THE TRILLION-DOLLAR ILLUSION?

What If Artificial Intelligence Succeeds — and Its Builders Still Fail?

A documented investigation into the AI financial loop, the 2026 IPO race, debt, energy, falling model prices and the question of who ultimately pays for intelligence.

ADAM KOHN

Reporting cut-off: 02 November 2026

INVESTIGATIVE NOTE

This report distinguishes four evidentiary levels: documented fact, attributed reporting, structural inference and hypothetical stress test. It does not allege fraud or predict bankruptcy without proof. It distinguishes binding commitments, announced investment intentions, revenue recognized, cash paid and company valuation; it also separates gross accounting losses from operating performance.

Executive summary

The industrial system under construction around artificial intelligence is larger, more interconnected and more financially demanding than the familiar image of a software startup suggests. OpenAI disclosed $122 billion in committed financing at an $852 billion post-money valuation in March 2026. Anthropic filed its U.S. IPO registration statement in October, revealing explosive revenue growth and a layered program of future computing commitments valued in the hundreds of billions. These are genuine developments. The dispute is whether future customer spending will justify the scale, timing and rigidity of the obligations being created today. [1][2][7]

An unusual financial pattern connects the major participants: cloud companies invest in laboratories that purchase their services; hardware manufacturers sell the processors and participate in financing demand; infrastructure operators borrow against long-duration customer contracts; alternative asset managers purchase data centers and lease them to technology tenants. The IMF now describes circular financing as a risk transmission mechanism rather than an anecdotal peculiarity. None of this proves fictitious demand. It does mean that investors should not mistake every signed contract, funded data center and equipment sale for independent confirmation of final-customer willingness to pay. [4][14][15][24]

The immediate economic contradiction concerns time. AI capabilities are becoming rapidly cheaper: Stanford documented a more than 280-fold decline in the price of GPT-3.5-level inference in approximately eighteen months; the OECD reported nearly an 80% decline in its quality-adjusted model price index through April 2026. Meanwhile the industry is signing leases and computing commitments extending across much of the next decade. If the cost per useful task declines faster than revenues grow, utilization, margins and debt service become decisive. Increasing token usage is not itself proof of increasing profits. [22][23]

The physical expansion is real. The IEA estimates data-center electricity demand at 485 TWh in 2025, with a central projection of 950 TWh by 2030. Equipment manufacturing, grids, permits, water and financing do not move at model-release speed. Project Jupiter in New Mexico demonstrates that power and construction delays can interfere with a sophisticated contract chain before the computing it promises has been delivered. A failed project can impose costs beyond its developer if grid upgrades or other public commitments were financed in anticipation of its demand. [17][19][21]

The inquiry finds a legitimate commercial market, extraordinary growth and serious structural vulnerabilities. It finds no basis for claiming that laboratory sales are broadly fabricated or that a coordinated financial fraud has been proved. The strongest conclusion is less sensational and more important: AI may create enormous economic value even if some leading model developers, lenders or infrastructure operators fail to capture it. Technological success and the financial survival of particular firms are separate propositions.

Contents

01. The most dangerous success story
02. The filing that changed the argument
03. The arithmetic of an $800 billion promise
04. OpenAI's capital machine
05. The disappearing boundary between investor and customer
06. The IMF has entered the investigation
07. CoreWeave: the balance sheet behind the cloud
08. The $50 billion debt warning
09. Nvidia sells the machines — and helps finance the buyers
10. Wall Street becomes the landlord of intelligence
11. The price of intelligence falls as its factories multiply
12. The competent cheap model
13. Proof of productivity is not proof of profit
14. The energy system keeps the receipts
15. The construction site that punctured the narrative
16. The depreciation problem
17. The valuation race is a separate market
18. How the bubble can be real without the technology being fake
19. The financing stress test
20. The public pays where contracts fail
21. Concentration masquerading as diversification
22. Regulation can become an economic moat
23. An alternative future: intelligence becomes a commodity
24. What the evidence cannot yet prove
25. The documents investors should demand
26. The final buyer of intelligence

Appendix A. Timeline of the financial build-out, 2022–November 2026

Appendix B. Claims, classifications and unresolved questions

Source notes and bibliography

Methodological note: all monetary amounts are in U.S. dollars unless specified. The date of each measurement is material. Forecasts, market estimates and attributed analyses are labeled as such.

01

The most dangerous success story

The technology could prosper while its financial pioneers disappear.

The central danger to the artificial-intelligence boom is not necessarily that the technology disappoints. A more disturbing possibility is that it works. Models could become cheaper, more reliable and sufficiently capable to enter almost every industry, while the companies that financed their development struggle to earn adequate returns on the infrastructure required to build them. The distinction between social usefulness and private profitability is a recurrent feature of industrial history. Railways made nineteenth-century economies richer without guaranteeing the survival of every railway financier; fiber-optic networks eventually transformed communications although many of their original owners suffered devastating losses. Intelligence could follow a similar path. What society inherits and what early investors recover are not the same question.

By late October 2026 the economics had become concrete enough to investigate rather than merely speculate about. Anthropic's public securities filing disclosed extraordinary commercial growth, complex governance and vast future spending commitments. OpenAI had announced $122 billion of committed financing in March, alongside a valuation of $852 billion. Nvidia was helping assemble capital-market arrangements to fund demand for its processors; specialist cloud operators were borrowing against multiyear computing contracts; energy authorities warned that electricity demand was growing faster than grid development could comfortably accommodate. These facts describe a new infrastructure economy, not just a software industry. [1][2][7][14][19]

They also reveal a mismatch of time horizons. A model can lose pricing power within months. A data-center lease may bind a company for a decade. Debt instruments may mature while facilities are still being commissioned. Electricity infrastructure can require years to permit and construct, while the software it supports changes several times over during construction. Revenue from customers remains contingent on adoption, price, competition and actual useful work; contractual payments to suppliers can be considerably less flexible. That mismatch is the subject of this investigation.

This report advances a testable proposition: parts of the AI industry may be constructing a long-lived, debt-financed production system around a product whose unit price is undergoing unusually rapid deflation. The proposition does not imply that the money is fictitious, that executives are conspiring or that frontier models have no commercial value. It asks who bears the risk if customers ultimately need less expensive intelligence than the builders have assumed, and whether losses could migrate from laboratories to lenders, utilities, pension funds and the wider economy. The strongest evidence is not an anonymous prediction. It is in regulatory filings, research on model prices and an IMF analysis that now explicitly identifies circular financing as a financial-stability channel. [23][24][25]

> INVESTIGATIVE RULE: a technologically successful industry can still contain overvalued businesses; a financing loop is not, by itself, proof of fictitious revenue.

02

The filing that changed the argument

Anthropic’s S-1 replaced venture-capital mythology with a paper trail.

The documentary turning point arrived on 26 October 2026, when Anthropic filed its registration statement with the United States Securities and Exchange Commission. Unlike the highly selective announcements typical of venture funding rounds, an S-1 places a company’s operations, risks, ownership rights, accounting judgments and contractual exposures in a common legal framework. The filing is not a government certification of its forecasts. It is, however, a much stronger basis for investigation than a conference presentation or a speculative valuation headline. Anthropic reported revenue of about $386 million in 2024, $4.59 billion in 2025 and $16.29 billion during the first half of 2026. These figures are exceptionally rapid growth, making the claim that frontier AI has no real paying customers untenable. [1][2]

The same document complicates the equally simplistic claim that Anthropic is merely burning limitless money. Its first-half 2026 reported net loss of around $80.7 billion was dominated by approximately $72.4 billion of noncash changes in the fair value of convertible instruments. Operating loss for the period was about $693 million, and Reuters reported positive operating income in the second quarter. The distinction matters. A noncash valuation adjustment can dwarf the actual resources consumed in operations; describing the former as if it were an operating cash loss is inaccurate. But positive operating income over one quarter does not settle whether an enterprise can sustain the cash requirements of a multiyear infrastructure expansion. The financial question shifts from today's accounting profit to tomorrow's contractual bill. [5][6]

The prospectus also demonstrates that 'valuation' is not synonymous with liquid wealth. An anticipated IPO valuation around $2 trillion is a proposed market judgment, not money sitting in Anthropic’s bank account. The amount raised in an offering, any secondary sale proceeds and the effective economic rights of different share classes are separate matters. Anthropic's unusual voting architecture grants a founder-controlled vehicle 50.1% of voting power on many corporate questions despite having no economic entitlement. A long-term benefit trust holds distinctive board-election rights. Thus a public investor may purchase an exposure to AI's economic future without purchasing proportionate power over its strategic decisions. [1][6]

The filing also provides a vital discipline: distinguishing already binding obligations from contemplated capacity and from the additional contracts expected in future periods. Numbers that appear to conflict across September and October reporting often cover different dates, counterparties or contract categories. It is irresponsible to add a September $518 billion total to an October $637 billion total and declare the company committed to more than $1 trillion. The $518 billion was an earlier disclosed perimeter; the later filing broadened or updated the picture. The exact dates and definitions must travel with every reported figure. [1][3][5][6]

03

The arithmetic of an $800 billion promise

The largest numbers concern tomorrow’s payments, not today’s revenue.

