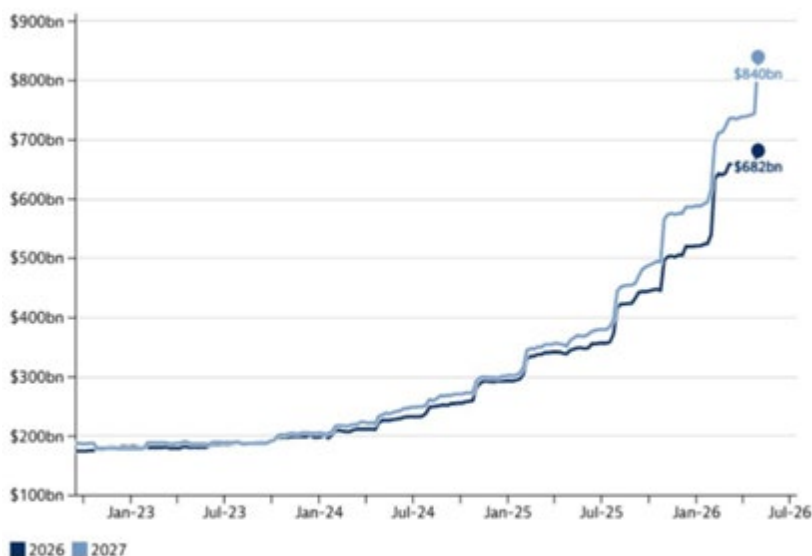
June 2026

1. The trillion-dollar question

The money being poured into artificial intelligence is staggering. The four biggest tech companies alone plan to spend somewhere around \$700 billion on AI infrastructure in 2026, and the industry's total spending is on track to pass \$1 trillion by 2027. Spending on that scale only makes sense if it comes back as profit — and that raises a simple question most headlines skip past: when the dust settles, who actually makes the money?

Industry analysts keep increasing their expectations for big tech spending

Consensus hyperscaler capex expectations

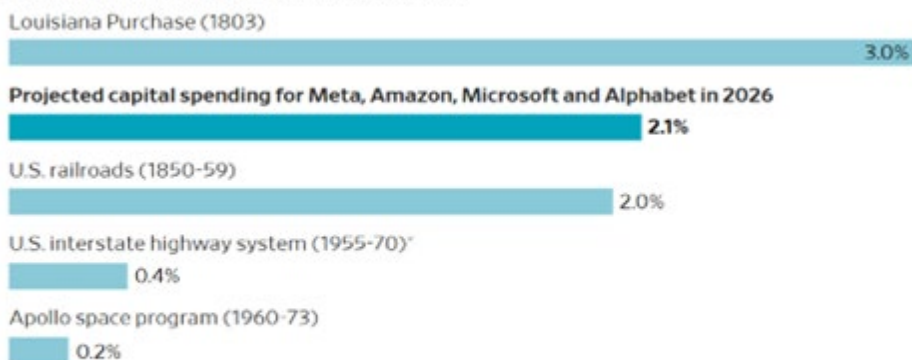


Source: Bloomberg, Goldman Sachs Asset Management • As of May 5, 2026
 The economic and market forecasts presented herein are for informational purposes as of the date of this presentation. There can be no assurance that the forecasts will be achieved. Please see additional disclosures at the end of this presentation.



The instinctive answer is “whoever builds the smartest AI.” This guide argues the real answer is subtle, and it can be stated in one line: **the profits flow to whoever owns the part of the system that stays scarce once AI gets cheap.** That single idea — follow the scarcity — is the thread running through everything below. As you’ll see, the companies famous for building models may not be the ones holding the scarce piece.

