HB Wealth

Bubbles and Breakthroughs: How Should Investors Respond?

By: Josh Rowe, CFA, PhD, Managing Director of Research & Family Office Investment Strategy, HB Wealth

This article is the fourth and final installment in our four-part series examining artificial intelligence (AI) through a wider economic lens. In the first three essays, we explored the history of AI and its recurring cycles of enthusiasm, the economics of technological adoption, and the implications for labor markets. If you have not yet read those pieces, we encourage you to start there for additional context.

In this final article, we turn to the question most relevant for investors: how to interpret current market movements, and how to position portfolios for a world of increasing rapid technological changes.

Executive Summary

Generative AI exhibits many of the characteristics of past technological revolutions—rapid capital deployment, rising valuations, stories about economic transformation. History is littered with examples of where revolutionary innovation was met with speculative excess and capital destruction. Where are there signs of a bubble? Which legacy businesses are at risk of disruption or obsolescence? Can investors still catch lightning in a bottle?

The question we face is not whether AI will change the ballgame – it is already doing so. The question is about who benefits and who pays the bill. Unlike railroads, electricity, or telecommunications, the AI's transformation of the economy is happening in real time, not over decades. Like the Internet before it, it is absorbing enormous amounts of capital and loosing tremendous experimentation in business models that will increase the mortality rate not just of startups, but of incumbent businesses. The scale of current investment in AI infrastructure is unprecedented, raising legitimate questions about whether future demand will justify the capital being deployed. While chipmakers and cloud providers have frantically increased spending to keep pace, there remain doubts concerning how value will ultimately be distributed across the AI ecosystem. It is possible that, on the whole, consumers benefit to a greater extent than investors.

Financial market innovation is evolving almost as rapidly as the technology it funds. Private credit, new forms of leverage, complex multi-party non-cash contracted purchase arrangements: all of these shift the risk and

opportunity calculus versus previous waves of technological change. We are, emphatically, in the early innings of the economic and commercial changes that AI and deep learning will wreak. Yet certain companies, both public and private, are valued as if the ultimate state of the world is guaranteed. Picking winners is a difficult game, even for venture capitalists and tech hedge funds. Investors should value flexibility and dry powder, even as they look to participate ratably in the early success stories.

About the Author

Josh Rowe, Managing Director of Research at HB Wealth, wrote a PhD thesis in the history and economics of technology, focusing on computer automation of office work in the 20th century. He has studied the history of AI, venture capital's funding of technological innovation, and the impact of technological change on financial markets—both as a resident of the ivory tower and as an investor. This surprising moment in history is the first time that he can say with any confidence that the years he spent in libraries and databases working on a doctoral dissertation might be of any practical use. He used AI in organizing and editing these essays, but the ideas (right and wrong) here are his own.

Introduction

The question of how investors should position themselves amidst a technological revolution is, in one sense, as old as capital markets themselves. The canal mania of the 1790s, the railway bubbles of the 1840s and 1870s, the electricity boom of the 1920s, the dot-com frenzy of the late 1990s—all are signature episodes in historians' study of the development of financial markets and their role in society. Each involved a transformative technology, fortunes made and lost, and a social process of disentangling hype, speculation, and the productive role of private capital in financing technological change. That generative AI fits into this lineage seems increasingly clear. The harder question is what, if any, lessons can today's investors infer from past stock market booms.

Since the public unveiling of ChatGPT in 2022, AI has been the word on every investor's lips. The theme of the AI trade has driven markets upward, turbocharged the market leadership of a narrow group of "AI stocks"—and increased valuation concentration among a few technology companies to historic levels. As the foremost mascot for the generative AI revolution, Nvidia's share price has risen nearly tenfold—an unprecedented gain for a company already among the world's largest. The "Magnificent Seven" technology giants have each added trillions of dollars in market capitalization. Venture capital has poured into AI startups at levels that would have seemed

hallucinatory a few years ago: OpenAI raised money at a $500 billion valuation, Anthropic has approached $350 billion, among a growing roster of unicorns in infrastructure, applications, and tooling.

The four largest American “hyperscalers”—Microsoft, Amazon, Alphabet, and Meta—are projected to spend nearly $650 billion on capital expenditures this year alone, most of it directed toward AI infrastructure, with commitments to spend substantially more in the years ahead. The four hyperscalers will spend more than the cost of the Apollo space program, adjusted for inflation.

To grasp the scale of what is underway, zoom in on a town called Ellendale in North Dakota—population 1,100, two motels, a Dollar General, a Pentecostal Bible college. It now hosts a half-built AI data center larger than ten Home Depots, with a price tag exceeding $15 billion—equivalent to a quarter of the state’s annual economic output. The mayor, Don Flaherty, recently took out loans to build sewers, sidewalks, and infrastructure for a planned neighborhood of twenty new houses to accommodate expected population growth. “We’re stepping out and taking a chance here," Flaherty told the *Wall Street Journal*, “and there’s a fear that everything could come crashing down” if the AI boom falters. Multiply Ellendale across dozens of similar projects in Texas, New Mexico, and the rural Midwest. The AI story will leave a substantial physical footprint for archaeologists of the future.

This exuberance for deploying capital is characteristic of the early part of what Carlota Perez calls a technological revolution’s “installation phase.” In Perez’s schema, transformative technologies pass through two broad periods: an initial installation phase marked by financial speculation, infrastructure buildout, and often a spectacular crash; and a subsequent deployment phase in which the technology diffuses broadly through the economy, its productivity-enhancing effects begin to be felt, and returns flow more to users than to the original investors. The canal companies that went bankrupt in the 1790s left behind a network of waterways that powered British industrialization for decades. The railroad investors wiped out in the Panic of 1873 had nonetheless laid the tracks that enabled continent-spanning markets. The fiber-optic companies that collapsed after 2000 had built the backbone of the modern internet.

The historical pattern suggests a sobering possibility: that transformative technologies can create enormous economic value while simultaneously destroying the wealth of those who financed their construction.

How The Pie Gets Sliced

For investors navigating the AI boom, the central question is not whether generative is vaporware, like previous hype cycles of consigned to the "trough of disillusionment" (wearables, VR, metaverse). It is how the value it creates will ultimately be distributed. To whom will the spoils flow? To the model builders, who develop the most sophisticated AI systems? To the cloud providers, who supply the computing infrastructure? To the power utilities who light the forges? To the chipmakers, who produces the specialized hardware? To the application developers, who builds the tools that enterprises actually use? To the enterprises themselves, who capture productivity gains? Or to consumers and workers, in the form of lower prices, better products, and new economic opportunities?

The most intellectually honest answer is that we do not yet know. Platform technologies often exhibit winner-take-most dynamics, network effects and high customer switching costs ensuring that value is concentrated among a few dominant players.[1] The history of digital technology offers numerous examples: Microsoft in PC operating systems, Google in search, Apple in smartphones, Amazon in e-commerce. If AI follows this pattern, the leading model providers or infrastructure players could capture enormous rents for decades.

But the current structure of the AI market looks more fragmented than these analogies indicate. Multiple frontier models jockey closely for performance leadership: OpenAI's GPT series, Anthropic's Claude, Google's Gemini, Meta's Llama, and a growing roster of open-source alternatives. The application layer is crowded and innovative. Enterprise customers, wary of lock-in after decades of experience with technology vendors, are actively managing their exposure across multiple providers.

Perhaps most importantly, large language models may lack the network effects that enabled previous platform monopolies. When I use Instagram, the value depends on who else uses Instagram; when stores list on Amazon, they go where the customers are. But whether I should use ChatGPT or Gemini does not depend on who else uses either—I will simply use whichever works better for my purposes. The switching costs are minimal. When ChatGPT goes down, I open Claude or Gemini and continue working. The interfaces are nearly identical. User feedback does not dramatically improve the models in ways that create lock-in.

[1] See our previous discussion of Metcalfe's Law: that the value of a network grows proportionately with the square of its number of users. This "law" suggests rapid, unpredictable and indeed unassailable advantages accrue to network owners who gain even a slight edge. It also suggests that network-driven technology platforms like social media, telecommunications infrastructure, Internet protocols or even app stores may be "natural monopolies" like railroads.