The scale of Anthropic's compute obligations is difficult to reconcile with ordinary software-industry intuition. Reuters reported in late September that the company's contemplated infrastructure arrangements included at least $518 billion over roughly ten years, with about four-fifths largely noncancelable or payable without regard to utilization. The report assigned approximately $111.1 billion to Google, $110 billion to Amazon, $31.4 billion to Microsoft and $161.2 billion to Broadcom leases. These are not cash expenditures already realized, nor are they necessarily all identical legal promises: different counterparties provide distinct services, hardware and capacity under distinct conditions. They show that access to computation, rather than the purchase of office software or conventional research equipment, has become the strategic bottleneck. [3]

After the public S-1, coverage reported approximately $637 billion of compute contracts in one classification, while a Reuters Breakingviews calculation discussed around $800 billion of leases and purchase commitments when considering a wider set of contractual obligations. The larger figure should be read in its proper context. It includes contracts not all immediately active; some capacity may be adjustable or cancelable; a contract's headline value is not equivalent to the amount due in any single year. The useful question is how much is contractually payable in a given year, what termination rights exist, and which payments become unavoidable if customers unexpectedly reduce usage. [5][6]

Reuters Breakingviews offered a transparent sensitivity test rather than claiming certainty about future revenues. Spreading the reported commitments across their contractual periods, the columnist estimated roughly $97 billion in 2027 compute payments. At a computing-cost-to-revenue ratio of 55%, that would imply around $177 billion in sales to maintain the ratio. That is not an audited cash-flow forecast or a representation of management guidance. It is an illustrative calculation using stated assumptions and known agreements, and its power lies in making the scale of the hurdle visible. Anthropic's $16.29 billion of first-half 2026 revenue is impressive. It does not by itself establish a pathway to an order of magnitude higher annual revenue. [5]

The imbalance is structurally dangerous because model customers may buy by consumption while developers acquire capacity with minimum commitments. Variable selling volume against less-variable purchased capacity is a familiar problem in airlines, shipping, telecommunications and electric utilities. It can create exceptional profitability at high utilization and punishing fixed-cost leverage at low utilization. The AI debate often asks whether models become more intelligent. A lender must ask whether reserved accelerators are actually being used by customers who pay enough to cover the bill. The same machine that appears as a strategic asset in an expansion scenario can become a liability when its utilization falls.

> FINANCIAL CAUTION: $518B, $637B and approximately $800B are differently scoped snapshots and analyses, not additive components.

04

OpenAI’s capital machine

A valuation is an expectation; committed funding is not all cash already spent.

OpenAI provides a different route to the same problem. On 27 February 2026, it announced $110 billion of prospective investment at a $730 billion pre-money valuation, identifying $50 billion from Amazon and $30 billion each from Nvidia and SoftBank. On 31 March it announced a completed financing round with $122 billion of committed capital and an $852 billion post-money valuation. These are sequential disclosures describing a financing process, not two independent funding rounds that can simply be summed into $232 billion. OpenAI stated that it was generating around $2 billion of revenue per month at the time of the announcement, with more than 900 million weekly active ChatGPT users and over 50 million subscribers. [7][8]

Such numbers demonstrate a formidable distribution advantage. Yet they are not evidence that all users pay, or that large consumer engagement translates automatically into high enough net income to support new computing generations. OpenAI itself characterizes compute availability as a compounding advantage: more computing supports better models; better models improve products; products attract usage; usage generates revenues for reinvestment. This 'flywheel' is a coherent growth thesis, not an empirical guarantee that incremental capital will always earn a competitive return. It assumes that customers value the next increment of model capability more than it costs to deliver. [7]

The strategic partnerships make the revenue chain harder to disaggregate. Amazon and Nvidia were named among the financial backers; Microsoft remains both an investor and infrastructure partner. OpenAI's expanding supplier portfolio includes Microsoft, Oracle, AWS, CoreWeave and Google Cloud, while its compute architecture incorporates Nvidia, AMD, Cerebras, Trainium and custom silicon with Broadcom. Diversification can improve resilience and negotiating power, but it also multiplies commercial relationships among enterprises that buy, supply, finance and compete with one another. The fact that investment partners also become sales counterparties does not make the transactions illegitimate. It increases the importance of tracing ultimate users and the sources of final cash receipts. [7]

The capital market began to question the story more visibly in late October. Reuters reported on 31 October, citing the Financial Times, that OpenAI had reduced a separate fundraising target by about $10 billion amid weakness in technology shares, and that valuation ambitions were being reconsidered. This was attributed reporting about negotiations, not a completed repricing. It nevertheless illustrates the dependence of the expansion model on continued willingness to supply capital at attractive valuations. If investors demand higher returns, the same projected spending becomes harder to finance; if the next model generation requires still more computation, the company may face a choice between dilution, borrowing, partnership concessions and slower expansion. [33]

05

The disappearing boundary between investor and customer

The capital relationship can become the commercial relationship.

Consider the classical distinction between a customer and an investor. A customer transfers money to receive a product or service; an investor transfers money in expectation of a future financial claim. Artificial intelligence has produced a landscape in which the same counterparty often performs both roles, while also supplying the manufacturing capacity needed to deliver the product. Amazon and Google distribute Claude through their cloud marketplaces, sell Anthropic computing services, have invested in the developer and operate their own AI research programs. The resulting relationships are commercially understandable yet analytically complicated. Reuters reported that marketplace sales through those two platforms reached approximately 47% of Anthropic's 2025 revenue, versus 11% in 2023. [4]

Marketplace distribution is a real service. It reaches enterprise customers with existing cloud accounts, procurement approvals and integrated billing. It can accelerate the adoption of otherwise unfamiliar technology. But the same arrangement means that the laboratory's distribution margin depends on the fee structure and business incentives of a company that may be trying to promote an alternative model. It also creates accounting questions about whether marketplace gross sales or revenue net of distribution fees provides the more useful basis for comparing laboratory economics. Reuters documented a disagreement between Anthropic and OpenAI concerning the comparability of gross versus net recognition. There is no reason to assume either presentation is inherently fraudulent. The investigation must identify who contracts with the buyer, who bears performance obligations, and who controls pricing. [4][6]

An investment may also support product demand indirectly. When a strategic partner supplies equity capital, that financing can fund purchases of cloud capacity and equipment from that same partner or its affiliates. This is not proof of wash trading: real computing services may be delivered and real engineering problems solved. It is, however, important to avoid mistaking the resulting revenue for fully independent evidence that unrelated customers have chosen to purchase equivalent volumes at the same price. The most probative evidence would be a breakdown of end-customer payments, intercompany transactions, refunds, credits, minimum purchase requirements and partner-specific margins. Public disclosures provide some of this information but not all. [1][4]

The political economy becomes especially sensitive when a supplier is also a financial rescuer. The laboratory may be reluctant to change providers because of contractual restrictions; the supplier may be reluctant to let the laboratory fail because it would impair expected utilization; lenders may rely on both parties remaining solvent. That web of incentives can sustain ambitious investment during a boom and transmit shocks during a downturn. It is not a secret conspiracy. It is a series of publicly observable incentives, and it deserves the same scrutiny that investigators apply to related-party transactions in conventional finance.

06

The IMF has entered the investigation

Circular finance is no longer merely a skeptical journalist’s metaphor.

In October 2026 the International Monetary Fund published an unusually relevant technical note: The AI Ecosystem: Structure, Financing, and Financial Stability Considerations, by Yingyuan Chen, Deepali Gautam and Jason J. Wu. The note describes an AI value stack and explicitly identifies circular financing arrangements linking equity investment, debt financing and future revenue commitments. It observes that advanced AI development was initially funded largely by profitable incumbent technology companies but was becoming more capital intensive and more reliant on external capital. Market concentration, technological obsolescence and correlated security prices could amplify shocks if expected AI profits or productivity benefits disappointed. The report does not proclaim an imminent financial crisis. It warns about a recognizable vulnerability. [24]

The IMF's April 2026 Global Financial Stability Report had already examined the financial co-movement of what it described as the AI 'circle', including listed firms exposed to deals with privately held AI developers. Its analysis connected circular dealmaking with higher asset-price correlations and possible systemic spillovers. This is a decisive distinction for the investigation. The concern is not simply that startup valuations might be exaggerated. It is that a correction in one part of the chain could change the perceived credit quality or commercial outlook of several ostensibly different firms at once. An investor who believes they own diversified exposure through a chip designer, a cloud company, an infrastructure lender and a model laboratory may hold positions that all depend on the same underlying AI spending cycle. [27]

IMF Managing Director Kristalina Georgieva sharpened the message in October, noting that hyperscalers' share of investment-grade bond issuance had grown from below 1% to around 10% over two years, and warning of the potential international propagation of an AI-related market correction. That matters because risk cannot be contained merely by pointing out that an individual cloud provider has a strong balance sheet. Strong firms can transmit volatility through suppliers, capital markets and common investor portfolios. The probability of a systemic crisis remains uncertain; the channels through which one could propagate are easier to specify. [26]

A rigorous account must also acknowledge the IMF's balancing conclusion. The current AI cycle is anchored in major profitable companies and many risks remain manageable. Equity investors, private-credit lenders and data-center operators are not all equivalent to highly leveraged subprime borrowers. Financial stability analysis therefore asks how exposures evolve: who is guaranteeing whom, what collateral can realistically be sold after a shock, whether debt matures before contracts generate cash and how much apparent diversification survives once shared customer dependencies are recognized.