Spending as percentage of GDP, annual average



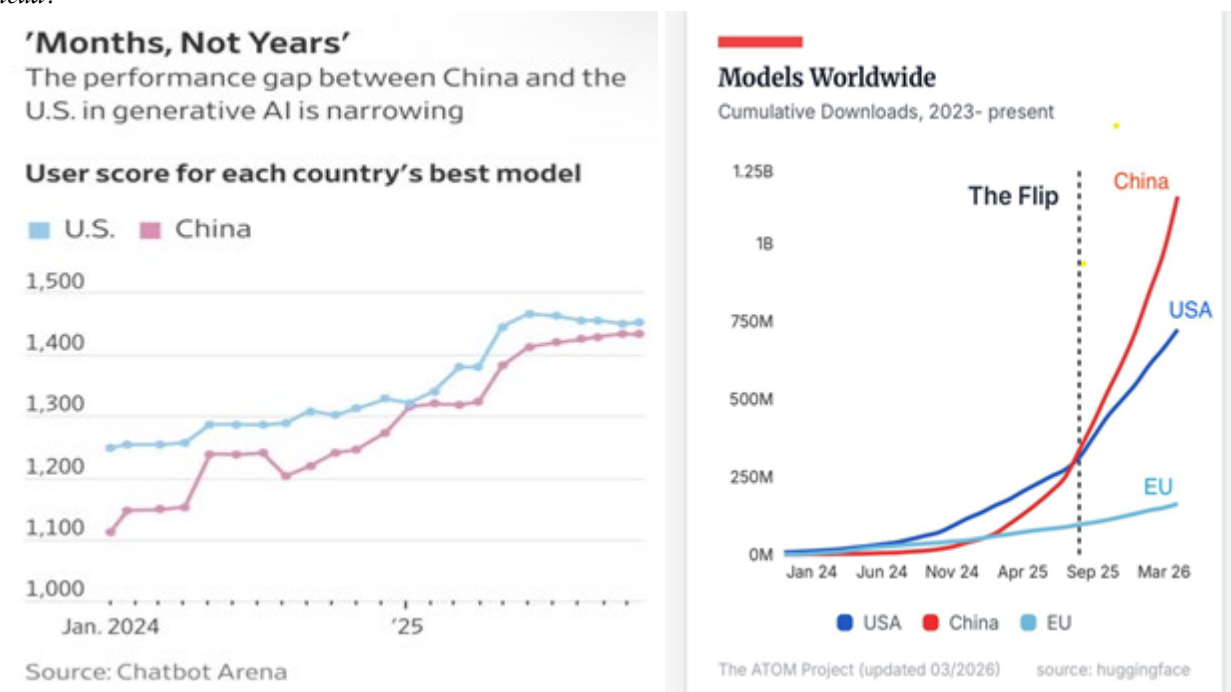
*Federal highway spending
 Sources: National Archives and Measuring Worth (Louisiana Purchase); the companies (Tech capital spending); Visible Alpha (projected Microsoft capital spending) Louis P. Cain (U.S. Railroads); Robert Gordon (U.S. interstate highway system); Planetary Society (Apollo); Wall Street Journal calculations (all ratios)

2. Why Building the Best Model May Not Be Enough

Imagine you invent the world's best recipe. If you're the only one who has it, you get rich. But if the recipe leaks and every restaurant in town can cook the same dish, the recipe itself becomes nearly worthless — and the money flows instead to whoever owns the *ovens*, the *prime real estate*, and the *loyal customers*. That is the risk hanging over the companies that build AI models, and the evidence suggests the recipe is leaking.

Two forces are at work. First, prices are collapsing: the cost to use a top-tier AI model has fallen dramatically from one generation to the next. Second, rivals are catching up cheaply — in early 2025, the Chinese lab DeepSeek built a competitive model at a small fraction of the usual cost, and the shock briefly erased roughly \$1 trillion in US AI-related market value in a single day. When the best capability becomes cheap and widely available, it stops being a goldmine and starts behaving like a commodity — something many suppliers can provide and nobody can charge a premium for.

This is the heart of the worry for “pure” model-makers: selling raw access to a model can become a brutal, low-margin business even as the technology itself changes the world. Strong demand and a money-losing business can coexist — a point we'll return to. The natural follow-up is the one investors should sit with: *if the model is the commodity, where does the value go instead?*



3. The Twist: Cheaper AI Grows the Whole Pie

There's an optimistic counter-argument, and it's a real one. When something gets cheaper, people often use far more of it — so much more that total spending actually rises. Economists call this the Jevons paradox, after a 19th-century thinker who noticed that more efficient steam engines led to more coal being burned, not less, because cheaper power kept finding new uses. Applied to AI, cheaper “thinking” could unlock enormous demand that was previously priced out.

