

Understanding the LLM Bubble

by Hubert Horan

The current debate on the existence of an “AI bubble” centers on a single question: is the current high level of investment into AI data centers a massive misallocation of capital or the key to future economic growth? The companies tightly linked to generative artificial intelligence (GenAI) drove 80 percent of the stock market’s growth across much of 2025.¹ The equity value of these seven companies increased by \$16 trillion in the last three years and now accounts for a third of the equity value of all listed U.S. companies. Even before potential trillion-dollar capital spending increases begin, data center investment accounts for about half of current U.S. GDP growth and is now a larger portion of GDP than all consumer spending; it may have also prevented the United States from falling into a recession.²

The tools at the center of the investment boom are Large Language Models (LLMs) that evolved from decades of language modeling based on the probabilistic pattern-matching of neural networks. Arguments that LLM investments are the key to future economic growth are based on the belief that growing LLM capabilities will lead to the rapid automation of a broad range of jobs, driving major improvements in business productivity. More importantly, advocates claim that in a few short years, today’s LLM apps will achieve artificial general intelligence (AGI), which could be one of the most important technological breakthroughs ever, as it will supposedly allow humanity to solve a wide variety of critical problems.

This essay aims to disprove and overturn that view: it is divided into three sections, each with a main counterargument to advance. The first is that the current LLM industry is not economically viable. Equity appreciation, repayment of the investors currently funding data center expansion, and the belief that the industry could drive huge national

Hubert Horan is forty-year veteran consultant in the transportation industry.

productivity growth all depend on the claim that LLM software will achieve superintelligence by (roughly) 2028. This section will explain why inherent, structural problems with current LLM software preclude the medium-term ability to achieve superintelligence; they also preclude the significant near-term improvements to their problem-solving capabilities needed to reduce the industry's huge current cash flow drain. The question is not whether LLMs have value—it is whether they create enough value to justify the orientation of today's industry, which is entirely focused on extremely large general-purpose LLMs that require trillions in investment. If the industry cannot provide end users with huge increases in productivity-enhancing capabilities, those end users will not provide the revenue the industry is forecasting, and the equity and investment boom will inevitably collapse.

The question of a possible AI bubble emerged in 2025, but this essay incorporates that issue into the more important but largely neglected question of how a boom this large could have been created and sustained in the face of major uncertainties about LLM technology and economics.

The second counterargument is that this bubble was deliberately engineered by investors pursuing the largest IPO payday in history, who managed to subvert the normal workings of capital markets; these markets are, after all, designed (at least in theory) to prevent trillions in capital from being allocated to companies that cannot demonstrate how they would fulfill core business promises or even generate positive cash flow.

The LLM bubble can best be understood with reference to the earlier development strategy pioneered by Uber over a decade ago, when its investors were pursuing a massive IPO payday despite \$33 billion in losses and the complete absence of a path to profitability over its first fourteen years. Uber is not relevant to today's LLM boom because of any economic similarities between rideshare services and data centers, but because both sets of investors wanted massive IPO paydays despite their failure to develop a viable business model. Uber is critical because it was the first private tech start-up to apply narrative techniques previously used in partisan political campaigns to corporate development, and it successfully used these to conjure a massive valuation out of thin air.

Corporate wealth extraction can take many forms,³ but this comparison highlights how both Uber and GenAI represent an especially pernicious version. Some extractive companies establish a concrete economic foundation, with profits earned in competitive markets based

on innovation and efficiency; these firms might then use artificial market power to increase those profits dramatically. Uber and GenAI, on the other hand, are almost entirely extractive, as they never offered products users were willing to pay for at the requisite scale needed to make a profit; these two firms have been strictly focused on transferring billions in wealth to their investors from the rest of society. These proven capital misallocation strategies raise broader questions which cannot be answered with an analysis narrowly focused on the LLM industry.

The third counterargument is that the LLM bubble is fundamentally different and more dangerous than past U.S. investment bubbles because there are three overlapping bubbles at play: the corporate (LLM industry) overinvestment bubble, the stock market overvaluation bubble, and a political expectations bubble about drivers of future economic growth. The political class in Washington has taken the claim that AGI would turbocharge national productivity growth at face value and is effectively close to giving the AI industry *carte blanche* when it comes to determining the direction of formative regulatory policies. All of these parties are currently making a trillion-dollar, one-way bet that superintelligence is on the imminent horizon. No previous U.S. investment bubble has had as many powerful forces working together to create and sustain a bubble based on such unrealistic technology and expectations. Needless to say, the stakes are enormous, as these same forces and the broader American economy are now threatened by the bubble's collapse.

THE LIMITS OF THE LLM INDUSTRY

The LLM industry has three major sectors: producers of the GPU (graphics processing unit) hardware required to create output for an LLM, the cloud services that purchase the GPUs and install them in data centers, and the companies that provide LLM apps to end users.

The chips side of the GenAI industry is dominated by Nvidia, whose valuation reached \$4.9 trillion in October 2025 and is now the most valuable company in the world.⁴ Nvidia currently sells over 90 percent of the GPUs required to create an output for an LLM.⁵ In a gold rush, most miners lose their entire investment, but the companies selling the picks and shovels to the miners can earn steady profits. Nvidia's valuation reflects its quasi-monopoly position selling "picks and shovels" to the very rapidly growing LLM industry.⁶

Cloud services are currently dominated by five "hyperscalers" (Microsoft Azure, Google Cloud, Meta, Amazon AWS, and new entrant

Oracle) and supplemented by smaller, more specialized “Neoclouds” (CoreWeave, Lambda, Anysphere, Nebius). LLM apps available to end users include OpenAI’s ChatGPT family, Anthropic’s Claude, Microsoft’s Copilot, Google’s Gemini, Perplexity, Amazon’s Alexa, xAI’s Grok, and Meta’s Llama. The current industry boom began with the November 2022 introduction of ChatGPT version 3.5.