DOCUMENTED: the IMF explicitly discusses circular financing and correlated vulnerabilities.

SPECULATIVE: asserting that those channels will necessarily produce a crash.

07

CoreWeave: the balance sheet behind the cloud

One customer once accounted for two-thirds of sales.

CoreWeave offers a rare opportunity to inspect an AI infrastructure company through an audited public filing. Its 2025 Form 10-K reports $5.131 billion in revenue, compared with $1.915 billion in 2024 and $229 million in 2023. It also reports a net loss of about $1.2 billion for 2025 and remaining performance obligations of $60.7 billion at year-end. The backlog figure measures contracted future obligations to deliver services, not cash held in a bank account. CoreWeave's filing explains that it finances infrastructure primarily through asset-level debt supported by take-or-pay customer contracts. This is a recognizable project-finance model with recognizable risks. [9]

The most revealing number is customer concentration: Microsoft accounted for approximately 67% of CoreWeave's recognized 2025 revenue. That is not an allegation from a short seller; it is the company's own disclosure. The company also acknowledged that a significant reduction in spending by one or more major customers could materially affect its business. The significance is obvious. Although CoreWeave presents itself as an infrastructure provider to the AI economy, a large share of its historical operating performance depended on one technology buyer. Investor exposure to CoreWeave was therefore partly indirect exposure to Microsoft's decisions about data-center expansion and its own suppliers. [9]

The customer base has been diversifying through substantial new contracts. Reuters reported that OpenAI agreements had reached up to $22.4 billion by September 2025. Meta added significant capacity orders, including a $21 billion agreement in April 2026 supplementing an earlier deal. Diversification is real but must be measured at the right stage: contracts signed today become revenues only when computing capacity is deployed and service obligations are met. Backlog diversification is not the same as current revenue diversification. The same contract may require years of construction and tens of billions of dollars in capital expenditure before the operator receives the corresponding cash flows. [12][13]

CoreWeave's own description of its business emphasizes multiyear committed contracts, typically with take-or-pay conditions. Those provisions can protect operators from sudden customer withdrawal. They also establish why the creditworthiness of the counterparty matters so much. A high-quality customer contract can support borrowing against expensive processors; if the customer experiences financial distress or if construction delays prevent timely delivery, the expected cash flows may no longer cover scheduled debt service. There is a contractual bridge between software usage and financial markets, but it is not invulnerable.

08

The $50 billion debt warning

A cloud contract is not the same as an investment-grade balance sheet.

On 15 October 2026 Reuters reported that CoreWeave had accumulated roughly $50 billion of debt in support of its expansion, prompting pressure in both its shares and bonds. The reporting cited elevated credit-default insurance costs and substantial required returns on some of its securities. These market prices can be interpreted as signs of heightened perceived risk, not as precise actuarial forecasts that default is inevitable. CoreWeave continued to argue that its contracts and backlog supported its expansion. Nvidia's chief executive publicly expressed confidence in the company's improving finances. These opposing assessments belong together in a responsible account. [10][11]

The mechanism is important. A neocloud operator may borrow to purchase accelerators or construct facilities, using signed customer commitments as evidence that future cash flows can service the debt. A lender evaluates the customer's ability to pay, the equipment's resale value, expected utilization, service margins and construction timetable. If the equipment becomes technologically obsolete sooner than expected or if a major customer changes direction, the security underpinning the loan may depreciate before the debt is repaid. Unlike a road or bridge, a GPU system is part of a rapidly changing technology cycle. Its economic useful life may be shorter than the real estate, cooling systems and electricity connections surrounding it. [9][10]

There is also a difficult mismatch between accounting and liquidity. A company can report contractual backlog and strong gross demand while experiencing cash stress because new facilities require upfront spending. Positive EBITDA does not necessarily imply positive free cash flow after capital expenditure and interest. Similarly, a customer commitment can appear strong in headline terms while imposing costly performance obligations or requiring expensive equipment upgrades. The relevant questions are not merely how many billions of backlog exist, but how they are distributed over time, whether capacity is delivered and which party bears costs when hardware generations change.

By late October, the broader neocloud sector had become a visible focus of concern. Reuters reported that Elon Musk warned of consolidation and failures among debt-dependent operators, while several listed firms' shares and bonds had weakened. These are attributed predictions and market observations, not proof that a specific company will fail. The investigative significance lies in the changed mood of financing. When capital is abundant, a customer contract looks like permission to expand. When capital becomes expensive, the very same contract is tested for its ability to generate cash soon enough to support debt already incurred. [35]

09

Nvidia sells the machines — and helps finance the buyers

The supplier is becoming part of the funding architecture.

Nvidia occupies an extraordinary position in the financial anatomy of AI. Its accelerators, interconnect technologies and software ecosystem underpin much of the industry's training and inference capacity. A conventional interpretation would identify Nvidia as a supplier whose revenue benefits when AI developers purchase more processors. Yet in August 2026 Reuters reported that Nvidia was partnering with Apollo, BlackRock, Blackstone, Brookfield, Goldman Sachs and KKR to create financing arrangements targeting more than $500 billion of third-party capital for AI infrastructure. The report described potential Nvidia support for part of the contemplated deals. It did not establish that the full amount was already financed or deployed. [14]

The financing concept matters because it brings the processor supplier closer to the business of securing its own downstream demand. In a simplified example, a financial institution lends money to a data-center operator that buys Nvidia hardware; the operator leases computing capacity to an AI customer; payments from that customer help repay the loan; Nvidia recognizes hardware revenue when the applicable accounting conditions are satisfied. If the chain works, productive infrastructure is built at enormous scale. If usage declines and operators fail, losses are distributed among lenders, equity owners and potentially guarantors. The transaction does not become illegitimate because a vendor participates in the financing. Vendor financing has a long history in heavy industry and telecom equipment. But it becomes a problem when investors assume every stage of the chain represents a separate, independent source of demand.

Nvidia also illustrates the concentration of the technology stack. A chip purchase may involve proprietary software, network fabrics, memory and specialized systems that make switching costly. Enterprises may diversify among competing architectures, but the short-term availability and performance characteristics of chips can strongly influence where AI applications are deployed. The more dependent financing arrangements become on a particular equipment generation, the more important it is to analyze hardware useful life and the residual value that lenders assume their collateral retains.

The best counterargument to excessive pessimism is that Nvidia's equipment is genuinely useful and global demand for computation continues to grow. Chips are not merely financial tokens. They carry real productive capability, and the company has substantial independent commercial relationships. The correct criticism is therefore not that Nvidia is 'financing fake sales'. It is that an infrastructure seller with exceptional market power may also be helping enlarge the pool of capital that purchases its infrastructure. That feedback mechanism deserves transparent disclosure and systematic stress testing. [14][24]

10

Wall Street becomes the landlord of intelligence

The private-credit story is a real-estate story, a lease story and a technology story.

The large technology companies are no longer the sole financiers of the AI construction boom. Reuters reported on 29 October 2026 that major alternative asset managers were expanding data-center investments based on long leases and the creditworthiness of tenants. Blue Owl described more than 130 data centers owned or under development, while Ares and Blackstone emphasized pre-leasing arrangements involving investment-grade hyperscalers. These structures can resemble infrastructure finance more than venture capital: an asset owner constructs or acquires a facility, a tenant signs a long lease and institutional investors receive returns backed by contracted rental payments. This is not necessarily fragile. Long-term leases with strong counterparties can be more predictable than speculative merchant computing operations. [15]

But the asset itself is a hybrid. The building and land can remain useful for decades, whereas the servers it contains may change generation several times. Cooling technology, power density, electrical equipment and networking design may become outdated faster than the financing assumptions anticipate. If a tenant renegotiates, relocates or fails, an operator may discover that a supposedly general-purpose data center is tailored to a specialized computing configuration. This is residual-value risk: what is the facility really worth when the initial contract ends? The question is especially important where lenders treat committed demand as a substitute for careful examination of technological obsolescence. [15][16]

Moody's warned in October that AI-related financing was expanding private credit and increasing refinancing exposure. Refinancing risk is not the same as current inability to pay. A borrower may be solvent under existing terms yet face difficulty when a debt maturity arrives in a higher-rate environment or when capital markets no longer value future AI demand as generously. The danger is that credit underwriting based on strong long-term narratives may become vulnerable to short-term changes in market confidence, construction schedules and equipment prices. [16]

The most revealing beneficiaries of the boom may thus be institutions with comparatively modest exposure to whether one model beats another. A landlord leasing space to an investment-grade tenant, or a financier with properly priced collateral and covenants, may capture returns without forecasting which model will dominate. Yet the protection is conditional. If a large share of apparently different tenants ultimately depend on the same model laboratories, cloud buyers or financing markets, counterparty diversification can be weaker than it looks. The investigation must therefore follow the tenant's revenue source beyond the signature on the lease. [15][24]

11

The price of intelligence falls as its factories multiply

A 280-fold price decline is an achievement and a threat.