This absence of network effects raises the possibility that many investors have been slow to consider: that the AI industry may come to resemble airlines more than it resembles software. As Peter Berezin, chief global strategist at BCA Research, has observed, large language models "have become increasingly indistinguishable from one another. They may end up functioning more like highly competitive airlines with thin profit margins rather than monopolistic social media platforms." The comparison is not flattering. Jet aircraft are a massively transformational technology, enabling a reshaping of global transportation patterns and providing critical arteries of commerce. They require organizational and infrastructural complements that rewired the global economy. Airlines are essential, demand for air travel has grown for decades, and yet the industry has been a graveyard for investor capital—a business where everyone agrees the product matters and almost no one earns adequate returns.

An Infrastructure Extravaganza

The scale of capital being deployed into AI infrastructure is difficult to overstate. McKinsey estimates that nearly $7 trillion will be needed for AI-related data centers alone by 2030. Private equity firm Brookfield's CEO sees $5 to $10 trillion in total spending across data centers, power generation, and transmission. OpenAI itself has announced $1.4 trillion in infrastructure commitments and would spend more if it could find the capacity. Over the past three years, leading technology firms have committed more toward AI data centers, chips, and energy as a share of gross domestic product (GDP) than it cost to build the interstate highway system over four decades. These are sums, as one executive remarked, "that have never been invested before."

The numbers are large enough to raise a straightforward question: Where will the revenue come from to justify such expenditures? David Cahn, a partner at venture capital firm Sequoia, has attempted to reconcile the gap between infrastructure spending and plausible end-market demand. His analysis suggests that the money sunk into AI infrastructure just in 2023 and 2024 will require consumers and companies to purchase roughly $800 billion in AI products over the useful life of these chips and data centers to produce a good investment return. Most AI processors have a useful life of only three to five years. Bain & Company estimates that the wave of infrastructure spending will require $2 trillion in annual AI revenue by 2030. For context, that figure exceeds the combined 2024 revenues of Amazon, Apple, Alphabet, Microsoft, Meta, and Nvidia, and represents more than five times the size of the entire global subscription software market. Morgan Stanley estimates that actual AI product revenue last year was approximately $45 billion. The current arithmetic is not encouraging. Either AI adoption must accelerate dramatically, or a substantial portion of current investment will prove uneconomic.

The hyperscalers—Microsoft, Amazon, Alphabet, Meta, and increasingly Oracle—are financing this buildout through a combination of operating cash flows and rapidly expanding debt issuance. Amazon raised $15 billion in bonds in late 2025, Alphabet $25 billion, Meta $30 billion, Oracle $18 billion. During the first half of the year, investment-grade borrowing by technology firms ran 70 percent higher than in the first six months of 2024. Capital expenditures at these companies have increased more than thirteenfold over the past decade. Add the capital expenditure of Alphabet, Meta, and Microsoft during the past year to that of Amazon and Oracle, and the sum exceeds the outlay of all America's listed industrial companies combined. Microsoft's capex now consumes 25 percent of revenue, more than triple the ratio a decade ago. Alphabet recently issued up to $32 billion in bonds to finance its ambitious datacenter plans. Companies that were long lauded for network-driven free cash generation—likened by investors to ATM machines—are now engaged in a progressively more capital-intensive business.

This transformation from asset-lite to high capital intensity business models represents a fundamental shift in the economics of technology investing. For two decades, the giants of Silicon Valley built their dominance on an enviable formula: create disruptive innovations, deliver exceptional growth, and keep capital requirements minimal. Software, at its best, is a marvelous business—once written, it can be copied infinitely at near-zero marginal cost, generating the extraordinary margins that made technology stocks the darlings of growth investors. The AI era threatens to invert this logic. Training frontier models requires billions of dollars in compute. Running those models for users—inference, in the industry parlance—requires ongoing expenditure on chips, electricity, and cooling that scales directly with usage. The data centers housing this infrastructure depreciate rapidly as newer, more efficient hardware becomes available; functional depreciation schedules measured in a few years compared to decades for traditional industrial equipment.

In fact, asset lives in this cycle may be shorter than in previous technology buildouts. The Economist estimates that the average American technology firm's assets now have a shelf life of just nine years, compared with fifteen for telecommunications assets in the 1990s. An unusually large share of current capital investment is being devoted to assets that depreciate quickly. Nvidia's cutting-edge chips will inevitably look clunky in a few years; the company itself has announced it will release new chip generations annually rather than every two years. Jensen Huang, Nvidia's CEO, remarked that "when Blackwell starts shipping in volume, you couldn't give Hoppers away"—referring to the company's latest chips and their predecessors. This dynamic creates a treadmill effect—spending large sums just to maintain, upgrade, or replace equipment that was purchased only a few years ago.

When the market is highly competitive—or at least, oligopolistic—technological arms races are likely to ensue. Spend more today or your competitor will—and in so doing make all your accumulated capital investment worthless.

Since 2020, hyperscalers have systematically extended the depreciation lives of their graphics processing unit (GPU) and networking assets—from three years to five or six years at most major companies. The rationale is that older chips remain useful for inference and other tasks even after newer generations arrive. Perhaps. But if the true economic lifespan of these assets is closer to two or three years, then reported profits are materially overstated. Jim Chanos, a veteran short-seller, has argued that if Meta's AI chips depreciate over two to three years rather than five and a half, "most of its 'profits' are materially overstated." Analysis by Barclays suggests that more realistic depreciation assumptions could reduce earnings per share at Alphabet, Amazon, and Meta by 5 to 10 percent. Apply the same logic across all five major AI hyperscalers, and the potential hit to combined annual pre-tax profits could reach $26 billion, or roughly 8 percent of their total. At current market multiples, that implies a valuation adjustment measured in hundreds of billions of dollars.

Financial Engineering: Follow the Cash

Technological innovation at the production possibilities frontier is often accompanied by parallel innovation in accounting and finance. How bankers, VCs, and accountants carve up the capital structures of the leading companies, and how they structure their contractual relationships can sometimes create or destroy as much economic value as the technology itself.

The financing structures emerging to support the AI buildout have begun to attract scrutiny from credit analysts and regulators alike. Meta financed a massive Louisiana data center through an off-balance-sheet joint venture with Blue Owl, a private equity firm, borrowing $27 billion through a special-purpose vehicle called Beignet Investor while taking only a 20 percent minority stake. The arrangement was structured to avoid affecting Meta's investment-grade credit rating—the periodic payments and residual value guarantee effectively function as a triple-net lease in most scenarios, though the complexity of the structure obscures whether Meta should be viewed as entirely on the hook for this borrowing. Fixed-income managers who can invest across public and private debt nonetheless view Beignet as a Meta subsidiary; having committed billions to the bonds, they found themselves needing to reduce risk by selling their direct Meta holdings.

Oracle presents an even more aggressive case. The company has made over $100 billion in capital commitments to lease data-center shells for the Stargate project, a collaboration with OpenAI and SoftBank to rapidly build AI infrastructure across the United States. These obligations are only beginning to appear on Oracle's financial statements; the company has more than $16 billion of operating lease liabilities on its balance sheet, plus almost $250 billion of lease commitments that have yet to begin. Capital expenditures are expected to reach 138 percent of operating cash flow in the coming fiscal year, a level of spending intensity that dwarfs even Meta, the next most aggressive hyperscaler at 84 percent. Oracle's debt has risen above $100 billion, pushing it toward the more indebted end of investment-grade borrowers. As a cuspy, near-junk rated issuer racing to catch up to peers in AI cloud hosting, Oracle is an unusually perfect bellwether for the bond market's appetite to finance the gen AI revolution: its credit default swaps (contracts that insure bond investors against losses) have become much more expensive, and continue to swing wildly in the winds of market sentiment.

Oracle's fate is now tightly bound to OpenAI, which accounts for more than half of the company's $523 billion in contracted future revenue. OpenAI has committed to pay Oracle $60 billion per year—an amount OpenAI does not yet earn in revenue—to provide cloud computing facilities that Oracle has not yet built, which will require 4.5 gigawatts of power (the equivalent of 2.25 Hoover Dams or four nuclear plants) not yet secured. But that $300 billion contract is reportedly spread over only five years, while Oracle's property leases typically run at least twice that long. Should OpenAI fail to meet its contractual obligations, a possibility that even bullish analysts acknowledge, given the startup's enormous cash burn and uncertain path to profitability—Oracle could find itself with costly rent payments to honor and a large hole to fill. Gil Luria, head of technology research at DA Davidson, called the OpenAI contract "fantasy" and said the likelihood that OpenAI will have $300 billion to spend is "negligible at this point." According to analyst projections aggregated by Bloomberg, Oracle's infrastructure bet is expected to pay off only after absorbing approximately $70 billion in free cash flow losses over the rest of the decade.