Any discussion of the future of the LLM industry can be usefully simplified into an analysis of OpenAI, which claims ChatGPT has ten times more users than the next four LLM apps combined. Thus, OpenAI revenue approximates the total revenue that can currently be earned from all LLM end users. This revenue must cover the cost of Nvidia’s GPUs, data center capacity, and OpenAI’s LLM software, if its business model is to be viable.

The LLM industry has been claiming much greater productivity-enhancing capabilities are imminent. While current software can quickly gather and summarize a large set of legal documents and translate blocks of text, expectations are that LLMs will soon have the ability to complete legal briefs accurate enough to be submitted in court and provide complete document translations that accurately capture the original intended meaning. According to the industry’s projections, LLM software would then quickly develop “agentic” capabilities, including the ability to perform complex multistep tasks and figure out what subsequent tasks might be required without human supervision, eliminating the need to hire human assistants.

But the central justification for the LLM boom is the claim that within the next few years, today’s LLM software will achieve AGI and outperform humans in most economically valuable forms of work. Superintelligent systems would be capable of independent decision-making and improving their own capabilities. The industry also claims AGI would be the biggest economic breakthrough since the industrial revolution and would turbocharge national GDP and productivity growth.

OpenAI CEO Sam Altman said, “Building AGI that benefits humanity is perhaps the most important project in the world.”⁷ “Whether we burn \$500 million a year or \$5 billion—or \$50 billion a year—I don’t care, I genuinely don’t. . . . As long as we can figure out a way to pay the bills, we’re making AGI . . . breakthroughs for mankind [that] are priceless.”⁸

Meta CEO Mark Zuckerberg stated that he would rather risk “mispending a couple of hundred billion” on AI investments than miss

the opportunity to develop superintelligence, emphasizing that the greater risk lies in building too slowly.⁹

Nvidia CEO Jensen Huang laid out how everything from “surgical rooms to data centers, warehouses to factories, even traffic control systems or entire smart cities will transform from static, manually operated systems to autonomous, interactive systems embodied by physical AI.”¹⁰

Anthropic CEO Dario Amodei has claimed that AI would surpass human capabilities by 2026 or 2027, and he has written that AI data centers would be “smarter than a Nobel Prize winner across most relevant fields”; he claims as well that AI could eventually end poverty, bring world peace, and double the life span of the average person to 150 years. As he sees it, this will be a world where “cancer is cured, the economy grows at 10% a year, the budget is balanced—and 20% of people don’t have jobs.”¹¹

The LLM boom is also tied to particular claims around AGI. These include a commitment to neural network and “deep learning” software, and the belief that “scaling”—dramatically increasing the quantity of data LLMs were trained on—would drive the ongoing improvements and the eventual achievement of AGI.¹² As OpenAI’s Altman said, “How do we get to the doorstep of the next leap in prosperity? In 15 words: deep learning worked, got predictably better with scale, and we dedicated increasing resources to it.”¹³

This AGI-centric approach demanded that the industry would be highly centralized around very large companies developing general-purpose models (such as ChatGPT) that would require massive capital funding. While some industry observers have suggested alternative development concepts,¹⁴ industry leadership and investments have been totally focused on the current approach, and there is no possibility of changing to a different development strategy at this point.

The table below summarizes publicly reported OpenAI P&L forecasts as of mid-2025, which do not include \$300 billion of Oracle compute costs subsequently announced. The forecast is consistent with OpenAI’s expectations that its software will rapidly become much more valuable to users and that it will remain the overwhelmingly dominant LLM software company. But it also shows that a company that claimed to have a \$500 billion valuation at the time of the forecasts will need to fund \$31 billion in losses through 2026 and \$113 billion through 2028.¹⁵

OpenAI’s revenues are forecast to grow at a compounded annual rate of 124 percent across this period and are expected to reach \$100 billion by 2028; at a compounded annual rate of 92 percent, they would reach

FIGURE 1. SUMMARY OF OPENAI PROFIT AND LOSS FORECASTS

	2024	2025	2026	2027	2028	2029	2030
Total Revenue (US \$bn)	4	13	30	60	100	145	200
<i>% Change YoY</i>		225%	131%	100%	67%	45%	36%
Total Expenses (US \$bn)	9	22	47	95	147	153	162
<i>% Change YoY</i>		114%	114%	102%	55%	4%	5%
Net Income (US \$bn)	-5	-9	-17	-35	-47	-8	38
<i>Profit Margin (%)</i>	-125%	-69%	-58%	-58%	-47%	-5%	19%

Source: see endnote 15.

\$200 billion by 2030. In 2024, OpenAI incurred \$2.25 in expenses for every \$1.00 of revenue; this ratio is projected to be closer to \$1.50 for every \$1.00 of revenue in 2026, assuming OpenAI can hit these ambitious revenue growth targets.

Expenses roughly double each year through 2027, grow 55 percent in 2028, and then only grow at single-digit rates thereafter. As a result, profit margins improve only gradually between 2025 and 2028 (from negative 69 percent to negative 47 percent, then dramatically improve to near breakeven in 2029 and a positive 19 percent in 2030).

OpenAI has not publicly articulated any of the major assumptions behind these P&L forecasts, but it is reasonable to speculate that OpenAI's forecasted sea change in profitability after 2028 reflects its expectation of when the massively greater capabilities of AGI reach the market.

In 2025, the cloud operators required \$360 billion in capital expenditure (for Nvidia GPUs and other data center infrastructure) to create the compute capacity used by the LLM software companies.¹⁶ Goldman Sachs has estimated that the capex needed for 2026 would exceed \$500 billion, and Nvidia CEO Jensen Huang predicted that annual capex will exceed \$1 trillion by 2028.¹⁷ These capex figures understate the cash flow problem, as they do not include necessary operating expenses, including utilities and the salaries of AI software engineers.