The Stanford 2025 AI Index documented an extraordinary reduction in the cost of model inference at a fixed performance threshold. For a system reaching GPT-3.5-level performance on the MMLU benchmark, the estimated price fell from $20 per million tokens in November 2022 to $0.07 by October 2024. That is a decline of more than 280-fold. The result does not mean every AI task has become 280 times cheaper or that benchmark performance exactly measures real business value. It does establish a profound trend: capabilities that were once expensive are migrating rapidly into much cheaper systems. [22]

The OECD's July 2026 analysis extended the observation. Its quality-adjusted language-model price index had fallen by nearly 80% between January 2024 and April 2026, with declines across coding and agent-related segments. The economist sees a tension that the AI enthusiast may overlook. Falling prices can expand demand and raise consumer welfare while simultaneously reducing the revenue available to providers per unit of intelligence. Whether the provider benefits depends on changes in cost, volume, market share and the value created by more complex applications. A thousandfold increase in usage can produce growth, but only if the price and marginal cost structure make that usage profitable. [23]

There is a tempting analogy to the Jevons paradox: efficiency can increase total resource consumption by making a product more affordable and widely useful. Such an effect is plausible for AI but not guaranteed to preserve the margins of any individual laboratory. Energy consumption can expand while electricity prices fall; internet traffic can soar while bandwidth becomes commoditized; model usage can explode while competition transfers most of the economic benefit to customers and application developers. More tokens are not automatically more profit.

The OECD also identifies a complication: sophisticated AI agents may consume orders of magnitude more tokens than simpler tasks. Consequently, a falling price per token may coexist with rising expenditure per completed task. The right unit of economic comparison is not a million tokens in isolation; it is the fully loaded cost of correctly completing work, including retries, verification, security, integration and failures. This is especially relevant for science, medicine and regulated industries, where an apparently cheap model may produce expensive downstream errors. The central question is whether improvements in cost per successful task keep pace with the financial commitments made to produce those improvements. [23]

12

The competent cheap model

Most customers do not buy intelligence for its own sake.

A fundamental assumption behind the race to ever-larger frontier models is that additional model capability will continue commanding a premium. This is true for some applications but not necessarily for the ordinary work that makes up a large share of enterprise software spending. Invoices require extraction and reconciliation. Procurement teams need classification and search. A hospital may need a reliable administrative summarizer with access controls rather than a model that excels at abstruse mathematics. A manufacturing company may prefer a smaller model that runs locally because data sovereignty, predictability and latency are more valuable than an extra fraction of a benchmark point. These are examples of economic choice, not claims that smaller models always perform as well. [22][23]

The significance of open-weight models is not simply that they eliminate licensing fees. They allow developers to select a fixed version, deploy it within a controlled environment, adapt it to a specific task and test it against local requirements. That can be attractive when a proprietary API changes frequently or when data cannot be transferred to an external provider. But operating a model locally is not automatically cheap: hardware, cybersecurity, monitoring, engineers, maintenance and quality assurance all matter. The comparison must be total cost of ownership rather than a marketing price for inference.

The OECD highlights that different segments of the AI value chain display different competitive conditions. Foundation-model markets can be dynamic and pricing may fall rapidly, while some upstream infrastructure and distribution layers are far more concentrated. If model capability becomes less scarce than cloud access, specialized chips, proprietary data or enterprise relationships, value may move away from the laboratory creating the model and toward companies that control the workflow in which it is used. The effect would resemble the fate of commodity suppliers in earlier technological transitions: they can underpin enormous economic activity without capturing proportionate profits. [23]

The counterargument is equally material. Frontier capabilities can create markets that did not exist before: long-horizon agents, automated scientific experimentation, advanced design and software systems that execute multi-step professional work. Businesses may pay far more for a reliable automated outcome than for access to individual tokens. The frontier is potentially where new demand is invented. But a commercial valuation should not assume that every enterprise customer needs that frontier or that frontier pricing power will persist indefinitely. The question is how much of the market requires the best model, how long the advantage lasts and how easily competitors can reproduce it.

13

Proof of productivity is not proof of profit

Economic value created and value captured are two different measurements.

A sound investigation cannot treat enterprise interest in AI as hype. Research increasingly documents adoption, time savings and changes in work. In an experiment involving 7,137 knowledge workers across 66 firms, researchers associated with NBER observed reduced time spent on email among users of an integrated generative-AI tool, although they did not find broad shifts in the composition of workers' tasks attributable to individual access alone. The findings are important precisely because they are more specific than sweeping predictions about universal automation. A tool can generate meaningful gains without immediately restructuring a company or doubling its revenue. [28]

Survey evidence also suggests that adoption is highly uneven. A 2026 NBER study using U.S. Census Bureau data found that about 18% of firms used AI in at least one function during November 2025 through January 2026, with much higher adoption in large and knowledge-intensive businesses. A broader survey of senior executives reported much wider AI use, but the two studies measured different populations and different concepts of adoption. Any headline claiming that 'most firms use AI' must specify whether it means occasional employee experimentation, a licensed tool, deployment in a business function or material operational transformation. The distinction changes estimates of future spending. [29][30]

The business model also depends on who captures the gains. Suppose an insurer saves $10 million annually from faster document processing. Competition among AI vendors may allow the insurer to keep most of that benefit, buying models for a small subscription or usage fee. Suppose a software company produces code faster using a sophisticated assistant. Higher productivity may accrue to its employees, customers or shareholders rather than to the model developer. Even transformative economic outcomes need not translate into equivalent payments to the supplier of computation. A company can help create a large social surplus while retaining only a small slice.

This is not an argument that AI developers cannot earn profits. Proprietary products, user trust, integration, switching costs, reliability and networks of developers can support durable business advantages. Yet a serious forecast must estimate the share of productivity value paid to the model provider instead of assuming that increased GDP automatically finances every new data center. The journey from macroeconomic benefit to laboratory cash flow has several toll gates, and not all of them are controlled by the laboratory. [23][28][32]

14

The energy system keeps the receipts

Electricity cannot be financed with a token benchmark.

Global data-center demand for electricity was about 485 terawatt-hours in 2025, according to the International Energy Agency, and its central 2026 projection puts demand at roughly 950 TWh by 2030. AI-focused data-center electricity demand is projected to triple over that period. The same report estimated that capital expenditures by five leading technology companies exceeded $400 billion in 2025 and could rise by approximately 75% in 2026. These are not merely impressive technology statistics. They describe large claims on physical generation, transmission, transformers, cooling equipment, water systems, labor and capital. In some localities a data center is becoming an industrial utility customer comparable in size to a heavy factory. [19]

The physical system cannot expand at the speed of model releases. Long transmission lines require land access and permits. Gas turbines and power electronics have manufacturing lead times. Electricity-system operators cannot simply accept unlimited proposed load without assurance that generation and network capacity will exist. The IEA notes that the power density of high-end AI facilities is rising rapidly and that data centers can induce unusually sharp changes in electrical load. Scarcity of high-bandwidth memory and other specialized components can delay the computing equipment even if buildings are completed. Power procurement is therefore both a technical constraint and a financial liability. [19]

The distinction between a gigawatt announced and a gigawatt reliably supplied matters. Developers sometimes cite planned megawatts in press releases long before substations, pipelines or permits are secured. Construction milestones and commissioning dates provide more informative evidence. Likewise, announced clean-energy procurement does not by itself reveal whether new generation physically matches a facility's hour-by-hour demand. Storage, grid rules and firm power need to be examined. The IEA's October 2026 report on powering AI infrastructure emphasizes the need to integrate data centers without transferring disproportionate costs to other consumers. [21]

An especially troubling possibility is that infrastructure costs migrate beyond the AI industry. When a utility builds capacity based on optimistic demand forecasts and a project is delayed or abandoned, it must still account for investment already made. Depending on tariffs and regulatory structures, the cost may be recovered from the developer, shareholders, other power customers or some combination. The fairness question is concrete: who pays for the grid connection, who supplies financial security against cancellation, and who bears the depreciation if demand disappears? A speculative AI expansion should not quietly become an involuntary public subsidy. [19][21]

15

The construction site that punctured the narrative

A contract for future computation is not computation delivered.

Project Jupiter in New Mexico offers a case study in the distance between financial promise and physical completion. Reuters reported on 24 September 2026 that Oracle had invoked a force-majeure clause relating to possible power-supply delays at a data-center development associated with Blue Owl and Stack Infrastructure. The project was intended to support computing services for OpenAI. Reporting described difficulties connected with a natural-gas pipeline as well as legal and permitting disputes involving water and air quality. Oracle's notice reportedly sought to delay payment obligations; Blue Owl said the notice did not alter Oracle's underlying financial commitments. These are attributed accounts of a disputed project, not a finding that the enterprise was insolvent or had cancelled the facility. [17]

The case is economically illuminating because financing events and physical milestones do not occur in lockstep. Banks can commit funds, tenants can sign leases, technology companies can announce partnerships and investors can mark securities higher before a single profitable AI query is served from the facility. The sequence is normal for major infrastructure. It is also risky. A delay may increase interest expenses, extend the period of low or no operating revenue and alter the economics of equipment ordered before the facility opens. If next-generation accelerators arrive before the building is ready, part of the original hardware business case may need revision.