To secure power quickly in a market where utilities have long waitlists, Oracle has turned to novel and costly expedients, including running entire sites with gas generators. For one data center campus in rural Shackelford County, Texas, this approach will cost more than $1 billion per year. The suspicion, as Bloomberg's Chris Bryant put it, is that Oracle "has effectively allowed OpenAI to lean on its balance sheet. By getting Oracle to do the hard and costly legwork of setting up and running the data centers that it will use, Sam Altman's startup doesn't have to raise as much cash."

How the Oracle story plays out may provide no more than a footnote in the historical record. But it is emblematic of the bold "bet-the-company" moment that is sweeping Silicon Valley. And it illustrates the fragilities of a financial and contractual structuring that works well so long as all investors are sold on the prevailing narrative of a profits bonanza just over the horizon.

Lenders are slicing and securitizing data-center debt, spreading risk through the financial system in ways that make exposures increasingly opaque. The market for debt securities backed by data-center borrowing has grown from almost nothing in 2018 to around $50 billion today. Some borrowers have reportedly sought loans exceeding 100 percent of construction costs, justifying the requests on anticipated valuation uplifts when facilities begin generating revenue. The Bank of England has begun reviewing lending to data centers amid concerns about the level of spending and financing. The asymmetrical logic of debt investing vs. venture capital invites questions. Howard Marks of Oaktree Capital, in a widely circulated memo, questioned whether it was "prudent to accept 30 years of technological uncertainty to make a fixed-income investment that yields little more than riskless debt." It is a truism of capital markets that, in boom times, deal volume, market share considerations, and bankers' arranging fees can drive capital allocation as much as intrinsic value analysis.

Credit markets, which historically have served as early warning systems for technology buildout cycles, are beginning to register these concerns. The spread on Oracle's 30-year bonds has roughly doubled since issuance, from about 1.2 percentage points over Treasuries to more than 2.1 percentage points. CoreWeave, a "neo-cloud" company that purchases advanced chips and rents them to AI developers, has seen its bond spreads widen to nearly 8 percentage points over Treasuries, with credit default swaps soaring in tandem. When Amazon sold $15 billion in notes in late 2025, investors withdrew 40 percent of their orders when the final pricing failed to offer enough yield. Bank of America analysts have noted that "today's hyperscalers, viewed in the same framework, look eerily similar to early-stage telcos of '98"—a comparison that should give pause to anyone who remembers how that episode concluded. Selloffs in telecom bonds preceded sharp corrections in stock prices by months.

"WeWork on Steroids: Neo-clouds and Duration Mismatch"

Among the most precarious participants in the AI buildout are the "neo-cloud" companies—firms like CoreWeave that lease data centers, fill them with Nvidia chips, and rent server capacity to AI developers. CoreWeave's trajectory illustrates both the opportunities and the vulnerability of this model. Six years ago, the company was an obscure cryptocurrency miner with fewer than two dozen employees, operating out of a drab New Jersey office park next to a Container Store and a waxing salon. Flooded with money from Wall Street and private-equity

investors after ChatGPT's release, it has metamorphosed into a computing goliath with a market value larger than General Motors or Target.

The company's CEO, Michael Intrator, a former commodities trader, has called debt "the fuel for this company." CoreWeave has accumulated roughly $15 billion in debt, with interest rates starting above 8 percent on deals with top technology companies and far more for upstarts. As of late 2025, CoreWeave had racked up over $42 billion worth of contracts with technology companies renting its servers.

But the company's filings reveal an eyebrow-raising duration mismatch. CoreWeave owes $56 billion in payments for data-center leases, which typically run around ten years. Its deals with technology customers are typically for two to five years. When those contracts expire, CoreWeave will still owe billions in lease payments regardless of whether it has found new customers. If the wave of building proves far more than needed, or if technology companies pivot away from third-party providers, CoreWeave's data centers could end up like the dormant fiber-optic cables that snaked through the United States after 2000—expensive monuments to misplaced optimism.

Intrator defends the approach. The high financing costs are "the tuition you pay when you build something new, and we paid that tuition to get in early," he told the *Wall Street Journal*. "I'm not going to tell you there's no risk. But we've been incredibly thoughtful about how we've mitigated that risk and structured that debt so that it's appropriate for this technology." Perhaps. But the company's stock has fallen more than 50 percent from its June 2025 peak, though it remains up 90 percent from its March IPO price—a pattern that suggests investors are reassessing the risks even as they remain hopeful about the upside. Tuition may be the price for early leadership—Intrator is consciously playing the game of establishing first-mover advantage in a network effect-driven competitive market. But if the economics prove different this time, or if the technology evolves in unpredictable ways, as many recent college graduates have found, tuition can also be wasted.

Oracle, the new AI datacenter bellwether, is the subject of similar investor skepticism. The structural mismatch between the amount and timing of the rents it pays—$248 billion in a recent data release—and its contracted revenues has invited comparison not to Cisco or AOL or other darlings of the Internet age, but to a yet more speculative business model: WeWork. When WeWork attempted to go public just before the COVID pandemic, its dressed-up arbitrage of renting and subleasing space was seen less as a technological innovation and more as a leveraged bet on continued investor enthusiasm and an unhedged duration and liquidity mismatch worthy of an early-2000s hedge fund. The build-it-and-they-will-come ethos of the spendthrift neoclouds, and the structural

asymmetry between leases and chip purchase contracts, and the (frequently non-cash) revenue promised by the model builders has many technology watchers feeling jittery.

Utilities, conservative by nature, have already begun to think twice about signing long-term energy contracts with these newer entrants. “You do not know which of these players will be around in five, ten or 15 years’ time,” comments Pankaj Sachdeva, a partner at McKinsey. In response, insurance policies, securitizations, and other risk-mitigation structures are being designed to reassure counterparties. Nvidia has pitched in with a web of vendor financings and cross-investments. But if the worst happens, such incestuousness will increase the vulnerability of the AI ecosystem as a whole.

A Circular Economy

More troubling than the scale of investment is the self-referential quality of much of the early financial activity. A significant portion of AI revenue, particularly that reported by the hyperscalers and model builders, comes from deals with other participants in the AI ecosystem itself: cloud credits extended to startups, infrastructure partnerships among the major players, and contracts where consideration often takes the form of equity or deferred commitments rather than upfront cash.

Such circularity is pervasive. Microsoft’s multi-billion-dollar investment in OpenAI included a substantial component of Azure cloud credits; OpenAI spends those credits on Azure compute, which then shows up as Microsoft’s AI revenue growth. Amazon’s $4 billion investment in Anthropic came with a commitment to make AWS the primary training partner, along with hundreds of millions in additional cloud credits to AI startups that will, in turn, spend on AWS. Google has invested up to $3 billion in Anthropic, locking in training and inference contracts. Nvidia has made equity investments in companies that are, simultaneously, its largest customers. The chip giant’s $6.3 billion backstop agreement with CoreWeave obligates Nvidia to purchase any unsold cloud capacity through 2032, while also being a major supplier and equity holder.

In November 2025, this circularity reached a new crescendo when Microsoft and Nvidia announced a joint $15 billion investment in Anthropic—which, in turn, committed to spend $30 billion on Microsoft’s Azure cloud platform, underpinned by Nvidia’s chips. At one point, markets celebrated these agreements as evidence of a demand lift-off. Now they are looked at with more intense scrutiny; Microsoft’s stock fell 3 percent after the announcement. A sell-off in technology shares dates back to September 2025, which coincided with announcements from OpenAI agreeing to spend $300 billion over five years on computing power from Oracle, and

Nvidia promising to invest up to $100 billion in OpenAI. That marked the start of $1.4 trillion in spending commitments by OpenAI, a plan that has belatedly focused attention on the growing bill to deliver the model training and inference capabilities required to stay competitive with rivals like Google and Anthropic.

These sorts of arrangements engender what might generously be called a web of dependencies or, less charitably, "round-tripping." When both sides of a transaction are owned or reported by overlapping sets of investors, distinguishing genuine customer demand from financial engineering becomes trickier. Unless investors scrupulously peel back layers of the revenue onion, there is a likelihood of collective double-counting. The situation conspicuously rhymes with the late 1990s' fiber-optic, when telecommunications carriers booked revenue by swapping capacity with one another, creating an illusion of demand that evaporated when sentiment shifted. The chain of mutual interdependence can amplify small movements affecting any link. As John Plender writes in the Financial Times, "There are echoes here of the behaviour of banks and insurance companies in credit derivatives before the financial crisis of 2007-09."