And there is no evidence that the massive sums invested have yielded discernible productivity gains: none of the ChatGPT upgrades in the last two years have been seen by users or independent researchers as providing major new benefits; new versions of the software have not driven the expected magnitude of unit revenue growth forecasts and certainly do not demonstrate strong progress toward super-intelligence.¹⁸

Only 3 percent of the users of ChatGPT and Microsoft 365 Copilot pay anything for GenAI capabilities (a \$20 monthly fee in most cases).¹⁹ As the P&L forecast demonstrates, these user fees fall well short of OpenAI's expenses; demand would presumably fall if businesses had to pay the full cost of their queries. LLM costs are already artificially low since the training of the LLMs depends on (what some consider) intellectual property theft at an industrial scale. In addition, local taxpayers have effectively provided major subsidies for data centers, and much of the huge utility costs these data centers require has been forced onto local ratepayers.²⁰

LLM providers have not found significant ways to monetize "free" users comparable to Google's or Facebook's advertising and data collection efforts. There is no "killer app" comparable to the tools that drove rapid mass market acceptance of PC software. The strategy of giving away LLM access for free or at highly discounted rates in the hope that consumers discover valuable uses has not driven anything close to the magnitude of revenue growth needed to achieve positive cash flow. Users clearly value some LLM capabilities, but the cost of creating that value goes wildly beyond what users are willing to pay.

Reports have shown that thanks to the saturation of media narratives predicting that LLMs will revolutionize the economy, a very large percentage of big companies say they are "using AI." But those reports also show that most common uses (summarizing reports and emails, responding to simple customer inquiries, etc.) only replace very low-value work and have little or no impact on overall company productivity or costs.²¹ A recent study documented that LLMs actually reduced the productivity of coders, the activity believed most susceptible to LLM automation.²² Other studies found no "tangible impact on enterprise-level EBIT from their use of Gen AI" and that the use of LLMs in big companies is generating backlash and is now trending down.²³

ChatGPT's expenses have been increasing due to much higher training costs, and the higher compute costs needed to handle more complex queries. OpenAI's major project ("Arrakis") to improve compute cost efficiency failed and was shut down.²⁴ The naïve expectation that LLMs would follow the pattern of past Big Tech software offerings, with each new version offering increased functionality at lower prices driven by economies of scale, was never true. Claims that LLMs would, like previous Big Tech companies, inevitably pivot to profitability after large early losses are highly suspect because GPUs do not show Moore's Law-type productivity gains, and LLMs never had rapidly declining unit costs like Google and Amazon.

Furthermore, it is not clear whether the industry can find the trillion dollars in funding for all the compute capacity it needs to support planned revenue growth. To put this funding requirement in context, the combined lending capacity of the entire venture capital industry, the top ten private capital firms, plus JPMorgan and Goldman Sachs, is less than \$700 billion.²⁵ Recent estimates suggest the hyperscalers would need over \$2 trillion in revenue by 2030 to justify their planned levels of capital spending.²⁶ Most of this capacity is currently unfunded and depends on companies who have never built an LLM data center. Detailed timelines showing funding and completion dates consistent with revenue forecasts have not been disclosed, and hundreds of billions in data center financing would need to be repaid long before those forecasts show profits.

These near-term financial challenges could seriously undermine industry economics. But much more importantly, there is no evidence that the industry has a path to commercializing superintelligence by 2028. There has been ample evidence in recent years that the industry's current path might not lead to AGI at any point or even achieve significant improvements in LLM reasoning and problem-solving capabilities.

The central problem is that LLMs are, by design, nothing more than statistical pattern matchers, making probabilistic guesses to answers. They have been marketed as having the reasoning abilities to replace human labor, and the industry has anthropomorphized LLM outputs to make users think they emerged from humanlike thinking. But for years, LLM critics have been pointing out that, unlike humans, they cannot try to understand the questions they have been asked, are processing symbols but not meaning, and are incapable of determining when an apparent correlation reflects a causal relationship. Humans simultaneously make decisions and update their mental models based on the results of those decisions. These processes are totally segregated within LLMs that have no ability to update their algorithms based on new information.²⁷

One major driver of these misguided expectations is the false assumption that all knowledge workers do is mechanically produce "output" (e.g., computer code, written words, translated words, answers to customer questions) and that an LLM could be trained to produce the same output. The LLM industry treats nearly all of human intelligence and decision-making as an engineering process that software designers can replicate. The mistaken belief "that every aspect of learning or any other feature of intelligence can in principle be so precisely described

that a machine can be made to simulate it” dates to the earliest days of AI research.²⁸ The actual value of a creative worker is the ability to think about what they are producing. They might ask certain questions: “Could the program be designed more efficiently?”; “Do the written or translated words convey the meaning intended?”; “What would solve a difficult customer service problem?” Since LLMs cannot think, they cannot achieve the promised mass automation of creative worker jobs beyond the simplest tasks, or ones with the lowest stakes.

One inevitable result of this underlying design is that LLMs often produce inconsistent and inaccurate answers. LLMs are designed to produce outputs that have a plausible statistical correlation with training data, while business users expect answers that are objectively accurate and can withstand the critical scrutiny of knowledgeable humans, which is something fundamentally different. The industry calls these “hallucinations,” which falsely implies the LLMs are sentient while being prone to freak aberrations instead of a structural, unsolvable problem that has long been documented by industry critics.²⁹

Accuracy declines as user queries become more complex, and user productivity declines since LLM output must be manually checked for fabricated or nonsensical claims. This means LLMs can only automate relatively low-value work, wiping out both the potential for large forecasted business productivity increases and any chance that current LLMs will inevitably (and rapidly) evolve into AGI systems.

If there is no path to superintelligence by 2028, and there is little prospect of the dramatic product improvements needed to drive major short-term revenue growth (including solutions to inaccuracy and unreliability issues), it will be impossible to sustain either the investment boom or the LLM industry, as currently organized and operated.

The burden of disproving this conclusion rests with the companies seeking the additional trillion dollars in capital. Meeting this burden will require vastly more concrete data about operating finances, capex construction, and funding plans as well as a coherent explanation for why the end users who paid \$4 billion in 2024 will soon be willing to pay hundreds of billions; these companies must also explain how the path from today’s error-prone LLMs to superintelligence can be completed within the next two to three years.