The construction timetable is subject to stakeholders who are not participants in the AI financing round. Municipal authorities, water regulators, electricity companies, landowners and environmental groups can change a project's timing. These actors are not necessarily opposing progress; many are enforcing ordinary legal requirements. Their decisions reveal why artificial intelligence cannot simply be 'scaled' through financial commitments. Physical capacity exists only when a sequence of engineering and public decisions succeeds. [17][19]

The larger investigative lesson is to distinguish four project stages: announced, financed, under construction and operating at commercial utilization. Financial press accounts often use the word 'capacity' for all four. Yet their expected future cash flows have radically different risk profiles. A signed lease may protect the landlord financially, but only if construction, performance and counterparty solvency all cooperate. A completed data center may carry significant value, but only if power is available and customers actually use it. Reporting that fails to separate these stages exaggerates certainty.

16

The depreciation problem

A GPU can age faster than its debt.

AI finance involves a physical asset that is unusually difficult to value over time. A contemporary accelerator may be expensive, scarce and revenue-producing, but each successive generation can deliver better price-performance, memory capacity or energy efficiency. The owner therefore faces two different depreciation processes: accounting depreciation selected under financial-reporting rules and economic depreciation determined by what customers are willing to pay for older machines. These can diverge. The problem is not new; telecom equipment, aircraft and industrial machinery also have uncertain residual values. AI chips simply make the issue unusually visible because technological performance is moving quickly. [9][22]

CoreWeave's 2025 Form 10-K recorded a sharp increase in technology and infrastructure expenses driven substantially by depreciation and amortization. The filing described approximately $2.3 billion in related depreciation and amortization, versus $843 million in the prior year, following its large-scale investment in platforms and equipment. That growth is not necessarily evidence of distressed assets. It is the normal accounting consequence of an enormous hardware expansion. But the choice of useful lives materially influences reported profit and the timing at which capital expenditures flow through the income statement. Financial analysis should therefore compare accounting policies across operators and test what happens if equipment must be replaced sooner than expected. [9]

An operator with a five-year customer contract and a similarly timed debt schedule may feel protected. Yet if prices for comparable computation fall significantly after two years, its customer may have little desire to renew at the original rate. If a contract gives the buyer price-adjustment or termination rights, the exposure arrives sooner. If the contract is firmly take-or-pay, risk shifts toward the customer's financial strength. The creditor must still ask whether outdated equipment has sufficient resale value to provide security in a downside case. Hardware collateral is not equivalent to a building that can be repurposed for decades.

An illustrative one-year change in assumed equipment useful lives could materially affect a technology operator’s reported profits. That is an analytical scenario, not a restatement of current earnings. It captures a key blind spot in the public narrative: financial success depends not only on model performance and demand, but on the difference between economic obsolescence and the financing schedule of the machines delivering the service. [9]

17

The valuation race is a separate market

Private valuations, IPO prices and exchange capitalization are not interchangeable.

Public discussion of the AI boom often mixes three quantities: a private financing valuation, the anticipated valuation of a future flotation and the daily market capitalization of a listed company. They may all be expressed in billions of dollars, but they answer different questions. OpenAI's March 2026 funding announcement placed it at $852 billion post-money. Anthropic's planned IPO was discussed in contemporaneous coverage at a potential value approaching $2 trillion; that was not a guaranteed offer price or a measure of proceeds actually received. Listed companies such as Nvidia, Microsoft or Alphabet have market values that fluctuate every trading day. Comparing a prospective flotation target with an established public-market capitalization as if both were simultaneous market-clearing prices produces misleading rankings. [2][7]

The valuation problem becomes more acute when companies issue securities with different rights, liquidation preferences or conversion formulas. A small funding transaction can imply a large enterprise valuation without proving that all existing shares would sell for that amount. Secondary transactions can also carry their own constraints. IPO investors confront additional risks from governance, lockups and the availability of future financing. Anthropic's S-1 makes this visible through its special voting structure and unusual range of stock classes. Public shareholders may receive substantial economic exposure while founders and the benefit trust preserve important governing powers. [1][6]

Financial markets nevertheless have reasons to ascribe enormous valuations to the AI laboratories. A system that reliably automates sophisticated professional labor could create a large market, and a company with defensible distribution or capability advantages might capture a significant share. But forecasts become fragile if they implicitly assume that enormous future demand will coexist with limited competition, stable margins and predictable infrastructure costs. A valuation constructed on heroic revenue growth can be undermined by modest changes in discount rates or terminal profitability even while absolute revenue continues to rise.

Late October 2026 brought a practical reminder of how quickly the discount rate and risk appetite can change. Reuters reported that OpenAI's new fundraising target was being reduced in negotiations as technology stocks weakened. At the same time, Anthropic's S-1 made public-market investors confront gross versus net revenue, future contract obligations and governance structures. These are ordinary elements of securities analysis, not proof of impending collapse. They show that as AI leaves private financing and approaches public ownership, its extraordinary expectations must increasingly be expressed in ordinary accounting language. [33][6]

18

How the bubble can be real without the technology being fake

The railway analogy is useful precisely because railways worked.

The phrase 'AI bubble' is often interpreted as a claim that artificial intelligence provides little genuine value. That interpretation is mistaken. A financial bubble occurs when prices and commitments become disconnected from the returns that assets can plausibly generate, not necessarily when the underlying product is useless. Railways reshaped industrial societies while many railway securities performed disastrously. Telecommunications infrastructure survived the dot-com collapse while some original owners did not. A communications revolution can be real, and the financing undertaken in its name can still be excessive. The distinction is central to understanding the present investment cycle.

Artificial intelligence may create immense increases in productivity. The IMF suggested in October 2026 that an effective diffusion of AI could raise global growth over time, but it also stressed that reorganizing firms, work and institutions is slow. This is a timing problem for investors. Infrastructure spending occurs now and is front-loaded; productivity gains may arrive gradually, be unevenly distributed and accrue to businesses that buy AI rather than those that sell it. A company can therefore correctly forecast the direction of technological change and still misjudge how quickly its own revenues will capture that change. [25][26]

Market structure amplifies the problem. A small number of semiconductor designers, foundries, cloud providers and frontier laboratories occupy strategically important positions, yet they do not all have identical competitive advantages. Some operate profitable existing businesses capable of absorbing temporary AI losses. Others are new companies dependent on capital markets. Consolidation could leave the technological infrastructure intact while wiping out substantial equity stakes or restructuring debt. The effect would resemble the aftermath of earlier infrastructure booms: overcapacity is eventually absorbed by new uses, often at prices lower than those assumed by original financiers.

The analogy has limits. Today's hyperscalers are not nineteenth-century speculative railroad promoters, and many are supported by exceptionally profitable software, advertising and cloud businesses. Nor can one assume that all AI investments will become stranded: inference demand, new agents and scientific applications may justify substantial additional computing. The correct analytical claim is narrower: technological inevitability is not a sufficient investment thesis. The terms of finance, the speed of physical construction, the decline in prices and the distribution of profits determine who survives. [24][25]

19

The financing stress test

Three plausible shocks can reinforce one another.

A useful stress test begins without predicting disaster. Imagine, first, that quality-adjusted model prices fall faster than forecast because smaller and open-weight models close the performance gap. Imagine, second, that businesses adopt AI widely but remain selective about spending, replacing expensive general-purpose models with cheaper specialized systems wherever performance is adequate. Imagine, third, that higher interest rates or disappointed investors increase the cost of refinancing infrastructure debt. Each scenario is compatible with AI becoming more useful. Their combination would nevertheless be difficult for an operator whose revenues vary with usage but whose leases and debt service remain fixed. [22][23][16]

To make the mechanism explicit, consider a purely illustrative operator with $10 billion in annual contracted capacity costs and $15 billion in revenue. If price competition cuts revenue per comparable task by 30% while paid task volume rises only 20%, revenue becomes $12.6 billion before mix effects. If its unavoidable capacity bill remains $10 billion, the amount left for all other expenses shrinks from $5 billion to $2.6 billion. If utilization grows faster, the result could be much better; if costs per task also fall, margins might improve. The point is not the assumed parameters but their sensitivity. The same demand increase can be financially welcome or disastrous depending on price and marginal cost.