While the possibility of fraud exists, and boom times always expand the space for what economist John Kenneth Galbraith called "the bezzle" (undiscovered losses that investors obliviously celebrate as revenue or assets), fraud is not our claim. Technology firms' accounting practices appear to be GAAP-compliant. Defenders argue that cross-investments represent legitimate ecosystem building—partnerships among big players are inevitable when technologies are immature and monetization lags. But non-cash deals, unearned revenue, and share grants also blur visibility into the underlying demand for AI services and create the possibility for valuation contagion. If confidence wavers in one part of the ecosystem, the effects could propagate quickly to others.

Everything Hinges on the Pace of Adoption

Assume, for the moment, that all the infrastructure is built and can draw power from the grid. Assume that the financing holds together, that the circular arrangements do not unwind catastrophically, that data centers come online and the supply of next-generation chips keep flowing. A deeper question remains: Is anyone actually using this technology in ways that would justify all the investment?

The early evidence is equivocal at best. A July 2025 study by MIT found that despite $30 to $40 billion in enterprise investment in generative AI, 95 percent of organizations are getting zero return. Successful deployments are moving forward at a small fraction of the rate of proof-of-concept (POC) programs and piloting. In 2025 we saw many companies play around with AI, finding it wanting for production use cases, and quietly shelving the

initiative. Data suggests that 42 percent of companies abandoned the majority of their AI projects in 2025, up sharply from 17 percent the year before. The share of CEOs who are "very confident" in their AI strategy has fallen from 82 percent in 2024 to 49 percent in 2025. Feedback from respondents was telling: "It's excellent for brainstorming and first drafts but doesn't retain knowledge of client preferences or learn from previous edits. It repeats the same mistakes and requires extensive context input for each session. For high-stakes work, I need a system that accumulates knowledge and improves over time."

Assaf Araki of Intel Capital has drawn attention to a pattern of corporate behavior that may distinguish AI adoption from the software-as-a-service businesses that defined the previous era of enterprise technology. SaaS revenues, once established, tended to be sticky—organizations-built workflows around systems, trained employees, integrated data, and renewed subscriptions year after year almost automatically. AI usage, by contrast, has often been exploratory: companies purchasing tokens or API access to experiment with use cases, running pilots, testing capabilities. This "experimental revenue" may not convert to the kind of committed, budget-line-item spending that sustains enterprise software businesses. Consequences for software pricing are intriguing: this terrain favors upstart, lean disruptors with competitive or even usage- and success-based pricing strategies. But it also means that the valuations attached to "contracted" recurring revenue should be adjusted accordingly. Organizations that cannot clearly demonstrate how agentic applications are driving measurable value and reducing costs will decline to make the long-term financial commitments necessary to validate vendors' heavy fixed costs.

Should the nascent experience of tinkering and piloting persist longer than the installation phase, if experimental usage fails to convert to recurring revenue at scale, then the demand projections underlying current infrastructure investment will prove overstated. David Cahn's estimate of $800 billion in AI product purchases that David Cahn calculates would be required to justify 2023-2024 infrastructure spending depends on a relatively rapid pickup in committed adoption. That transition is not yet evident in the data.

Indeed, surveys point to stagnating corporate adoption. According to Stanford University researchers, 37 percent of Americans used generative AI at work in September 2025—down from 46 percent in June. A tracker by the Federal Reserve Bank of St. Louis found that 12.1 percent of working-age adults used generative AI every day at work in August 2024; a year later, 12.6 percent did. Ramp, a fintech firm, reported that AI use at American firms soared to 40 percent in early 2025 before leveling off. The growth in adoption really does seem to be slowing.

One possible explanation is economic uncertainty—trade wars, immigration restrictions, and an unclear interest-rate outlook may be causing businesses to defer investment until conditions stabilize. History is replete with adoption cycles that proceed in fits and starts. But there is also ample evidence of the familiar lags in enterprise user penetration. Almost everyone in senior management sings the praises of AI; in recent earnings calls, nearly two-thirds of executives at S&P 500 companies mentioned AI. At the same time, the people actually responsible for implementing AI may be less keen early adopters—perhaps because they worry about the technology putting them out of a job. A survey by Dayforce found that while 87 percent of executives use AI on the job, just 57 percent of managers and 27 percent of rank-and-file employees do. Perhaps middle managers set up AI initiatives to satisfy their superiors' demands, only to wind them down quietly at a later date.

There is also evidence that changing perceptions of AI's usefulness may be dampening adoption. A poll of executives by Deloitte found that 45 percent reported returns from AI initiatives were below expectations; only 10 percent reported expectations exceeded. McKinsey has argued that for most organizations, AI use has not yet significantly affected enterprise-wide profits. Goldman Sachs produced an index of companies with the "largest estimated potential change to baseline earnings from AI adoption via increased productivity." Thus far, these firms have failed to see significant earnings uplift and their share prices have lagged the market. Goldman expects that the phase of earnings benefit from AI users is yet to begin.

Markets have entered into a high-stakes, highly leveraged gamble. When forward contracts are capitalized into value, when exuberant stock markets and late-stage venture rounds embed aggressive expectations in valuation, when the building of data centers is financed with enormous amounts of debt (that will require refinancing at maturity), and when vital equipment is depreciated on schedules ranging from two to six years, the risks of near-term disappointment are magnified. Much of the financial apparatus supporting the generative AI rollout is exposed to the outcome that the anticipated end-user revenue shows up, and shows up on schedule. In a previous section we discussed productivity lags and the pace of technological adoption. Current-generation AI applications have advantages versus previous technology platforms with respect to the speed of diffusion. But if history is any guide, markets are prone to getting ahead of themselves. Discount rates reflect confidence in the future. When they are too low, capital flows too recklessly and ends up being misallocated.

The Uncertain Path to Profitability

Let's make another assumption. Not just that the capital-intensive installation phase succeeds in bringing online all the infrastructure required for an AI rollout. But, let us further assume that enterprise customers find legitimate

value and productivity enhancements from embedding AI-native and agentic workflows into their core activities. Will the generative AI paradigm shift then usher in a golden era for investors? There is every reason to suspect that even in this blue-sky scenario, today's investors fail to reap long-term rewards. The historical record is littered with transformative technologies that have generated enormous economic surplus while leaving their financiers with mediocre or worse returns. In the 19th century railroads reshaped the American economy, enabling continent-spanning markets and accelerating industrialization. But consumer surplus from rail transport dwarfs any reasonable estimate of the profits earned by railroad companies themselves. Railroad investors, famously, did poorly—a chronicle of overbuilding, rate wars, bankruptcies, and consolidations extending across decades. The technology was revolutionary; associated investment returns were not.

Andrew Odlyzko, an emeritus mathematics professor at the University of Minnesota who has studied financial manias from nineteenth-century railways to the fiber-optic boom, calls the unbridled optimism in such episodes "collective hallucinations"—periods when investors assemble in herds, willfully blind to obvious risks. Odlyzko's interest in technology bubbles stems from firsthand experience. From the 1970s through the early 2000s, he spent time as a researcher at the greatest innovation factory America has ever produced—Bell Labs—during a period in which telecom giants and upstarts alike raced to bury tens of millions of miles of fiber cable into the ground, spending the equivalent of around 1 percent of U.S. GDP over half a decade. The prevailing belief was that internet traffic was doubling every hundred days. In reality, for most of the 1990s boom, traffic doubled every year—a yawning delta for investors discounting future revenue to the present.

A historical parallel holds for electricity, automobiles, commercial aviation, as for other telecommunications technologies. Each industry overhauled the economy's physical backbone, leveled-up its core infrastructure, generated vast economic value, and enriched society broadly. In every case, the investors who gamely financed that installation phase captured at best a tiny fraction of the surplus, after losses to competition, obsolescence, and miscalculation. The fiber-optic cables laid in the late 1990s—much of which sat dark for years after the telecom bust—eventually carried the traffic of the modern internet. Investors betting on their construction were largely wiped out. These cables' value flowed to the likes of Google and Netflix, and to the billions of people streaming video and ordering pet food all over physical toll-roads later scavenged out of bankruptcy proceedings for pennies on the dollar.