**THE UBER PLAYBOOK:
FINANCIAL EXTRACTION AND NARRATIVE CONTROL**

OpenAI was founded in 2015, with funding from some of the biggest names in Silicon Valley (including Altman, Elon Musk, Peter Thiel, Reed Hoffman, and Amazon) who shared the belief that superintelligence was inevitable. It was originally chartered as a nonprofit in 2018, “unencumbered by the need to produce financial returns . . . to ensure that artificial general intelligence (AGI) . . . benefits all of humanity.” But faced with the skyrocketing costs of scaling training data and the difficulties in matching competitor compensation, OpenAI abandoned its “greater good” mission and shifted to the Big Tech model based on emphasizing revenue hypergrowth and investor enrichment within an oligopolistic industry structure.³⁰

OpenAI became singularly focused on equity appreciation, believing, despite the current lack of a profitable product or positive cash flow, it could achieve a valuation comparable to the other “Magnificent Seven” Big Tech companies. OpenAI rapidly increased its claimed value from \$80 billion in early 2024 to \$157 billion in October 2024 to \$300 billion in April 2025 and to a \$500 billion valuation in September 2025. Reports now indicate that OpenAI is pursuing a \$1 trillion valuation for an IPO in 2026 or 2027, which would be the largest IPO in history.³¹

The only way the equity and investment boom can be justified is if OpenAI can repay the trillion dollars it hopes investors will provide in capex funding: this is, of course, a scenario in which its software lives up to all the promises made about discovering the cure for cancer, ending global poverty, solving the climate crisis, and turbocharging national economic growth. Absent these achievements, OpenAI must be seen as pursuing an explicitly extractive strategy. Its investors want to extract a trillion dollars from capital markets without establishing a viable business or doing anything else that would increase overall economic welfare.

OpenAI cannot follow the financial blueprint Google, Facebook, and Amazon used to achieve outsized investor returns. While current Big Tech profits rely heavily on artificial market power, each of them initially developed strong core businesses based on genuine product and service breakthroughs (Google’s search engine, Facebook’s social network, Amazon’s ecommerce infrastructure) that led to significant value and profitability. User revenue fully covered the costs of these services and generated very strong scale and network economies, creating strong positive cash flow that financed ongoing growth.

OpenAI (and the other LLM software companies) have none of those economic attributes, but their investors are demanding comparable equity returns. They claim to be emulating the last generation of successful Big Tech companies, yet their actual plan seems to be to skip all the hard parts Amazon, Google, and Facebook went through and go straight to exploiting market power in an oligopolistic LLM industry.

Fortunately for OpenAI and its investors, another firm managed to do just that: in the early 2010s, Uber and its investors were also pursuing explicitly extractive returns and demonstrated how it could work. Indeed, the LLM industry has practically adopted Uber's hypergrowth strategy on an almost copy-and-paste basis.³²

Uber accumulated \$33 billion in losses between 2010 and 2022, consistently lost \$4–5 billion a year prior to the pandemic with negative 45–55 percent GAAP margins. Between 2015 and 2019, revenues grew at a compound annual rate of 63 percent, but losses kept increasing. Uber's investors were seeking a \$100 billion IPO valuation in 2019 even though its S-1 gave no indication of what it planned to do to eventually achieve profitability aside from the (false) claim that they would become the "Amazon of Transportation."³³ Uber would not generate a dollar of positive cash flow until 2023, the firm's fourteenth year of operations, after the business model the S-1 had been based on was completely abandoned.

The \$80 billion valuation Uber's IPO ultimately achieved was less than what they had targeted, but it was a huge accomplishment since it was conjured out of thin air. There was absolutely no credible evidence in that period that future cash flows justified any valuation above zero. Nonetheless, capital markets in 2019 established it as the most valuable transportation company in the world and the second most valuable tech start-up ever (after Facebook).

Uber's key innovation had nothing to do with technology. It was the first private company to base its development entirely on the type of manufactured narratives previously only found in partisan political campaigns. All of Uber's narratives had been taken from a previous neoliberal propaganda effort which called for replacing all taxi regulations with a laissez-faire regime, financed by Charles Koch-aligned groups. Substantively distinct but functionally similar narratives have created and sustained the LLM bubble.³⁴

To distract attention from its actual financial results, Uber and OpenAI alike argued that they were leading a moral struggle with highly emotive and tribal dimensions. Uber's Manichaeic narrative pitted heroic tech innovators against the "evil taxi cartel" and corrupt

regulators who would block the larger social benefits ridesharing would create. The LLM industry's analogous narrative has pitted messianic data scientists on the verge of one of the greatest economic breakthroughs of all time against the foolish people too shortsighted to see the importance of funding every bit of investment the industry demanded.

Foreshadowing the LLM industry's later spectacular projections of a world transformed, Uber claimed its business model was so efficient that it would rapidly dominate car services across the globe, eventually displace private car ownership, and become the leading company in the autonomous vehicle industry. Its claims of near-term benefits were also designed to mislead the public. Uber claimed its popularity and rapid growth were due to customers freely choosing its lower fares and increased service in competitive markets.

As with LLM data center investments, Uber's huge initial investments were not based on any evidence that customers were willing to pay for the actual costs of their service. Uber's original investors provided \$13 billion in funding by 2015 and \$20 billion by 2019. This was 2,300 times greater than Amazon's pre-IPO funding because Uber's only path to Amazon-like growth was using massive amounts of investor funding to subvert markets by driving lower-cost competitors out of business. Uber customers were only paying 56 percent of the cost of their service in 2016; in 2020, they were still only paying 70 percent.

OpenAI heralded similarly strong demand growth even though its users were paying an even smaller portion of their costs. Unsubstantiated claims about rapidly automating knowledge work were designed to divert attention from the LLM industry's refusal to pay for intellectual property or other major portions of their actual data center costs.