A lender's stress test should examine five linked channels. Does the AI customer still generate enough cash to purchase reserved computing capacity? Does the operator have enough margin and liquidity to service debt? Does the equipment retain value if the customer fails? Does the landlord have a viable replacement tenant if the operator cannot pay rent? Does the grid operator recover connection and network costs without spreading them unfairly among other consumers? If one answer changes, the risk can propagate through counterparties rather than remaining with the original laboratory. The IMF's examination of circular financing provides an institutional rationale for asking these questions together. [24][26]

There is also a strong upside case. Frontier agents may unlock high-value tasks whose complexity and reliability requirements support substantial pricing, and improvements in hardware may reduce unit costs faster than customer prices decline. In that environment, present infrastructure may look cheap in retrospect and scarce capacity may earn extraordinary returns. An investigation must present this possibility rather than treating failure as foreordained. A stress test demonstrates vulnerability under assumptions; it does not deliver a forecast. [23]

> MODEL WARNING: all numerical stress-test amounts in this chapter are hypothetical and are not financial forecasts for any named company.

20

The public pays where contracts fail

Electricity bills, pension funds and local infrastructure sit outside the laboratory.

The companies that directly sign AI contracts are not the only parties exposed to their consequences. Pension funds may own listed technology shares, private-credit vehicles or real-estate investments backed by data centers. Insurance companies may lend against long-term leases. Banks may underwrite bonds that finance infrastructure. Communities may approve land and water use or extend electricity networks in anticipation of promised investment and employment. These exposures create a larger distributional question: if the AI financing boom falters, how much loss will be absorbed by private risk capital, and how much might be transferred to institutions serving the general public? [15][16][25]

The electricity dimension is especially important because network investment can become unavoidable before the demand forecast is realized. A developer might request a large grid connection, prompting a transmission upgrade. If the project is delayed or never opens, the accounting treatment of sunk costs depends on contracts and regulation. Robust rules can require developer deposits, enforce financial guarantees and allocate costs to those creating them. Weak arrangements can leave remaining consumers contributing to infrastructure whose expected principal user has vanished. The IEA explicitly raises concerns that uncertainty over data-center load can translate into local electricity-price pressures without appropriate policy design. [19][21]

The social effects are also mixed. A new data-center campus can generate construction work, local tax revenue and business activity. Once operational, however, it may employ fewer people per dollar invested than a conventional industrial plant. Electricity and land use can carry opportunity costs. If a local economy grants subsidies on the assumption that AI will produce permanent high employment, the promises should be measured against actual operating jobs, wages and fiscal transfers. A useful investigation would collect utility filings, zoning approvals, subsidy agreements and local environmental assessments, not just press releases by developers.

The distribution of AI benefits adds another layer. Workers and consumers can gain substantially from cheaper tools even if financiers lose money. Conversely, shareholders may profit from infrastructure investment while some workers experience job displacement or wage pressure. Those effects cannot be reduced to a single public-interest score. They require separate analysis of productivity, labor bargaining, competition and fiscal outcomes. The question 'who pays for intelligence?' is therefore both a cash-flow question and a political question about how the costs and benefits of infrastructure are divided.

21

Concentration masquerading as diversification

Owning five AI companies may still be one financial bet.

Contemporary AI portfolios can appear diversified while sharing an underlying dependence on the same expenditure cycle. An investor might hold shares in a chip designer, a hyperscaler, a specialist cloud operator and a data-center landlord, and also own a credit fund financing GPU equipment. On paper these exposures span technology hardware, software, real estate and private debt. Economically, they may all depend on the continued belief that advanced AI will attract enough end-customer spending to justify additional data centers. Correlation matters most when optimism fades, because securities that behave differently in ordinary markets may move together during a shock. [24][27]

The IMF's examination of the AI value stack is particularly valuable here. It shows why traditional sector labels are insufficient for mapping vulnerability. A cloud provider's investment in a model laboratory may strengthen the laboratory's purchasing power; that purchasing power can support processor orders and data-center leases; those orders create expectations of revenue for public companies. When the market capitalizations of those public companies rise, additional financing can become available on favorable terms. The process may reinforce genuine economic development, but it can also amplify disappointments. If the first cash-flow premise weakens, several financial claims can be reassessed at once. [24][27]

This does not establish that AI finance resembles a Ponzi scheme. Such a claim would require evidence of deception and a specific mechanism in which old investors are paid primarily with new investor money under fraudulent pretenses. The documented AI arrangements generally involve actual chips, real leases, legitimate services and publicly reported investment relationships. The more precise concern is an interdependent capital cycle in which independently priced assets share contingent exposure to one optimistic technological narrative. Similar concentration appears in property booms and commodity expansions without requiring a central fraud.

A serious investigation should therefore build an ownership-and-obligations graph, not merely a ranking of market capitalizations. Nodes would include investor entities, laboratories, clouds, equipment manufacturers, lessors, lenders and utilities. Edges would identify actual equity, debt, leases, take-or-pay purchases, guarantees and distribution fees, each dated and marked as binding, proposed or conditional. Such a network would reveal where risk is concentrated and where attractive-looking transactions effectively depend on the financial health of the same small group of companies.

22

Regulation can become an economic moat

Safety may be necessary and still favor the richest incumbents.

The AI industry now faces governance debates concerning cybersecurity, model evaluations, data access and autonomy. These debates are not merely about public relations. Agentic systems interacting with real computer networks can create direct operational risks, especially where they have access to financial systems, personal information or critical infrastructure. Stronger validation and security requirements may be legitimate. Yet compliance has a cost, and that cost can affect market structure. A large laboratory can maintain dedicated safety teams, external evaluators, regulatory lawyers and audited infrastructure; a smaller developer may struggle to afford equivalent systems. Rules that protect the public can simultaneously raise barriers to entry.

This tension matters to the economic story because concentration can influence pricing power. If expensive compliance standards restrict competition, leading laboratories may retain a premium despite falling technical costs. If open-weight and smaller models remain viable, customers may exert more pricing pressure. Neither outcome is automatically desirable in every application. A sufficiently dangerous capability may justify strict oversight regardless of commercial effects. But policy should identify the safety benefit, quantify compliance costs where feasible and avoid assuming that a rule is neutral merely because it applies to all firms in formal terms. [23]

The ownership question is also present in Anthropic's governance. Its S-1 describes mechanisms intended to preserve a public-benefit mission alongside stockholder interests, including special founder rights and the Long-Term Benefit Trust. These structures may help resist short-term pressure. They also concentrate authority over consequential decisions. The source establishes the governance architecture; it does not establish that the founders will misuse it. The investigative question is whether independent oversight can be strong without simply entrenching the organizations already controlling frontier models. [1][37]

Economic incentives and sincere safety commitments are not mutually exclusive. Executives may believe a risk is real and support rules that also advantage their businesses. A regulation can reduce harm while increasing compliance costs and limiting new entrants. The appropriate journalistic response is neither conspiracy theory nor unquestioning acceptance. It is to trace who helped draft standards, who pays for certification, who can comply and whether independent evaluators can publish adverse findings without financial or contractual retaliation.

23

An alternative future: intelligence becomes a commodity

The biggest beneficiaries may not be the companies that invented the frontier.

Imagine a future in which highly capable models are widely available at low marginal cost, where many tasks can be handled on local or specialized systems, and where competition makes model-level margins modest. Society could gain enormously. Universities might gain inexpensive research assistants; hospitals could access reliable clinical documentation tools; manufacturers might deploy advanced scheduling and quality-control systems; software companies could automate routine engineering. None of those gains requires every foundation-model vendor to remain independently profitable. The value of intelligence can migrate into the products, organizational processes and proprietary data that use it. [22][23][28]

This possibility resembles earlier technological waves. Semiconductors produced vast economic value across consumer electronics and industrial automation, but value did not remain evenly distributed among all chip designers. Internet connectivity became ubiquitous while many network operators struggled with heavy capital costs. Commodity cloud resources enabled profitable software businesses that did not own the original physical infrastructure. AI could become a similarly foundational input whose largest economic winners emerge in downstream applications rather than upstream model training.

There are reasons this future may be incomplete. Proprietary frontier models can be meaningfully differentiated in reliability, multimodality, agentic performance and scientific reasoning. Models integrated into trusted consumer products can benefit from distribution scale and network effects. Specialized enterprise systems can create switching costs even if their underlying model can be replaced. Cloud providers, hardware suppliers and model laboratories may each retain defensible positions in different parts of the market. It is premature to assert that intelligence will become a perfectly competitive commodity.

Still, the commodity scenario is a useful financial stress case because it asks whether contractual spending remains affordable if a substantial share of tasks moves to much cheaper systems. Investors betting on a permanent frontier premium should identify exactly which tasks command it, how often customers upgrade, what alternative suppliers can offer, and whether the provider's share of business value increases or declines over time. A technology may become indispensable while the identity of the winning suppliers changes repeatedly. The implication is not that AI is worth little. It is that a large portion of its value might accrue somewhere other than today's best-capitalized laboratories.

24

What the evidence cannot yet prove

A financial loop is not a completed account of where every dollar goes.