The bull case rests on the assumption that AI will follow the example of earlier software platforms—that network effects, high switching costs, and economies of scale will accumulate value among a small cadre dominant player

who can sustain high margins and fend off competition. But the current structure of the AI market, as we have seen, suggests an altogether different precedent.

Consider the economics already visible in the industry. When Oracle's GPU rental margins became public, they revealed a worryingly thin 14 percent gross margin on Nvidia-powered cloud instances, with some operations running at a loss. The operating expense of renting and running the physical equipment is nearly as high as reported revenue, the reality of a land grab where providers sacrifice profitability for market share. Google, which has invested heavily in custom TPU chips optimized for its own models, has managed to reduce the cost of an AI query to roughly twice that of traditional search—impressive engineering, but still uncomfortable margin compression when compared to its legacy business. OpenAI, despite approximately $13 billion in annual revenue, continues to operate at a substantial loss. Meanwhile, GPU hourly rental rates paid by neo-cloud and hyperscaler customers have declined by 20 to 25 percent over the past year—a trend that suggests pricing power may be eroding even as the infrastructure buildout accelerates.

Subscription software is defined by wonderful unit economics—vanishingly small variable costs required to ship an incremental product. Yes, customer service, maintenance engineering, and go-to-market activities imply some ongoing operating expense, but, like fixed costs, these are proportionately reduced when spread over higher volume. High retention, increased cross-sale, and wallet penetration opportunities meant a strong positive relationship between size and profitability. Hence, software companies, even during their growth stage, were rewarded with extravagant valuation multiples—expectations that combined the revenue stickiness and predictability associated with monopolistic utilities but with near unlimited pricing power and uncapped growth rates. When Marc Andreessen remarked that "software is eating the world," he implicitly was talking about more than the ubiquity of code in our everyday economic lives—he could equally have been referring to the stock market.

Generative AI at present has none of these investor-friendly attributes. As we have seen, users are fickle and conservative with committing their capex budgets. More importantly, the unit economics, expressed in gross margins, are decidedly poor when compared to what the tech giants of the last decade were used to. Training new models requires immense compute, is insatiably thirsty for power, and has swallowed nearly all the data the Internet can provide. These demands run up the fixed cost tab, as we have written. But reasoning and inference—processing user queries or agent's activities—into useful output lights up those stacks of Blackwell processors anew. As more users sign up, the loads on the system increase, as do the model providers' costs. There is no

scaling effect, at least not yet observed. As the old joke goes, “We lose money on every sale, but we make it up in volume!”

Bulls counter that these are start-up economics. Amazon lost money on shipments until they were able to build scale and cross route-density thresholds; Uber and Lyft lost money on every ride until their offerings became so ubiquitous that they could establish pricing power. To the AI optimists in Silicon Valley, margins will reliably improve as the technology matures and economies of scale begin to kick in. Training costs are one-time expenditures that will eventually be amortized across an enormous user base. “Blitz-scaling” is how the global tech mega-platforms of the last generation were built. It’s hard-wired into the Silicon Valley ethos of huge risk, huge losses, but even huger upside.

The problem with this argument is that it presupposes a war of attrition where the capital requirements never level off. It ignores the constant upgrade cycle and the competitive arms race for model performance and data aggregation. Yes, more innovative, leaner training algorithms and inference engines are in development, but that is not where OpenAI, Oracle, Meta and others are placing most of their chips. One could easily imagine a counterargument where competition actually intensifies as the technology commoditizes. Efficiency gains are likely to be contested most aggressively by the lowest-cost providers. In this type of competitive industry surplus will flow to users rather than shareholders. AI could easily that the AI industry will come to resemble other essential-but-unprofitable sectors. Venture capitalists do not have a lot of recent experience in markets that look this way, and their memories might be shorter than their rhetoric.

AI Beyond the Screen

Most discussion of generative AI focuses on its applications in knowledge work—writing, coding, research, analysis. But some of the most consequential long-term value may be found in domains that receive less attention: neural nets and deep learning are poised to make major strides in automating or accelerating parts of the physical world.

Healthcare offers many compelling current examples. AI systems are already demonstrating capabilities that meaningfully improve clinical outcomes and reduce costs. A language model designed for synthesizing medical research reproduced a systematic review in two days that would have taken human researchers twelve work-years; it identified relevant studies more accurately than human counterparts and extracted data more reliably. In administrative functions—revenue cycle management, claims processing, prior authorization—early adopters

report substantial savings. Provider networks are reporting benefits. Beacon Health System saved $95 million using AI to streamline medical necessity reviews and documentation. University of Pittsburgh Medical Center saved $6.2 million using AI tools to analyze patient risk factors, reducing time spent on ventilators by more than 2,000 cumulative days.

Drug discovery and clinical trial design represent longer-term opportunities. The ability to simulate molecular interactions, predict drug efficacy, and identify promising compounds could substantially reduce the time and cost of pharmaceutical development—a process that currently averages over a decade and more than $2 billion per approved drug. Computer-aided surgery, pathology analysis, and diagnostic imaging are additional domains where AI capabilities are advancing rapidly. If these applications mature as proponents hope, the societal value created could dwarf the productivity gains from automating white-collar tasks.

Physical-world applications remain in early development, but AI will progressively be embedded in offline activities. Manufacturing, logistics, and mobility are areas of significant AI deployment. Advances in robotics, including specialized industrial robots, drones, and advanced mobile robots, are opening new frontiers of AI penetration into sectors that have resisted digitization. Warehouse automation, agricultural drones, and autonomous vehicles are examples of low-latency inference use cases where costs are likely to come down and feasibility increase.

For investors, the implications are twofold. First, the AI story should not be circumscribed to office tasks, chatbots, and virtual agents. The need for computing infrastructure will likely be more pervasive than from software alone. Enterprise adoption, then, could disappoint while industrial take-up surprises. Second, the companies best positioned to capture value from AI in healthcare, manufacturing, and logistics are unlikely to be the current leaders in foundation models—they may be specialized application developers, incumbent industry players who successfully integrate AI into production lines, or disruptive firms that do not yet exist.

AI Beyond San Francisco

Virtually every discussion of AI investment assumes that the current leaders—American hyperscalers and chipmakers—will maintain technological supremacy and pricing power. Data centers' voracious appetite for power and compute will continue to drive growth and occupy an increasing share of the U.S. economy. Economist Jason Furman of Harvard pointed out that, without data center capex, U.S. real GDP growth would have been 0.1%

in the first half of 2025. Or, as veteran technology watcher Paul Kedrosky jibed, "Honey, AI capex is eating the economy." These trends merit further examination.

In a recent podcast, tech pundit Scott Galloway argues that cost leadership and pricing will prove more important in 2026 than product leadership. He predicts that Chinese firms will engage in "dumping"—flooding Western markets with capable, cheap models that undercut incumbents. "Dumping" is a bit of a euphemism: in trade law, the terms refers to goods sold below cost in trade competition, and WTO anti-dumping rules do not cover services like AI APIs. But Galloway points out a structural similarity between China's AI strategy and its goals for industrial supremacy such as steel, solar panels, and electric vehicles. This behavior, with its non-economic rationale, may undermine that capital intensive model pursued by firms like OpenAI and Anthropic.

There is evidence for Galloway's thesis. In early 2025, DeepSeek, a Chinese AI lab spun out of a quantitative hedge fund rather than a major research institution, demonstrated that competitive reasoning models could be built at dramatically lower cost[2]. When DeepSeek-R1 launched in January 2025, Nvidia shed approximately $600 billion in market value in a single day—the largest one-day loss in stock market history—on fears that demand for its frontier chips might soon run dry. The shares later partially recovered, but the episode illustrated how sensitive valuations are to assumptions about pricing power and competitive moats.

DeepSeek claimed it could train competitive models on roughly 2,000 H800 GPUs with a headline compute bill of $5 to $6 million. Later analysis suggested that total spending—including servers, operations, and multiple training runs—was far higher, with estimates ranging toward $1.3 to $1.6 billion. Even at that cost, DeepSeek represents a proof point that endless capital investment might not be the optimal competitive strategy. DeepSeek's API pricing is frequently cited at below $0.50 per million input tokens—an order of magnitude cheaper than OpenAI's GPT-4o at $2.50, or Anthropic's Claude Opus at $5.