Both companies played fast and loose with accounting practices in order to mislead investors about actual finances. Uber inflated its 2018 earnings by \$5 billion; the company did this by combining earnings from current operations with the alleged appreciation of untradeable securities it received after abandoning unprofitable overseas operations. This gave readers of Uber's 2019 S-1 document the false impression that its ongoing operations were rapidly approaching breakeven. The hyperscalers of the LLM industry have their own accounting trick: projecting the depreciation of GPUs as a five- to six-year process instead of the actual three-year useful life, thereby overstating earnings by \$26 billion.³⁵

Uber was always openly pursuing (and eventually achieved) quasi-monopoly dominance of urban car service, even though, in the previous century, the industry had never shown tendencies toward even mild

concentration. The firm's ruthless behavior enabled it to raise fares and cut driver compensation at will since there was no possibility new competition would emerge to discipline anticompetitive behavior.

OpenAI's development strategy has always targeted a similar dominance of LLM/AGI software, even though robust competition is arguably a better way to drive technical innovation and commercial discipline in an emerging, immature industry. Both companies knew that investors would reward the focus on artificial market power.

Thorough due diligence could have shown investors hard truths: Uber's costs were much higher than the traditional taxis they had driven out of business; likewise, the data centers needed to fulfill OpenAI's growth plan had not been properly financed, could not be completed on schedule, and would suffer huge losses if they ever opened. Until late 2025, many investors ignored LLM bubble risks, believing that markets would correct any serious problems, oblivious to how the conventional wisdom the markets were reflecting had been manufactured by the investors hoping to profit from such narratives.

Corporate development traditionally involves a symbiotic relationship between investors and the rest of society. Consumers, workers, suppliers, and capital markets would normally evaluate the performance of a new company against competitive alternatives using price and P&L signals. In large-scale cases like these, primary marketplace signals would be augmented with analysis by objective outsiders: competitors, industry experts, financial analysts, journalists, etc. If successful in generating productive and profitable business models, the investors would receive financial gains commensurate with the risks they had taken, just as overall economic welfare increases thanks to positive externalities like job growth, technology spinoffs, product and efficiency improvements, increased market competition, and tax payments.

Uber and GenAI represent the polar opposite of this model. Both recognized the need to overcome traditional capital market discipline. If the narrative claims about benefits for society had ever been true, their public statements would not have been dominated by vague, pie-in-the-sky assertions about a wonderful Jetsons-like world at some indefinite point in the future. Instead, they would have focused on concrete steps toward those major benefits; they would have been clear about exactly what it would take to achieve Level 5 vehicle autonomy or about what was needed to solve the hallucination problem. These firms would have invited credible outsiders to review both their technical work and their progress toward sustainable profitability. But their actions suggest zero interest in defining real objectives or accepting critical scrutiny.

This artificial equity demand bubble also transformed Big Tech. In 2022, the combined market cap of Microsoft, Google, Meta, and Amazon had fallen 42 percent, going from \$6.88 trillion to \$3.98 trillion. After the initial 2022 market splash from ChatGPT 3.5, Big Tech joined the LLM scaling and data center arms race, hoping to rekindle equity appreciation and protect itself against anything the LLM app companies might do to threaten existing quasi-monopoly cash flows. The strategy of convincing Wall Street that they were now high-growth “AI companies” succeeded, and their market caps grew back to \$6.9 trillion by the end of 2023, and to \$12.21 trillion by November 2025. Oracle profits had been floundering, but its equity value rose 40 percent after it announced a \$300 billion deal to provide compute to OpenAI, even though it did not have any feasible plan to actually provide that compute.³⁶

Just as narrative power meant that Uber investors never demanded a path to profitability in the 2010s while \$33 billion in losses mounted, investors never demanded an explanation of what Uber had actually done to reverse the losses. The manufactured conventional wisdom prevented scrutiny of underlying economics both pre-pandemic (e.g., predatory subsidies, huge losses) and post-pandemic (using artificial market power to repress wages, cut service, and raise fares). It remains to be seen if the LLM industry’s assertions of future monumental growth can survive major data center financing challenges or a serious deflation of asset values.

EXTRACTIVE COALITIONS AND THE COMING RECKONING

Both Uber and OpenAI worked to lock in the loyalty of powerful supporters. The interests of Silicon Valley tech utopians, ideological neoliberals, and large-scale capital accumulators are not identical, but they make for a supremely effective political alignment. Indeed, the ability of these groups to wield their concentrated influence is precisely what accounts for the financial and political stranglehold of the two firms in spite of all the documented defects of their business models. The question now hanging over the U.S. economy is whether this coalition and the extractive economic system it sustains can withstand the catastrophic consequences that a bursting LLM bubble would deliver.

The hypergrowth strategy shared by Uber and OpenAI reflects the capital accumulators’ view that stock market prices are the only relevant measure of national economic health and that capital markets exist to facilitate outsized returns—not to allocate capital to its most productive

uses. Uber's \$33 billion in losses, OpenAI's projected \$131 billion in losses through 2028, and the common failure to deliver on core promises (improving transport efficiency or achieving superintelligence) can all be ignored because traditional capital market discipline has been neutralized in their eyes, and it will not be an obstacle to a huge IPO payday.

The ideological neoliberals support the view that companies have no obligation to improve overall economic welfare and applaud regressive wealth transfers from labor to capital. Uber's fight to block basic labor law protections for its already weak drivers was the decisive factor in its post-pandemic P&L recovery.³⁷ The LLM industry's development approach denigrates knowledge workers as readily replaceable drones, and the most enthusiastic customers of LLMs are companies focused on eliminating jobs as soon as possible.

Tech leaders believe that they are the predominant driver of the economy, that they should have a central role in the leadership of the country, and that AGI is key to this vision of the future.³⁸ These assertions reflect Silicon Valley's faith that any problem can be solved by applying the right technology and algorithms. As they see it, industry disruption only requires programming skills and processing power, rather than any genuine understanding of urban car service economics or how knowledge workers actually add value.