The investigation has important limits. Public SEC documents can disclose material obligations and partner relationships, but they do not expose every invoice, transfer price, GPU utilization profile or end-customer contract. Many AI laboratories remain private, and even public filings aggregate economically different business lines. Reported annualized revenue is not the same as audited annual revenue, just as an announced partnership is not the same as a binding purchase order. An expert analysis must mark these distinctions rather than collapsing them into the largest available headline number. [1][7][9]

The evidence also does not establish that the majority of AI revenue is fictitious or that strategic investors have manufactured demand as part of a coordinated fraud. Anthropic and OpenAI have real products and identifiable customers; CoreWeave delivers actual computing services; Nvidia sells useful hardware; data centers consume measured electricity. The IMF identifies potential vulnerabilities from circular financing without declaring that those relationships are themselves improper. Allegations of market manipulation would require transaction-level evidence about concealed side agreements, misleading accounting, fabricated counterparties or intentionally false disclosures. None is proven by the material reviewed here. [24]

Some widely circulated claims about imminent collapse rely on aggressive extrapolations from a single year's losses or on treating decade-long obligations as bills due immediately. Others take rising market capitalizations as evidence that the same dollars have been counted repeatedly. Those shortcuts make for dramatic headlines but weaken the case. The stronger story is more subtle: the economic system can be highly productive and financially vulnerable at the same time. Real assets and legitimate contracts can be overfinanced. Rapid growth can coexist with an unfavorable balance between payment obligations and eventual customer spending.

There are further unknowns at the model level. Reliability and capability may improve enough to expand demand by far more than price decreases reduce revenue. Training costs could be moderated by algorithmic breakthroughs; alternative chips could cut costs; public institutions might purchase large amounts of AI services. Conversely, legal liability, privacy requirements, safety events and organizational inertia could slow deployments. The uncertainties point in opposite directions. A documented investigation should treat each causal path separately and decline to substitute narrative certainty for missing data.

> EVIDENTIARY RULE: no reviewed source supports a blanket allegation of fabricated AI sales, coordinated fraud or inevitable laboratory bankruptcy.

25

The documents investors should demand

Disclosure should follow the money all the way to the final user.

An effective transparency regime would begin with a simple reconciliation of commercial flows. For each large AI laboratory, investors should see revenue from independent end customers separated from revenue booked through strategic investors' marketplaces, together with gross-to-net distribution reconciliations and the financial significance of credits or minimum purchase obligations. For each material capacity agreement, disclosure should identify the duration, scheduled payments, termination options, utilization requirements and any investment ties between buyer and seller. These requirements are not an argument for publishing confidential customer data or trade secrets. They are a proposal for comparable financial categories that would allow outsiders to distinguish genuine paid adoption from the financing conditions that make capacity available. [1][4]

Infrastructure operators should disclose the bridge from backlog to cash flow: committed capacity, expected delivery dates, revenue-recognition schedules, capital expenditure still required, debt maturities and the economic life of purchased processors. CoreWeave's 10-K offers a partial model by reporting remaining performance obligations, customer concentration and its reliance on asset-level financing. Lenders and investors still need carefully defined stress tests for customer default, equipment depreciation, construction delays and power costs. The proper time to ask those questions is before a financing market closes, not after collateral becomes difficult to sell. [9][10]

Electric utilities and local governments should similarly identify which proposed data centers have firm financial security, secured permits, and realistic load schedules. They should disclose how connection and network investment costs are allocated if a facility is delayed or cancelled. Long-term power procurement can be beneficial, but only if projects are not permitted to externalize large portions of their failure risk. The IEA's emphasis on integrated planning and better disclosure makes this a matter of infrastructure governance rather than simply investor sophistication. [19][21]

Finally, the ownership map should include who controls major decisions. Anthropic's founder rights and trust structure, OpenAI's partnership dependence, Nvidia's role in financing arrangements and the influence of large private-credit funds all affect the distribution of risk and authority. The public debate often asks who 'owns AI' as though ownership were a single share register. In practice it consists of overlapping rights to models, compute, capital, distribution, collateral and electricity. A company might own a model yet rent the machines that make it commercially usable. Following the money means following those overlapping dependencies.

26

The final buyer of intelligence

Every financing chain must eventually end in a paying customer.

At the end of the financial chain is not a processor manufacturer, a cloud landlord or another AI laboratory. It is a person or organization prepared to pay for useful work. A hospital may pay to reduce clerical burden, a pharmaceutical company to accelerate a research task, a manufacturer to improve maintenance, a business to write software, and a consumer to obtain assistance. Such purchases represent genuine demand. But the price each buyer is willing to pay is constrained by alternatives: internal software, smaller models, human workers, ordinary automation and the option not to deploy AI at all. The value produced by an AI system may exceed the revenue received by its supplier by orders of magnitude.

If the AI industry becomes increasingly competitive, its technology could permeate economic life at low prices. This outcome would be a social victory. It would not necessarily be a financial victory for the firms carrying the largest fixed commitments. If frontier intelligence remains scarce and reliably delivers extraordinary high-value outcomes, those commitments might instead be supported by expanding margins and enormous demand. The evidence by 2 November 2026 does not establish which market structure will prevail. It does establish that firms are committing money in advance of knowing the answer. [19][23][24]

The strongest conclusion is therefore not that OpenAI and Anthropic are doomed, nor that the AI boom is a scam. It is that the contemporary investment system has created a sharp asymmetry between commitments made now and economic value that must be captured later. The financial contracts exist; the construction schedules are real; electricity must be supplied; debt must be serviced. Yet the price and competitive structure of intelligence are evolving faster than the infrastructure surrounding it. If expected profitability disappoints, the losses could extend well beyond the original model developers through connected suppliers, creditors and asset owners. [24][26]

A final paradox remains. AI could advance medicine, science, education and industry so thoroughly that no reasonable observer would call the technology a failure. Governments and consumers might come to rely on systems descended from today's frontier models. Some of the corporations financing those systems might nevertheless disappear, merge or surrender control to creditors. The world might inherit affordable intelligence after a restructuring of the financial institutions that first made it possible. The decisive economic test is not how many billions the industry can announce. It is how much independent customers will pay, year after year, for successful outcomes — and whether that amount can ultimately support the enormous physical system now being built to deliver them.

CONCLUSION: technological success and financial success must be evaluated separately; neither a boom nor a collapse is established in advance.

27

Appendix A — Chronology, 2022–2026

A record of dated transactions and reporting, not a single causal narrative.

Nov 2022 Stanford benchmark baseline later uses roughly $20 per million tokens for GPT-3.5-level inference. [22]

2024 CoreWeave revenue reaches $1.915 billion; Anthropic 2024 revenue is approximately $386 million. [9][2]

Oct 2024 Stanford benchmark shows inference at comparable MMLU threshold priced around $0.07 per million tokens. [22]

Apr 2025 IEA publishes Energy and AI; estimates 2024 global data-center electricity use at 415 TWh. [20]

Sep 2025 OpenAI–CoreWeave contract expansion brings reported agreements to up to $22.4 billion. [12]

Dec 2025 CoreWeave ends year with $60.7 billion in remaining performance obligations. [9]

27 Feb 2026 OpenAI announces a $110 billion investment package at $730 billion pre-money valuation. [8]

2 Mar 2026 CoreWeave files 2025 10-K, revealing 67% revenue concentration in Microsoft. [9]

31 Mar 2026 OpenAI announces $122 billion committed funding at $852 billion post-money. [7]

9 Apr 2026 Meta signs further $21 billion CoreWeave AI-cloud agreement. [13]

16 Apr 2026 IEA projects 485 TWh data-center electricity demand in 2025 and 950 TWh in 2030. [19]

10 Jul 2026 OECD publishes analysis of model-price declines and market concentration. [23]

10 Aug 2026 Reuters reports Nvidia and six Wall Street partners planning major compute-financing platforms. [14]

24 Sep 2026 Reuters reports Oracle force-majeure notice concerning Project Jupiter power delays. [17]

29 Sep 2026 Reuters reports Anthropic $518 billion infrastructure commitments in pre-IPO documents. [3][4]

9 Oct 2026 IMF publishes technical note on AI financing and financial-stability risks. [24]

13–15 Oct 2026 IMF financial-stability publications emphasize AI market and circular-financing risks. [25][26]

15 Oct 2026 Reuters reports roughly $50 billion CoreWeave debt and higher credit-market stress. [10][11]

26 Oct 2026 Anthropic files public S-1 with SEC; Reuters details its financial results. [1][2]

27–29 Oct 2026 Reuters reports Moody’s credit warnings, Anthropic cost analyses and Wall Street data-center financing. [5][15][16]

30–31 Oct 2026 S&P revises Meta credit outlook; Reuters reports OpenAI financing-target revision. [18][33]

28

Appendix B — Claims and evidence matrix

What is documented, what is interpretation, and what remains unknown.