Nor was DeepSeek a flash in the pan. Alibaba's Qwen family and other Chinese models have followed a similar path: solidly capable, cheap, and increasingly used by Western firms. Brian Chesky, Airbnb's CEO, said publicly that the company "relies a lot" on Qwen for its customer service applications because it is "fast and cheap"—running a mix of models including OpenAI, Google, and Alibaba depending on the task. Chinese models like

[2] The true costs of training DeepSeek's R1 are contested, maybe by several orders of magnitude. But the fact remains that the model performed acceptably well on a range of reasoning challenges, and did so without access to Nvidia's state-of-the-art Blackwell chips.

Qwen tend to be “open-weight,” which means that their parameters are publicly released, allowing firms to configure and fine-tune them to their own needs, run the model locally for better security, use compressed or less compute-intensive versions, and avoid being locked into a single vendor on a price-per-token contract. Galloway cited a striking statistic: “80 percent of a16z startups use open-source Chinese models.” The figure requires context— Martin Casado, the Andreessen Horowitz partner whose quote was paraphrased, later clarified he meant 80 percent *of the 20 to 30 percent* of applicants that use open-source models, so more like 15-20 percent of all startups. Nevertheless, if the most innovative firms in the U.S. are relying on low-cost foreign models, we should not be surprised to see enterprise behavior follow. In a setting where piloting and experimentation predominate, cheaper, more flexible inputs are likely to be preferred.

China has pursued AI development with the intensity of a strategic national priority, not merely a commercial opportunity. Chinese government subsidies can reach 15 percent of industry profits in the semiconductor sector—a level of state support that American and European competitors cannot match. China’s industrial policy spending as a share of GDP dwarfs all other major economies, employing tools including state investment funds, state-subsidized credit, compute vouchers, tax grants, and “AI-plus-manufacturing” plans that encourage AI to permeate across the real economy. These policies emphasize open-weight and low-price models by design; the goal is not to maximize profits for any individual Chinese company but to establish technological capacity and global market position. There are precedents for openness leading to more widespread adoption of technologies and standards that were technically inferior to proprietary designs: VHS’s broader licensing agreements than its rival Betamax, and the IBM PC’s open developer architecture versus Apple’s walled system. Both encouraged low-cost competition and established dominant industry standards. This may be China’s aim.

Across the AI value chain, Chinese competition will have an uneven impact. Upstream suppliers—chipmakers like Nvidia and AMD, memory producers like SK Hynix and Samsung, advanced packaging providers like TSMC—may continue to benefit from volume demand even as model prices fall. HBM memory remains tight, with SK Hynix’s 2026 capacity largely sold out; TSMC’s advanced packaging is booked through the planning horizon. Falling model prices do not ease these chokepoints quickly. If cheaper models encourage greater usage (a Jevons-like effect), infrastructure demand could remain strong or even accelerate. This was explicitly what Microsoft’s Satya Nadella predicted when the DeepSeek shock was rattling the Nasdaq.

Pain will likely be more acute in the middle of the value chain and downstream. Commodity API margins face compression if “good enough at one-tenth the price” becomes a viable option for many companies’ needs. Neo-

clouds with heavy debt loads and thin competitive moats—CoreWeave is a paradigmatic example—face uncertain demand if the heaviest compute tasks are rationed. High-end model vendors like OpenAI and Anthropic would need to migrate their value proposition from racing to the AGI singularity toward customer support, enterprise assurances, integrated tooling, and data security and governance.

For diversified hyperscalers, lower model prices might actually increase usage over time even as unit prices fall—much as cloud storage did over the past decade, though the supply-demand equation is unlikely to be as unbalanced as forecasts suggest. Their risk is timing. Enterprise take-up of heavy compute inference needs to match debt maturities and amortization schedules.

The U.S. policy response has focused on export controls—restricting chips, high-bandwidth memory, EDA software, and advanced packaging equipment—rather than anti-dumping measures, precisely because AI models are services and code, not tariffable goods. Entity List designations and government procurement restrictions are the major policy levers. Coordination on export controls is essential if Western governments hope to choke off Chinese progress. But they are unlikely to thwart a country that graduates several million engineers a year and is intent on winning technology wars in sunrise industries, at almost any cost. Huawei and SMIC are creating indigenous computing clusters and lithography capabilities, and Chinese chips already achieve performance scores competitive with some Nvidia offerings. The realization that microchip security is national security has emboldened both China and the West to attempt to build their own, autarkic supply chains in hardware and software.

China is a difficult factor for investors to accurately discount. The country's geopolitical strategy and industrial policy introduces real price pressures for Western firms. In an apocryphal quote, the Chinese Premier Zhou Enlai in 1972 answered a question about the impact of the French Revolution saying that it was "too soon to tell." This might as well be our guidepost when predicting the market caps of the various AI players five years, let alone ten years hence.

We should instead think hard about where in the value chain durable competitive advantages are most likely to persist. Upstream suppliers with genuine bottlenecks—advanced memory, leading-edge fabrication, specialized packaging—may prove more resilient than model vendors competing on capability that is increasingly matched at lower price points. Generally, their recent blowout earnings, while doubtless more cyclical than they currently appear, are testament to some underlying competitive scarcity. Software and cloud hosting were winners in the

technology wars of the last decade. These are much more uncertain, and should be evaluated on gross margins without overly hasty extrapolation of the future benefits of scale.

The Software Shock

2025's DeepSeek episode was not an isolated tremor along the unstable faultline of AI disruption. In late 2025, a second shockwave rippled through technology markets—this time originating from a Western, not Chinese source. The latest release of Anthropic's Claude Code assistant had progressed so remarkably in its long-term memory and user responsiveness that it became possible to envision writing functional software entirely in natural language. The music streaming app Spotify, to take just one example, recently shed 17% of its workforce, and its CEO announced that most of his top programmers had stopped writing code altogether. A team of journalists from CNBC without technical backgrounds was able to replicate the core codebase of Monday.com, a company with a $5 billion market cap, from scratch in hours. What had recently been an esoteric hobby was becoming standard practice: "vibecoding," or the art of describing what you want in plain English and letting the AI produce working code. The term, coined by Andrej Karpathy (a co-founder of OpenAI), describes an unfolding process of democratization of a once rarified skill—software engineering.

With progressive leaps in the usability of no-code tools like Claude Code and Anysphere's Cursor (the fastest tech company to reach $1 billion in revenue), the implication for legacy enterprise software became suddenly apparent. Salesforce bills $35 billion in annual subscription revenue. Low-level IT staff or even salespeople now possess at their fingertips the ability to write their own custom CRM app. Operations leads can instruct Claude in natural language to build a bespoke inventory management system, threatening the market dominance of ServiceNow. CNBC reported that the most exposed companies were those whose products sit "on top of the work" like Atlassian, Adobe and HubSpot—that is, they are not embedded in multiple layers of a customers systems and data. The market's answer arrived in the form of a "SaaSpocalypse": enterprise software stocks have suffered double-digit declines. On the February release of a new version of Claude, the BVP Nasdaq Emerging Cloud Index, a benchmark for software-as-a-service companies, dropped more than 12 percent.

When the production function for software is running an AI agent, SaaS competitive moats rapidly dry up. Costs come down dramatically. The entire enterprise business model—sticky subscriptions, high switching costs, annual price increases, customer success teams ensuring renewal—depends on the assumption that building bespoke software is expensive and difficult. The new economics favor model builders and model users, but not application providers.

That was the prevailing thesis in the winter of 2025-2026, anyway. As is its custom, the market may be getting ahead of itself. To take one example, Salesforce and its legions of features and customer service agents, and deep integration in its customers' critical data, is not going to disappear overnight. Enterprise software companies possess advantages that a weekend vibecoding project cannot easily clone: integration with legacy systems, compliance certifications, audit trails, support organizations, and the institutional knowledge embedded in decades of customer deployments. Indeed, incumbents may well end up incorporating AI capabilities into proprietary features—agentic workflows operating inside their systems that function more reliably than anything a casual user could vibecode from scratch. Salesforce's Agentforce, for instance, is precisely such an attempt to turn the threat into a competitive wedge.

The upheaval in enterprise software illustrates how long-germinating but abrupt leaps in AI's effectiveness can unanticipated consequences across industries. Enterprise software has been an unambiguous economic winner—high margins, fast growth and rich valuations. Out of the blue, its value proposition was under existential threat. Subscription-style software has become essential in almost every business, from accounting firms to hospitals to retailers. It is possible now to conceive of a world where it becomes nothing more than a cheap commodity, with near-zero replacement cost.