As noted, the LLM boom is not a single bubble; it is three distinct but overlapping bubbles: a corporate bubble, a stock market bubble, and a political bubble. This means the potential bursting of the LLM bubble would combine the effects of those three forces, making it more dangerous than previous large bubbles, including the pre-1873 railroad construction bubble, the 1990s telecom and dot-com booms, and the 2008 housing crisis. Most recent discussion on the subject has focused narrowly on the corporate bubble, where near-term forces that could drive asset deflation include data center overexpansion, non-Big Tech company cash flow problems, and independent financial firms deciding to increase their assessments of LLM risks.³⁹

A critical difference is that the technology and basic business logic of railroads, telecom, and housing had been proven, while the current LLM business model, based on spending hundreds of billions to scale training models to develop very large general models, has not. Significant capital was destroyed after the previous bubbles burst, but most investment was not and (unlike GPUs) could remain in place until greater revenue potential emerged. Arguments that current LLM capital could be reorganized around totally different products and software models

ignore how lengthy and painfully complicated this hypothetical restructuring process would be.

The problem underlying the stock market bubble is the unprecedented size of the LLM investment boom. LLM investors have been seeking very large but low-risk returns. Reallocating this much capital under any conditions would be highly disruptive, and finding alternative investments with even vaguely similar risk-return outlooks that this capital could shift to after a major market downturn might be impossible. Forced mergers, such as the acquisition of the LLM app companies and neoclouds by Big Tech, would not address any of the underlying problems and would significantly weaken Big Tech financials.

Most importantly, the LLM industry also rests on a major political bubble. Much of the Washington political establishment has uncritically embraced the idea that superintelligence is imminent and will produce the productivity growth to fund all of their policy objectives. The Democrats have long been Big Tech supporters, but the Trump administration has openly positioned itself as an especially enthusiastic AI cheerleader. Trump, who is highly focused on increasing stock market valuations, established an “AI and Crypto Czar” position on his staff, provided \$60 billion in tax subsidies for data center construction, has openly dismissed concerns about either AI-driven job losses or bubble-driven market declines, and sees the LLM industry as critical to stronger ties with the Gulf States and U.S. competition with China.⁴⁰

Just as it would be extremely difficult for tech investors to admit major error and find comparable alternative investments after a sizeable market decline, it is difficult to imagine Trump and his allies admitting their faith in AI did not have a rational basis and convincing voters they could find some other way to produce the economic growth LLMs were supposed to deliver. Trump would not be the proximate cause of any collapse, but he would be the focus of public anger. The level of that anger would depend on whether his supporters saw him fighting to protect them from the economic fallout, or whether they saw him fighting to save tech billionaires and preserve the tax cuts for the rich and other unpopular policy priorities that AGI-driven growth was supposed to fund.

The possibility of using federal bailouts to mitigate market declines if LLM assets deflate will not solve any of the fundamental business problems and would amount to shoveling taxpayer dollars into a bottomless pit. Since AGI is not imminent, claims that massive tech industry bailouts are justified by the need to keep China from controlling the next industrial revolution are obviously specious; more honest

demands for massive taxpayer bailouts to keep China from controlling the technology that can quickly summarize office memos or make it easier to create AI slop videos on Tik Tok would nonetheless strike voters as unpersuasive.

Data center capex funding has been highly opaque and has increasingly come from private capital, so it is impossible to tell whether the fallout might lead to the kind of financial contagion that wrought havoc across the U.S. and global economy in 2008. But the impact of a major stock market decline would surely spread well beyond direct investors. Current consumer spending is largely driven by the wealthier Americans who would be most directly affected by a major equity decline, especially since most, if not all, of the post-2023 appreciation of the Big Tech stocks would be reversed. Overall economic confidence, which is already weak, would surely take a major hit as less wealthy Americans feared for the job and investment cuts that would inevitably follow major market declines.

MAGIC BULLET OR MIRAGE?

Modern financial bubbles are not due to inadvertent, unintended errors. They are created because politically well-organized interests believed they would benefit from them. The LLM equity and investment boom was engineered by overlapping interest groups in the tech and finance sectors (and their political allies) who came to believe that AGI was the magic bullet they and the rest of the economy desperately needed. In their shared vision, AGI would finally prove beyond all doubts that Silicon Valley tech utopians deserved to have a preeminent political role leading the country into an abundant future; it would revive Big Tech valuations and power after years of stagnant growth, reversing the trend by which quasi-monopoly power had significantly slowed meaningful new innovations; and it would, of course, allow LLM insiders to achieve staggering levels of personal wealth without ever producing a business with positive cash flow.

Despite a total disconnect from the real economy, AGI would also keep overinflated stock market valuations growing and sustain the claim that stock prices were the only important measure of national economic wellbeing. And most important for maintaining political consensus, AGI would fund the Washington political establishment's favored projects without having to deal with slumping productivity, skyrocketing debt, the precarious situation of many Americans, or myriad other serious problems.

Previous bubbles had some political components, but past instances did not have as many powerful forces working together to create and sustain the bubble. Should the AGI bubble burst, it is unclear how catastrophic the fallout would be or how wide the financial contagion might spread. A lesson of the 2008 financial crisis is that, despite massive costs and the widespread recognition of destructive behavior, absolutely no one was held accountable and the extractive forces that caused the crisis quickly resumed as if nothing had happened. Nevertheless, both investors and policymakers should recognize that the artificial, Uber-like narratives that have driven the LLM bubble are unsustainable. ^A