Claim	Classification	Basis and caution
Anthropic filed for an IPO on 26 October 2026	VERIFIED	SEC Form S-1; exact filing date and terms. [1]
Anthropic revenue grew exceptionally fast	VERIFIED	2024, 2025 and H1 2026 figures in filing. [1][2]
Anthropic lost $80.7B of operating cash in H1 2026	INCORRECT	Reported net loss mostly noncash valuation change; operating loss much smaller. [6]
All $800B of Anthropic capacity is due now	INCORRECT	Multi-year, different scopes and payment terms. [3][5]
Cloud partners are also investors and distributors	VERIFIED	Reuters analysis of cloud marketplaces and prospectus. [4]
OpenAI March funding was $122B committed	VERIFIED	Company announcement; not all means spent cash. [7]
CoreWeave depended on Microsoft for 67% of 2025 revenue	VERIFIED	2025 10-K. [9]
CoreWeave held around $50B debt in October	ATTRIBUTED REPORT	Reuters and lender analyses; categories matter. [10][11]
Nvidia $500B financing was fully deployed	INCORRECT	Reported target of financing platforms, not disbursed proceeds. [14]
AI financing connections may amplify shocks	SUPPORTED RISK	IMF modeling and technical note. [24][27]
AI end-customer revenue is broadly fabricated	NOT ESTABLISHED	No reviewed evidence supports this accusation.
Model inference at fixed capability has become dramatically cheaper	VERIFIED	Stanford and OECD measures; not all tasks. [22][23]
Electricity demand will exactly double by 2030	FORECAST	IEA central projection, not a certainty. [19]
AI laboratories are inevitably going bankrupt	NOT ESTABLISHED	A speculative scenario; outcome depends on margins, demand, financing.
A successful AI revolution can coexist with financial losses	STRUCTURAL INFERENCE	Economic mechanism; not a prediction for an individual firm.

29

Source notes and bibliography

Primary filings and institutional reports are preferred; media findings are attributed and dated.

1. Anthropic, PBC. “Form S-1 registration statement.” 26 October 2026. https://www.sec.gov/Archives/edgar/data/2130357/000162828026067168/anthropicpbc-sx1.htm

2. Reuters. “Anthropic’s IPO: rapid growth, spiraling costs.” 26 October 2026. https://www.reuters.com/legal/litigation/anthropics-ipo-close-look-ai-frontrunners-rapid-growth-spiraling-costs-2026-10-26/

3. Reuters. “Anthropic’s $518 billion buildout hinges on non-cancelable deals.” 29 September 2026. https://www.reuters.com/business/anthropics-518-billion-ai-buildout-hinges-largely-deals-that-cannot-be-canceled-2026-09-29/

4. Reuters. “Anthropic IPO prospectus lays bare dependence on Big Tech partners.” 29 September 2026. https://www.reuters.com/world/anthropic-ipo-prospectus-lays-bare-deep-dependence-big-tech-partners-2026-09-29/

5. Reuters Breakingviews / Karen Kwok. “Anthropic’s IPO is all about ignoring the future.” 28 October 2026. https://www.reuters.com/commentary/breakingviews/anthropics-ipo-is-all-about-ignoring-future-2026-10-28/

6. Reuters / Krystal Hu. “Should you invest in Anthropic? Five takeaways from its S-1.” 28 October 2026. https://www.reuters.com/technology/artificial-intelligence/should-you-invest-anthropic-five-takeaways-its-s-1-2026-10-28/

7. OpenAI. “OpenAI raises $122 billion to accelerate the next phase of AI.” 31 March 2026. https://openai.com/index/accelerating-the-next-phase-ai/

8. OpenAI. “Scaling AI for everyone.” 27 February 2026. https://openai.com/index/scaling-ai-for-everyone/

9. CoreWeave, Inc.. “Form 10-K, financial year 2025.” 2 March 2026. https://www.sec.gov/Archives/edgar/data/1769628/000176962826000104/crwv-20251231.htm

10. Reuters. “CoreWeave’s AI contracts and lenders financing its build-out.” 15 October 2026. https://www.reuters.com/legal/transactional/coreweaves-ai-contracts-lenders-financing-its-build-out-2026-10-15/

11. Reuters. “CoreWeave shares and bonds slide as AI-debt fears mount.” 15 October 2026. https://www.reuters.com/legal/transactional/coreweave-shares-bonds-slide-further-ai-debt-fears-mount-2026-10-15/

12. Reuters. “CoreWeave expands OpenAI pact with $6.5 billion contract.” 25 September 2025. https://www.reuters.com/business/coreweave-expands-openai-pact-with-new-65-billion-contract-2025-09-25/

13. Reuters. “Meta and CoreWeave deepen AI cloud partnership.” 9 April 2026. https://www.reuters.com/business/coreweave-signs-21-billion-ai-cloud-deal-with-meta-2026-04-09/

14. Reuters. “Nvidia partners with Wall Street giants on $500 billion AI financing effort.” 10 August 2026. https://www.reuters.com/technology/wall-street-giants-partner-with-nvidia-500-billion-ai-financing-deal-ft-reports-2026-08-10/

15. Reuters. “Wall Street’s AI bet: own the infrastructure, collect the rent.” 29 October 2026. https://www.reuters.com/business/finance/wall-streets-ai-bet-own-infrastructure-collect-rent-2026-10-29/

16. Reuters. “AI financing driving private-credit growth as refinancing risks rise.” 27 October 2026. https://www.reuters.com/legal/transactional/ai-financing-driving-private-credit-growth-refinancing-risks-rise-moodys-says-2026-10-27/

17. Reuters. “Oracle invokes force majeure over Project Jupiter power delays.” 24 September 2026. https://www.reuters.com/business/oracle-cites-force-majeure-shield-itself-controversial-data-center-bloomberg-2026-09-24/

18. Reuters. “S&P revises Meta outlook to negative on AI cash-flow worries.” 30 October 2026. https://www.reuters.com/business/sp-global-revises-outlook-meta-negative-cashflow-concerns-2026-10-30/

19. International Energy Agency. “Key Questions on Energy and AI — Executive Summary.” 16 April 2026. https://www.iea.org/reports/key-questions-on-energy-and-ai/executive-summary

20. International Energy Agency. “Energy and AI — Executive Summary.” April 2025. https://www.iea.org/reports/energy-and-ai/executive-summary

21. International Energy Agency. “Powering Artificial Intelligence.” 14 October 2026. https://www.iea.org/reports/powering-artificial-intelligence

22. Stanford HAI. “The 2025 AI Index Report — Research and Development.” 2025. https://hai.stanford.edu/ai-index/2025-ai-index-report/research-and-development

23. OECD. “Artificial Intelligence Markets: Recent Developments and Competition Issues.” 10 July 2026. https://www.oecd.org/en/publications/artificial-intelligence-markets_d531d73f-en/full-report.html

24. International Monetary Fund / Chen, Gautam and Wu. “The AI Ecosystem: Structure, Financing, and Financial Stability Considerations.” 9 October 2026. https://www.imf.org/en/publications/tnm/issues/2026/10/06/the-ai-ecosystem-structure-financing-and-financial-stability-considerations-579504

25. International Monetary Fund. “Global Financial Stability Report: Resilience, Risk, and Transformation.” October 2026. https://www.imf.org/en/publications/gfsr/issues/2026/10/13/global-financial-stability-report-october-2026

26. IMF / Kristalina Georgieva. “The World Can’t Afford to Get AI Wrong.” 15 October 2026. https://www.imf.org/en/blogs/articles/2026/10/15/the-world-cant-afford-to-get-ai-wrong

27. IMF. “Global Financial Stability Report, Chapter 1.” April 2026. https://www.imf.org/-/media/files/publications/gfsr/2026/april/english/ch1.pdf

28. NBER / Dillon, Jaffe, Immorlica and Stanton. “Shifting Work Patterns with Generative AI, Working Paper 33795.” 2025. https://www.nber.org/papers/w33795

29. NBER / Bonney and colleagues. “The Microstructure of AI Diffusion, Working Paper 35141.” April 2026. https://www.nber.org/papers/w35141

30. NBER / Yotzov and colleagues. “Firm Data on AI, Working Paper 34836, October revision.” 2026. https://www.nber.org/papers/w34836

31. NBER / Bick, Blandin and Deming. “The Rapid Adoption of Generative AI, Working Paper 32966.” 2024/2025. https://www.nber.org/papers/w32966

32. NBER / Bick and colleagues. “Mind the Gap: AI Adoption in Europe and the U.S., Working Paper 34995.” March 2026. https://www.nber.org/papers/w34995

33. Reuters. “OpenAI cuts latest funding target amid technology selloff.” 31 October 2026. https://www.reuters.com/business/openai-cuts-funding-target-by-10-billion-ft-reports-2026-10-31/

34. Reuters. “Neocloud firm Argentum AI in talks for $5 billion financing.” 27 October 2026. https://www.reuters.com/business/finance/neocloud-firm-argentum-ai-talks-5-billion-ai-chip-financing-sources-say-2026-10-27/

35. Reuters. “Musk’s warning hits neocloud stocks as debt concerns mount.” 26 October 2026. https://www.reuters.com/business/musks-warning-hits-neocloud-stocks-debt-concerns-mount-2026-10-26/

36. Reuters Breakingviews. “AI construction crunch widens credit fault lines.” 8 September 2026. https://www.reuters.com/commentary/breakingviews/ai-construction-crunch-widens-credit-fault-lines-2026-09-08/

37. Reuters. “Anthropic can do a better job of helping humanity.” 2 November 2026. https://www.reuters.com/commentary/breakingviews/anthropic-can-do-better-job-helping-humanity-2026-11-02/