Bubbles as Engines of Innovation

We highlight Chinese competition, physical-world systems, and vibecoding to underscore the point that the path of AI progress and its social and economic aftereffects are difficult to underwrite, even in the short term. Capital markets have a way of condensing this infinite complexity to simple narratives. In recent years, anything that was AI-inflected was rewarded with a premium and seemingly able to raise substantial sums. We live in a world of radical uncertainty. Disruptive innovation, à la Schumpeter and doesn't just alter markets; it creates and destroys them. It is impossible to value something that rewrites the entire economy. Equally impossible is choosing which firm will dominate the market.

So we investors guess, we extrapolate, we lump companies together, we elide the nuance. Given the murkiness of our crystal ball, it's understandable. Are we all, collectively if not individually fools? Schumpeter and Perez would say "no."

The debate over whether there is an AI bubble, rational or irrational, has raged for over a year. Partly a question of semantics, since there is no universally agreed definition of "bubble," this argument is really about whether the

massive shifts in capital allocation—to datacenter construction, to power utilities, to microchips, cloud hosts, and model builders—will earn cash returns commensurate with their risks. Conventionally, any enormous investment boom will cause a great deal of value destruction somewhere in the constellation of hopes and dreams. And that is the point.

One of the more subtle insights from the history of technological revolutions is that speculative excess can serve a productive function—that the very irrationality of bubble-era investment can accelerate the development and deployment of transformative technologies in ways that sober cost-benefit analysis never would.

Venture capitalist William Janeway, drawing from Schumpeterian economics and practical experience financing the frontier of innovation, bubbles are not merely unpleasant side effects of high-tech capitalism, but essential to it. Under conditions of true uncertainty—what Schumpeter's contemporary Frank Knight called "unmeasurable" risk, where no probability distribution of potential outcomes exists—rational calculation is useless. Financial models cannot tell us whether a nascent technology will transform the economy or disappear into the dustbin of history. Successful markets will finance it anyway.

What Keynes called "animal spirits," the restless speculative instinct that is rooted in the emotional rather than the rational centers of our human brains are what drive economic growth forward. Animals often congregate in herds, stampede off cliffs, and fight to the death over mates. The inescapable result of a lack of objective knowledge, willing financial capital, and collective psychology is the inflating and popping of bubbles. A counterintuitive conclusion is that bubbles are good for society, that speculation is essential for technological development, that blind-risk taking is the wellspring of productivity growth[3]. Hedge fund billionaire and amateur philosophy George Soros has argued convincingly that financial markets manifest their own reality. Because tech investors imagine a particular future, they actually bring it into being in a kind of self-fulfilling prophecy.

[3] There are several caveats to this simplification, the most important being that the productive value of bubbles depends significantly on how they are financed. Equity-financed bubbles leave many speculators poorer, but not bankrupt. Debt-financed bubbles, like the mortgage boom and crash of 2007-2009, leave behind the wreckage of unpaid bills and an overhang of balance sheet repair.

This two-way "reflexivity"—the feedback between reality and markets' observation of that reality—describes very well the massive capital account needed to bring OpenAI and Anthropic's latest models from obscure science project to economic forces of nature[4].

Bubbles, in Janeway's view, are how the future gets funded. The excess capital that flows into speculative ventures during manias constructs infrastructure that could never be justified by a cold-hearted discounted cash flows analysis. That infrastructure often proves essential to the technologies and companies that emerge afterward. The fiber-optic cables laid during the telecom bubble, the railroad tracks that connected a continent, the data centers built during the cloud computing boom: all represented waste from the perspective of the investors who financed them, and all created foundations on which subsequent generations built fortunes.

The mechanism is wasteful almost by design. Most bets fail. Capital is misallocated. Fortunes are lost.
But the venture capital industry is built on the expectation that the few successes that emerge from chaos can generate returns that dwarf the losses. The fiber-optic cables laid during the telecom bubble sat dark for years, a monument to speculative excess. But their existence, and the bargain prices at which they could be acquired after the bust, enabled video streaming, cloud computing, and mobile connectivity that now seem indispensable. The money lost by the original investors was, in effect, a subsidy to the companies and consumers who came after. For careful investors, there is an important teaching in this argument. We should not decry the AI bubble and sit on the sidelines. Nor should we pretend that we know the victors in advance. We want, like most VCs, a diversified portfolio of options – relatively small bets that can pay off convexly (that is, with future returns wildly disproportionate to the initial capital commitment), and to keep some powder dry for subsequent generations of beneficiaries. Debt, on the whole, would be the wrong instrument for followers of this philosophy. A basket of equities with asymmetric earnings upside and positive unit economics would be more in keeping with Janeway's approach.

Legendary investor Howard Marks, in memo to Oaktree clients, draws a further distinction between what he calls "inflection bubbles" and "mean-reversion bubbles." The latter—think subprime mortgages or portfolio

[4] Sociologist Donald MacKenzie describes economics as an "engine, not a camera" with respect to financial markets. While academic studies of markets purport to merely describe financial reality, they actually have an active influence on the markets they study. We might extend this metaphor from financial markets to the deeper level of technology. Markets don't merely value existing innovations; they summon new ones.

insurance—involve financial innovations that promise returns without risk but create no tangible value; when they collapse, nothing remains but a paper trail of defaulted claims. The former involves productive breakthroughs that capture the imagination, attract excessive capital, and crash spectacularly, but leave behind useful stuff and new skills that recombine to foster a new wave of innovation. Marks would place canal mania, railway bubbles, and even the dot-com frenzy as inflection bubbles, building foundations for economic progress.

Generative AI has the hallmarks of an inflection bubble. The technological breakthrough is a rupture from what came before. Its power and utility have ramped at unforeseen speeds. The potential applications span virtually every sector of the economy. Whether today's valuations prove justified is another question entirely. If history is our guide, many (not all) companies will fail to reward investors' enthusiasm. Both propositions can be true simultaneously.

It is of course salutary for investors to bear in mind extreme tail scenarios, so-called "black swans" or six-sigma events, but the market reaction is unlikely to reflect a clear consensus for some time. The current structure of the AI market remains fragmented at many links in the value chain; it's not yet obvious who commands disproportionate market power. Multiple frontier models compete closely on performance. Open-source alternatives have proven surprisingly effective. The application layer is dynamic and crowded. Enterprise customers are wary of lock-in and cautiously managing vendor relationships. It is not yet clear that any single player will develop the kind of durable moat that characterized platform winners like Google in search, Amazon in e-commerce, Microsoft in enterprise software, or Facebook in social networking.

Meanwhile, the infrastructure arms race is imposing enormous capital requirements that show little sign of slowing, while new equipment's functional depreciation is incredibly rapid. Fiber-optic cable laid in 2000 is still in use today. Graphics chips from the same time are not. Today's cutting-edge data center may be tomorrow's stranded asset in a few years as models become more efficient, chip architectures evolve, and demand patterns change. The history of technology infrastructure is littered with investments that looked essential at the time and worthless shortly after. The fiber-optic boom of the late 1990s created the backbone of the modern internet, but the investors who financed that buildout largely lost their shirts. Most small investors in the railway boom that covered England in thousands of miles of track within the span of a few years were wiped out entirely.

Investing in AI today looks as precarious as when some of these earlier bubbles were inflating. A fundamental seed lies at the core of market enthusiasm – but markets are prone to extrapolation in the "installation" and "frenzy" phases and disappointment later. The relationship between investor returns and the magnitude of a given technology's social import is weaker than our intuition tells us.

The present moment feels all the more vertiginous when executives most responsible for the AI buildout have begun issuing cautionary notes about all the capital flowing their way. Jensen Huang, NVIDIA's CEO and the primary beneficiary of the chip boom, acknowledged in mid-2024 that "there's a different set of risks" when customers are buying infrastructure ahead of proven demand, and that the AI trade had elements of "speculative investment." Satya Nadella, boss of Microsoft, who has bet hundreds of billions on OpenAI and Azure AI infrastructure, has warned that "there's hype, there's froth" in the market and emphasized that "we have to be grounded in real usage and real value creation." Sam Altman, OpenAI's chief executive and perhaps the most prominent evangelist for AGI (artificial general intelligence, or a future where algorithms are far better at every cognitive task than humans), told an interviewer in late 2024 that "some of this is going to end in tears" and that valuations in parts of the ecosystem were "disconnected from near-term reality." When the principal players absorbing all the capital start sounding notes of caution, outside observers might reasonably take heed.