A related idea pushes back on the fear that AI simply destroys jobs. That fear assumes there's a fixed amount of work to go around — economists call that the “lump-of-labor” fallacy. In reality, when a service gets cheaper, demand for it can expand: far cheaper legal help, for instance, could finally reach the millions of people and small businesses currently priced out of hiring a lawyer at all. The work doesn't vanish; it grows and shifts toward higher-value judgment that people still want from a human.*

But here's the catch that ties this section back to the last one. The engine driving all that extra demand is the *falling price* — and a falling price is exactly what turns models into a commodity. So cheaper AI is wonderful news for the total size of the market, and tougher news for the model-makers' profit margins. More demand doesn't rescue them; it can even deepen their problem. The growth is real — but it tends to reward whoever sells the scarce ingredient underneath, not the model itself. Which brings us to where the money actually pools.

**One honest qualifier: this works best in markets with lots of pent-up, price-sensitive demand. In heavily regulated, trust-dependent fields like medicine and law, price often isn't the main barrier — liability, licensing, and the simple desire for an accountable human are — so cheaper AI clears less of the backlog there, and some specific jobs can still be displaced before new work appears.*

4. Three Places the Money Coalesces

If the model itself is becoming a commodity, the value has to settle somewhere. It pools in at least three places — and the smartest players try to sit in more than one at once.

Layer 1 — The infrastructure (chips, data centers, power)

This is the “picks and shovels” layer. A chipmaker gets paid no matter which AI lab wins the race — it simply sells more shovels to everyone digging. That's a less risky bet than backing one specific lab to stay ahead forever, and it's why the dominant AI-chip company (Nvidia) became, for a time, the most valuable company on earth. The caveat: shovels only stay valuable while they're scarce. Rivals are designing competing chips, and the big tech giants are increasingly building their own — so the same commoditization that threatens model-makers could eventually creep down to the chipmakers too.

Layer 2 — The applications (software built on top of the models)

This is the layer that's easy to overlook, and it may capture a surprising share of the durable value. These are the tools people and companies actually pay for: coding assistants, customer-service agents, industry-specific software, workflow automation. They can hold pricing power because they own the workflow, the data, and the customer relationship — things that don't commoditize as fast as the model humming underneath. For example, leading AI coding tools like Claude Code and Cursor became one of the fastest software products ever to reach a billions of dollars in annual revenue. But the app layer has its own squeeze, often called the “token tax”: these apps usually *buy* their intelligence from the model labs, paying for every query. That cost can cap their profits well below those of traditional software — and it's why some app companies are now trying to build their own models to escape the toll. This is also why the line between “model company” and “software company” is blurring from both directions: app makers are reaching down to build models, while model labs are reaching up to build apps.

That blurring answers a question worth posing directly: *do the model labs have to become software companies to survive?* Largely, yes — and they know it. Selling raw model access alone is the weak hand, so the labs are racing to build their own products (coding tools, consumer assistants, enterprise agents) where the stickier moats live. But it isn't the only escape, and it isn't a guaranteed one. A lab could instead win as the lowest-cost producer, or by staying far enough ahead at the frontier that only two or three players can compete at all. And moving “up the stack” doesn't end the fight — it relocates it, since a lab that ships its own app now competes with the very customers it sells model access to.

Layer 3 — The integrated giants (who own several layers at once)

This is the most important correction to a too-simple “labs versus chipmakers” story. The biggest players aren't purely one thing. Companies like Microsoft, Google, and Amazon own cloud infrastructure, back or build their own models, and fold AI into already-profitable products that hundreds of millions of people use. Because they capture margin at several layers and plug AI into businesses they already dominate, they're naturally hedged — and arguably the best-positioned of all. The famous standalone labs grab headlines, but they're only one slice of the picture.

UPCOMING IPOs IN 2026				
RANK	COMPANY	SECTOR	WHAT TO KNOW	EST. VALUATION
1	 SpaceX	 Aerospace	Pricing Jun 11, trading Jun 12 on Nasdaq SPCK	\$1.75T to \$1.8T
2	 Anthropic	 AI	Raised \$65B in May	\$965B
3	 OpenAI	 AI	After \$122B round in March	\$852B
4	 ByteDance	 Social Media	TikTok parent	\$480B to \$500B
5	 Databricks	 AI Infrastructure	Enterprise data and AI platform	\$134B to \$165B
6	 Stripe	 Fintech	Global payments platform	\$105B to \$120B+
7	 Revolut	 Digital Banking	Super app for banking and investing	~\$75B

5. Two Real-World Bets at The Extremes

Two of the most striking strategies in AI sit at opposite ends of the “own the scarce layer” spectrum. Together they show the whole framework in action.