NOTES

- ¹ Microsoft, Nvidia, Apple, Amazon, Alphabet (Google), Meta and Tesla had a combined market capitalization of \$22.2 trillion in November 2025, 33 percent of the combined \$67.3 trillion market cap of all traded U.S. companies. Other major LLM companies such as OpenAI and Anthropic are privately held and not included. See: Piero Cingari, “Magnificent 7 Market Cap Tops \$22 Trillion—and Nvidia Just Got Bigger Than Japan,” *Benzinga*, October 29, 2025.
- ² Ruchir Sharma, “America Is Now One Big Bet on AI,” *Financial Times*, October 7, 2025; Derek Thompson, “This Is How the AI Bubble Will Pop,” *Derek Thompson* (Substack), October 2, 2025; Wolf Richter, “AI Makes Past Hype Cycles Look Tame: PitchBook on the Stunning Venture Capital Mania for AI,” *Wolf Street*, September 30, 2025; Victor Tangermann, “Deutsche Bank Issues Grim Warning for AI Industry,” *Futurism*, September 24, 2025; “What If the AI Stockmarket Blows Up,” *Economist*, September 7, 2025; Greg Ip, “The AI Boom’s Hidden Risk to the Economy,” *Wall Street Journal*, August 2, 2025; Christopher Mims, “Silicon Valley’s New Strategy: Move Slow and Build Things,” *Wall Street Journal*, August 1, 2025.
- ³ Tim Wu, *The Age of Extraction: How Tech Platforms Conquered the Economy and Threaten Our Future Prosperity* (New York: Knopf, 2025).
- ⁴ Niket Nishant and Rashika Singh, “Nvidia Hits \$5 Trillion Valuation as AI Boom Powers Meteoric Rise,” *Reuters*, October 29, 2025.
- ⁵ Adam Hales, “Can Jensen Huang Maintain Nvidia’s Grip on AI as Competitors Rise amid Geopolitical Tensions,” *Windows Central*, October 2, 2025.
- ⁶ John Rapley, “Is the AI Bubble about to Burst?,” *UnHerd*, August 12, 2025.
- ⁷ Altman quoted in: Karen Hao, *Empire of AI* (New York: Penguin, 2025): 144.
- ⁸ Christiaan Hetzner, “OpenAI’s Sam Altman Doesn’t Care How Much AGI Will Cost: Even If He Spends \$50 Billion a Year, Some Breakthroughs for Mankind Are Priceless,” *Fortune*, May 3, 2024.
- ⁹ Lee Chong Ming, “Mark Zuckerberg Says He’d Rather Risk Misspending a Couple Hundred Billion Dollars Than Be Late to Superintelligence,” *Business Insider*, September 19, 2025.
- ¹⁰ Huang quoted in: Matt Stoller, “Nvidia’s Trillions: Why We Keep Hearing About AI,” *BIG by Matt Stoller* (Substack), January 13, 2025.

- ¹¹ Allison Morrow, “The ‘White-Collar Bloodbath’ Is Part of the AI Hype Machine,” CNN, May 30, 2025.
- ¹² OpenAI described the critical role of scaling in: Jared Kaplan et al., “Scaling Laws for Neural Language Models,” Arxiv, January 23, 2020.
- ¹³ Sam Altman, “The Intelligence Age,” SamAltman.com, September 24, 2024.
- ¹⁴ Gary Marcus has advocated alternative approaches including neurosymbolic or world models and more narrowly focused LLMs at smaller scale with vastly reduced training costs. See: Gary Marcus, *The Algebraic Mind: Integrating Connectionism and Cognitive Science* (Cambridge: MIT Press, 2003); Gary Marcus, “The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence,” Arxiv, February 14, 2020. In addition to these writings, Marcus’s *Marcus on AI* Substack publication was a major source for the argument presented here that the LLM industry is not on a path to superintelligence. No one has demonstrated that the revenues from a smaller, significantly restructured industry using alternative software development approaches would be able to cover their costs.
- ¹⁵ Sri Muppidi, “OpenAI Says Its Business Will Burn \$115 Billion Through 2029,” *Information*, September 6, 2025; Ed Zitron, “OpenAI Needs A Trillion Dollars in The Next Four Years,” *Where’s Your Ed At?* (Substack), September 26, 2025. Zitron’s analysis is another major source for the arguments presented here about the viability of the LLM industry.
- ¹⁶ Karen Weise, “Big Tech’s A.I. Spending Is Accelerating (Again),” *New York Times*, October 31, 2025.
- ¹⁷ Beth Kindig, “Nvidia CEO Predicts AI Spending Will Increase 300%+ in 3 Years,” IO Fund.com, March 19, 2025; 2025 capex data is from: Ed Zitron, “The Haters Guide to The AI Bubble,” *Where’s Your Ed At?* (Substack), July 21, 2025; “Why AI Companies May Invest More than \$500 Billion in 2026,” Goldman Sachs, December 18, 2025.
- ¹⁸ OpenAI’s original GPT (generative pretrained transformer) product was released in June 2018, based on technology introduced by Google in 2017.
- ¹⁹ Ed Zitron, “OpenAI Is a Bad Business,” *Where’s Your Ed At?* (Substack), October 2, 2024.
- ²⁰ Jeremy Hsu, “When It All Comes Crashing Down: The Aftermath of the AI Boom,” *Bulletin*, December 5, 2025.
- ²¹ “New Workday Research: Companies Are Leaving AI Gains on the Table,” Workaday, January 14, 2026; Kate Niederhoffer et al., “AI-Generated ‘Workslop’ Is Destroying Productivity,” *Harvard Business Review*, September 22, 2025; Steve Lohr, “Companies Are Pouring Billions Into A.I.; It Has Yet to Pay Off,” *New York Times*, August 13, 2025.
- ²² Joel Becker et al., “Measuring the Impact of Early 2025 AI on Experienced Open-Source Developer Productivity,” METR, July 10, 2025.
- ²³ Victor Tangermann, “Economist Warns the AI Bubble Is Worse Than Immediately Before the Dot-Com Implosion,” *Futurism*, July 21, 2025; “The State of AI: A Global Survey,” McKinsey & Company, March 20, 2025.
- ²⁴ Ed Zitron, “How Does OpenAI Survive?,” *Where’s Your Ed At?* (Substack), July 29, 2024.