Most troubling is the self-referential quality of much of the early deal activity. A significant portion of AI revenue, particularly among the hyperscalers and model providers, comes from deals with other participants in the AI ecosystem itself: cloud credits extended to startups, infrastructure partnerships among the major players, and contracts where consideration often takes the form of equity or deferred commitments rather than upfront cash. This pattern rhymes uncomfortably with the fiber-optic buildout of the late 1990s, where carriers booked revenue by swapping capacity with one another, creating an illusion of end-demand that evaporated when investors got cold feet. Numerous large companies filed for bankruptcy amid discoveries that their financial numbers had been massively inflated.

Whether today's circular arrangements constitute the same species of accounting legerdemain or are simply hallmarks of an ecosystem in the early stages of maturity is a matter of much investor dispute; what is less debatable is that round-trip deals create tight interdependencies among multiple companies' valuations, leaving the entire complex particularly vulnerable to contagion when sentiment inevitably shifts. When Company A's revenue depends on Company B's spending, and Company B's valuation depends on contracts with Company A, a wobble anywhere can propagate everywhere.

We will examine these investor implications in more detail in a later section. For now, the takeaway is simply that the AI capital market today has outrun the reasonable forecasting ability of anyone familiar with past technology cycles. “Can OpenAI go bankrupt?” is the thrust of recent editorials in The Economist magazine, and, conspicuously, by venture enthusiast Sebastian Mallaby in *The New York Times*. The answer is, of course, “yes,” but the most interesting questions concern what happens next.

Patience is a Virtue

The recurrent cycles of technological adoption identified by economists tell us that much of the AI story remains unwritten and warn us against undue certainty. Excitement is rational, but we hardly need to point out the absurdity of investors believing that AI will bring about a critical disjuncture in history (AGI perhaps), akin to the industrial revolution, and *simultaneously* believing that they can today forecast the economic winners. Profound change in the way businesses process information is doubtless ahead of us. Humility is the watchword for anyone today trying to anticipate what this means.

For firm managers, the implication is that AI adoption is a process, not an event. Pilots are fine; full integration will require much more comprehensive organizational evolution. The firms that capture the largest productivity windfall will be those that treat AI implementation as a strategic investment in complements—data infrastructure, workflow redesign, training, governance—rather than a bolt-on to existing processes. The historical record strongly suggests that the spoils go to those who reorganize when it makes competitive sense, not those who automate around the edges. Sometimes being a fast follower is preferable to being an early adopter. To take only one example—Apple’s business faces challenges on many fronts, but demurring to pour a trillion dollars into cloud infrastructure may come to be seen as prudence rather than “fumbling the future.”

For policymakers, the implication is that the productivity effects of AI are coming, but not on any schedule that neatly fits electoral cycles. Investments in education, infrastructure, and institutional adaptation should pay dividends over decades. The temptation to take short-term positions for or against the “AI revolution” should be tempered by the recognition that the technology’s significance will not be perceived or understood for a generation. Policy should focus on future-proofing the economy and the labor force, on enabling the critical complements from power to interconnection to accounting and legal rules, and on ensuring AI safety guardrails.

For investors, the implication is that great fortunes are to be won and lost surfing the waves of market sentiment. But these are, at heart, leveraged beta trades. Alpha—at least the durable kind—comes when investors know

something about the shape of future fundamental demand and place bets that pay out in the form of cash flows, not paper-trading profits. It is very early to assess the shape of the opportunity for this sort of endeavor. Much of the value, we suspect, will ultimately accrue at the capillaries of the AI ecosystem, not the heart. AI tools that transform mundane businesses like gas stations and staffing companies, rather than shiny data centers, may be the economic story of 2035.

This generation of AI may fail to achieve its proponents' forecasts of superintelligent abundance. And yet despite our cautions, it will not fizzle. This is not tulipomania. All our historical analogies (railroads, electricity, computing, the Internet) delivered on their transformative promise, eventually. Hoped-for productivity improvements arrived—less in a big bang than in a steady, year-over-year drumbeat. The economy gradually reorganized to take advantage of the latent technological potential. It will do so again –perhaps too slowly for investors betting on a revolution, but more swiftly than in past episodes. How can we look for early signals of who will win, and how should we position portfolios across major asset classes? This is the subject of our next essay.

If you have any questions, please reach out to your client service team, visit us at hbwealth.com, or call 404.264.1400.

Notes

1. Robert Solow's remark appeared in his review of Manufacturing Matters by Stephen Cohen and John Zysman, published in the New York Times Book Review, July 12, 1987. The phrase "everywhere but in the productivity statistics" entered the economics lexicon almost immediately.
2. The formal treatment of general purpose technologies appears in Timothy F. Bresnahan and Manuel Trajtenberg, "General Purpose Technologies: 'Engines of Growth'?" Journal of Econometrics 65, no. 1 (1995): 83–108. A more accessible overview appears in the Brookings Institution's analysis of AI's economic effects: https://www.brookings.edu/articles/the-effects-of-ai-on-firms-and-workers
3. Paul David's canonical treatment of the productivity paradox is "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox," American Economic Review Papers and Proceedings 80, no. 2 (1990): 355–361. The paper remains essential reading for understanding technology diffusion lags.
4. Carlota Perez's framework is fully developed in Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages (Edward Elgar, 2002). Her taxonomy of installation/frenzy/turning point/deployment has become standard vocabulary among technology historians and venture capitalists alike.
5. The Noy and Zhang study is Shakked Noy and Whitney Zhang, "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," working paper, MIT, March 2023. Available at economics.mit.edu.

6. The Brynjolfsson, Li, and Raymond study examines customer service agents and was published as Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond, "Generative AI at Work," NBER Working Paper 31161 (2023), subsequently appearing in the Quarterly Journal of Economics (2025).
7. The BCG/Harvard study is Fabrizio Dell'Acqua et al., "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Harvard Business School Working Paper 24-013 (2023).
8. Estimates of AI maturity among enterprises vary by survey methodology but consistently find that only 10–20 percent of large firms have moved beyond pilots to scaled deployment. See McKinsey's annual "State of AI" surveys and Deloitte's enterprise AI reports for representative findings: McKinsey state of AI 2025, https://www.deloitte.com/content/dam/Deloitte/cr/Documents/consulting/2024/the-state-generative-ai-enterprise.pdf
9. On the measurement challenges for productivity in service sectors and knowledge work, the literature is vast. A useful starting point is the work of Chad Syverson, particularly "Challenges to Mismeasurement Explanations for the U.S. Productivity Slowdown," Journal of Economic Perspectives 31, no. 2 (2017): 165–186.
10. The comparison between railroad investor returns and the broader economic value created by railroads is a recurring theme in the economic history of transportation. See, among others, Robert Fogel's work on the social savings from railroads and the extensive literature on nineteenth-century railroad finance.

Important Disclosures

All information is as of the date above unless otherwise disclosed. The information is provided for informational purposes only and should not be considered a recommendation to purchase or sell any financial instrument, product, or service sponsored by HB Wealth or its affiliates or agents. The information does not represent legal, tax, accounting, or investment advice; recipients should consult their respective advisors regarding such matters. This material may not be suitable for all investors. Neither HB Wealth nor any affiliates make any representation or warranty as to the accuracy or merit of this analysis for individual use. This information contains forward-looking statements, predictions, and forecasts ("forward-looking statements") concerning our belief and opinions in respect to the future. Forward-looking statements involve risks and uncertainties, and undue reliance should not be placed on them. There can be no assurance that forward-looking statements will prove to be accurate, and actual results and future events could differ materially from those anticipated in such statements. Certain information herein is based on third-party sources believed to be reliable, but which have not been independently verified. Past performance is not a guarantee or indicator of future results; inherent in any investment is the risk of loss. Specific investments described herein do not represent all investment decisions made by the above date. The reader should not assume that investment decisions identified and discussed were or will be profitable. Specific investment advice references provided herein are for illustrative purposes only and are not necessarily representative of investments that will be made in the future. Investors are advised to consult with their investment professional about their specific financial needs and goals before making any investment decision. HB may hold positions (long or short) in the companies mentioned in this paper.