Apple: own the customers, rent the brain

Apple sat out the spending race almost entirely. Rather than build a frontier model or buy one outright, it chose a third path: rent. As of mid-2026, a rebuilt Siri runs partly on a custom version of Google’s Gemini model, in a deal reported at roughly \$1 billion a year across Apple’s two-billion-plus devices. This fits the thesis cleanly, because Apple already owns the scarcest asset of all — the customers: a billion-plus iPhones, the default-assistant slot, the payment rails, and a trusted brand. So the commoditization that frightens model-makers is, for Apple, the strategy: let models get cheap and interchangeable, then rent the best one and plug it into hardware people already own.

The catch is dependency. Apple handed the “brain” to a direct competitor. If users come to feel that the good experience is Google’s doing, the goodwill flows to Google — which may also earn a cut of purchases made through Siri, letting a rival monetize inside Apple’s own walled garden. The one-line version: Apple bet it can always rent a brain, but no one can rent its billion customers.

xAI / SpaceX: own everything, and out-fund the burn

At the opposite extreme is Elon Musk’s strategy of owning the entire stack. After SpaceX absorbed the AI company xAI in early 2026, the combined business now spans every layer at once: it’s designing its own chips, building enormous data centers (with a stated ambition for solar-powered “orbital” ones), running its own model (Grok), and owning a built-in distribution platform in X. In mid-2026 bought an option to acquire the coding app Cursor for about \$60 billion, adding the application layer too. On paper, this is the purest version of the integration bet — own the chips, the compute, the model, and the app, so you never pay a toll to anyone.

But the same case reveals integration’s limit: owning a layer doesn’t automatically fix its economics. Routing the new app onto in-house data centers doesn’t change the basic math — it still has to earn enough to cover its costs. And where the diversified giants integrate from a position of profit, this empire is integrating from a position of heavy cash burn, reportedly around \$30 billion over four quarters against a tiny sliver of AI revenue, even as its funding engine’s economics weakened. Tellingly, it built chips and data centers but still had to spend \$60 billion to buy the one scarce thing Apple already has for free — paying customers. The lesson: vertical integration is a bet that you can out-fund the losses long enough for one layer to turn profitable. “Own everything” and “make money” are different problems.

6. The hard questions an honest investor should ask

None of this is settled, and the strongest claims on both sides have real holes worth pressing on. Four questions matter most.

1. Is the moat actually **profitable**? The eye-watering valuations of the top labs — hundreds of billions of dollars — only make sense if you believe they own a durable moat, most likely a consumer ecosystem (huge user bases, habit, brand). But many of those popular subscriptions reportedly run at a loss today, subsidized to win market share, with heavy users costing more to serve than they pay. That leaves a vise: raise prices and risk losing price-sensitive users to a near-free rival; or wait for costs to fall — but falling costs help every competitor equally. So the real question isn't "do they have users?" but "can those users ever be turned into profit?"
2. **Are the neat "layers" actually merging?** The three-layer map is a useful snapshot, not a stable structure. The "token tax" is pushing app companies down into building models and model labs up into building apps, while the giants already span everything. The likely destination is consolidation — a handful of integrated players — which means today's tidy categories may blur into a few vertically integrated winners.
3. **Is the chipmaker's moat deeper than "scarce shovels"?** Quite possibly. The dominant chip company's real advantage isn't just hard-to-make silicon — it's a software ecosystem that developers have built on for over a decade. Overthrowing it means breaking a software habit, not just manufacturing a cheaper chip, which history says is far harder. That makes the infrastructure moat sturdier than "shovels stay valuable only while scarce" suggests — though the giants' in-house chips are aimed squarely at breaking exactly that lock-in.

Notice the common thread: in every case, the durable winner owns a **lock-in that doesn't commoditize** — customers, workflows, data, software habits, capital, or (in that last scenario) a breakthrough too hard to copy — wherever in the stack it happens to sit.