- ²⁵ Ed Zitron, “The Case Against Generative AI,” *Where’s Your Ed At?* (Substack), September 25, 2025.
- ²⁶ David Cahn, “AI’s \$600B Question,” Sequoia Capital, June 20, 2024; Ed Zitron, “Big Tech Needs \$2 Trillion In AI Revenue By 2030 or They Wasted Their Capex,” *Where’s Your Ed At?* (Substack), October 31, 2025.
- ²⁷ Gary Marcus pointed out these structural shortcomings over the years; see: Gary Marcus, “Deep Learning: A Critical Appraisal,” Arxiv, January 2, 2018; Gary Marcus, “Deep Learning Is Hitting a Wall,” *Nautilus*, March 10, 2022; Gary Marcus, “Game Over. AGI Is Not Imminent, and LLMs Are Not the Royal Road to Getting There,” *Marcus on AI* (Substack), October 18, 2025; Emily M. Bender et al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?,” *ACM Digital Library*, March 1, 2021; Emily M. Bender and Alex Hanna, *The AI Con*, (New York: Harper Collins, 2025).
- ²⁸ Nils J. Nilsson, *The Quest for Artificial Intelligence* (Cambridge: Cambridge University Press, 2009).
- ²⁹ Ed Zitron first argued that hallucination and inconsistency problems precluded projected LLM revenue growth in: Ed Zitron, “The Rot Com Bubble,” *Where’s Your Ed At?* (Substack), June 3, 2024. An increasing number of papers from industry insiders are confirming that the accuracy problems cannot be solved. See: Parshin Shojaee et al., “The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity,” *Apple Machine Learning Research*, June 2025; Gyana Swain, “OpenAI Admits AI Hallucinations Are Mathematically Inevitable, Not Just Engineering Flaws,” *Computerworld*, September 18, 2025; Mike Brock, “Why I’m Betting Against the AGI Hype: An Engineer’s (Philosophical) Perspective,” *Circus* (Substack), November 28, 2025.
- ³⁰ Hao, *Empire of AI*, 13, 25, 61, 105, 133.
- ³¹ Echo Wang et al., “Exclusive: OpenAI Lays Groundwork for Juggernaut IPO at up to \$1 Trillion Valuation,” *Reuters*, October 30, 2025.
- ³² All of the Uber analysis is taken from papers that I have been publishing since 2016. The best overview of my analysis of Uber’s economics during the first fourteen years of its operations can be found in: Hubert Horan, “Uber’s Path of Destruction,” *American Affairs* 3, no. 2 (Summer 2019): 108–33. An earlier, more in-depth analysis can be found in: Hubert Horan, “Will the Growth of Uber Increase Economic Welfare?” *Transport Law Journal* 44 (2017): 33–105. I also authored a series entitled “Can Uber Ever Deliver?” at *Naked Capitalism*. See: Hubert Horan, “Can Uber Ever Deliver? Part Thirty-Five: What Drove Uber’s Recent \$8 Billion P&L Improvement?,” *Naked Capitalism*, November 30, 2026; Hubert Horan, “Can Uber Ever Deliver? Part Thirty-Two: Losses Top \$33 Billion But Uber Has Avoided The Equity Collapse Most ‘Tech’ Startups Experienced,” *Naked Capitalism*, February 27, 2023.
- ³³ Uber’s Board was so determined to file an IPO that in 2018 they sacked CEO Travis Kalanick who felt Uber was not ready for serious capital market scrutiny. See: Hubert Horan, “Can Uber Ever Deliver? Part Ten: The Uber Death Watch Begins,” *Naked Capitalism*, June 15, 2017.

-
- ³⁴ Hubert Horan, "The Uber Bubble: Why Is a Company That Lost \$20 Billion Claimed to Be Successful?" *ProMarket*, November 20, 2019.
- ³⁵ Hubert Horan, "Can Uber Ever Deliver? Part Thirty-Five: Uber's IPO Prospectus Overstates Its 2018 Profit Improvement by \$5 Billion," *Naked Capitalism*, April 15, 2019; "The \$4trn Accounting Puzzle at the Heart of the AI Cloud," *Economist*, September 18, 2025.
- ³⁶ The OpenAI deal briefly made Oracle CEO Larry Ellison the wealthiest person in the world. Most of Oracle's equity gains were subsequently lost after data center bubble concerns increased. See: Berber Jin, "How Sam Altman Tied Tech's Biggest Players to OpenAI," *Wall Street Journal*, October 20, 2025. A further, larger P&L boost came after Uber and other gig companies spent \$200 million to nullify California laws that extended labor law protections to gig drivers, which immediately raised the market capitalization of Uber by \$36 billion (over 60 percent) at a point when passenger payments were still covering less than 70 percent of Uber's actual costs.
- ³⁷ Another contributing factor was Uber's 2023 introduction of first-order price discrimination that tailored fare and driver pay offers to the lowest amount it thought individual drivers would accept, or the highest level individual passengers would pay, something companies in competitive markets would not be able to impose. See: Hubert Horan, "Can Uber Ever Deliver? Part Thirty-Five: What Drove Uber's Recent \$8 Billion P&L Improvement?," *Naked Capitalism*, February 25, 2025.
- ³⁸ Marc Andreessen, "The Techno-Optimist Manifesto," Andreessen Horowitz, October 16, 2023; Iain Davis, *The Technocratic Dark State: Trump, AI and Digital Dictatorship* (London: Papercut Publishing, 2025).
- ³⁹ For one detailed discussion of these corporate issues as of the end of 2025, see: Ed Zitron, "Premium—How The AI Bubble Bursts In 2026," *Where's Your Ed At?* (Substack), December 22, 2024.
- ⁴⁰ In addition to \$6 billion in direct new federal spending on AI, the AI industry gained roughly \$50 billion through a "One Big Beautiful Bill" provision that would allow data center companies to write off all investment expense in the first year. See: "The One Big Beautiful Bill Act: A Catalyst for AI Investment and Tech Growth," *Growth Shuttle*, July 2025. An amendment to that bill to ban any state laws regulating AI failed by one vote, but Trump later issued executive orders attempting to do that without congressional support. See: "Tony Romm and Colby Smith, "Chasing an Economic Boom, White House Dismisses Risks of A.I.," *New York Times*, December 24, 2025; Guy Laron, "Trump's Road to Riyadh: The Geopolitics of AI and Energy Infrastructure," *American Affairs* 9, no. 3 (Fall 2025): 3–24.