7. The Bottom Line

- The AI boom is real, and demand has grown faster than the early skeptics expected — the live debate has shifted from "will anyone pay?" to "do the profit margins work?"
- The same force that grows the market — falling prices — is the force that squeezes the pure model-makers, because it turns their product into a commodity.
- The money pools in at least three places — infrastructure, applications, and the integrated giants — not a simple "labs versus chips" split. The integrated giants look the most hedged.
- The two extremes are instructive: Apple rents the brain and owns the customers; xAI/SpaceX owns everything and bets it can out-fund the burn. Neither is risk-free.
- No layer is permanently safe. The edge belongs to whoever owns the part that **stays scarce** — and that can shift over time.
- The whole bear case on model-makers assumes progress keeps *diffusing*. A single durable, hard-to-copy breakthrough would flip it — handing one lab the scarcity this paper says the model layer lacks. That, more than the price of chips, is the swing factor to watch.

So the question to keep asking isn't "will AI work?" It almost certainly will. It's the harder one: **when AI gets cheap, who's holding the part that stays scarce — and who's positioned in more than one layer at once?**

This is general information, not investment advice. It is not a recommendation to buy or sell any security, and it isn't personalized financial or legal guidance. Figures are as reported by the sources listed and change quickly; do your own research and consider speaking with a licensed professional before making investment decisions

Sources

Figures are as reported by the outlets below, as of mid-2026. Several are self-reported or single-source estimates that move quickly — treat specific numbers as approximate.

1. Big Tech AI capital spending projected to top \$1 trillion in 2027 — CNBC — <https://www.cnbc.com/2026/04/30/ai-boom-big-tech-capital-expenditures-now-seen-topping-1-trillion-in-2027-.html>
2. 2026 hyperscaler capital spending (~\$700B) — Futurum Group — <https://futurumgroup.com/insights/ai-capex-2026-the-690b-infrastructure-sprint/>
3. DeepSeek shock and ~\$1 trillion single-day market loss; the AI revenue gap — Tool Directory — <https://tooldirectory.ai/blog/ai-capex-bubble-2026-where-the-revenue-actually-is>
4. The Jevons paradox applied to AI (“a commodity we just can’t get enough of”) — Outlook Business — <https://www.outlookbusiness.com/explainers/why-satya-nadella-thinks-jevons-paradox-will-shape-ai-industry>
5. Jevons paradox explainer and skepticism — NPR Planet Money — <https://www.npr.org/sections/planet-money/2025/02/04/g-s1-46018/ai-deepseek-economics-jevons-paradox>
6. Leading labs’ rapid revenue growth (Anthropic and OpenAI annualized revenue) — Trending Topics — <https://www.trendingtopics.eu/anthropic-overtakes-openai-in-revenue-hitting-30-billion-run-rate/>
7. AI coding tool fastest to \$1B+ annual revenue; application-layer value — TechJack Solutions — <https://techjacksolutions.com/ai-brief/the-software-layer-is-winning-what-cursors-reported-50b-valuation/>
8. The “token tax” squeezing application-layer margins — TechTimes — <https://www.techtimes.com/articles/317542/20260601/ai-agent-economics-token-tax-locks-gross-margins-30-points-below-saas-baseline.htm>
9. Dominant AI-chip company’s valuation and scale — Tech-Insider — <https://tech-insider.org/big-tech-650-billion-ai-infrastructure-capex-2026/>
10. Apple–Google Gemini deal (~\$1B/year custom Gemini powering Siri) — Fortune — <https://fortune.com/2026/01/13/apple-ai-deal-with-google-gemini-means-for-google-apple-openai/>
11. Apple–Gemini deal detail (~2B devices, “bridge strategy”) — Kavout — <https://www.kavout.com/market-lens/apple-and-google-ai-partnership-2026-everything-you-need-to-know-about-gemini-powered-siri>
12. SpaceX–xAI merger, vertical-integration framing, and orbital data centers — CNBC — <https://www.cnbc.com/2026/02/03/musk-xai-spacex-biggest-merger-ever.html>
13. SpaceX–Cursor \$60B deal and AI cash burn (~\$30B over four quarters) — MarketWise — <https://marketwise.com/investing/spacex-cursor-deal-ai-cash-burn/>
14. AI build-out scale and assumptions — Goldman Sachs — <https://www.goldmansachs.com/insights/articles/tracking-trillions-the-assumptions-shaping-scale-of-the-ai-build-out>