Why AI investment looks like a bubble but may just be replacement spending

Jonathan Rice

European Systemic Risk Board

Combined capital spending by Amazon, Alphabet, Microsoft, and Meta hit $97 billion in the third quarter of 2025 alone, six times the quarterly run rate of 2018. Meta has guided $115–135 billion for 2026. Headlines frame this as a bubble. But a large share of that spending is not building new capacity. It is replacing hardware that has already become obsolete.

That distinction matters. AI accelerators—the specialised chips that train and serve large models—lose economic value far faster than conventional capital. GPU rental rates have fallen roughly 70 per cent in two years. Secondary-market prices for an H100 purchased eighteen months ago sit well below half the original cost. Taking rental-market and resale evidence together, economic depreciation of AI hardware runs at something like 30–40 per cent per year, an order of magnitude faster than the 6 per cent annual rate used for conventional equipment in standard macro models.

Most AI spending is standing still

A simple identity makes the point concrete. In any period, gross investment equals replacement of worn-out capital plus net additions to the stock:

$$Investment = Replacement + Net\ additions$$

When annual depreciation is 6 per cent, replacement is a modest background flow. When it is 33 per cent, replacement *is* the headline number. In a steady state, a firm that needs to maintain a $100 billion compute stock at 33 per cent annual depreciation must spend $33 billion per year just to stand still. A 20 per cent cut to gross spending in that world wipes out net additions entirely, without the stock falling by more than a few per cent.

Figure 1 shows quarterly capital expenditure for the four largest cloud infrastructure firms from 2015 to late 2025. The acceleration since 2023 is dramatic. But Figure 2 decomposes that spending into replacement and net additions using an illustrative perpetual-inventory calculation. Industry estimates suggest roughly 60 per cent of hyperscaler capex goes to servers and compute hardware, with 40 per cent on buildings and infrastructure (based on Alphabet CFO disclosures and TrendForce data). The decomposition applies high depreciation (∼33 per cent annual) to the compute share and conventional depreciation (∼5 per cent annual) to structures. Even with this more conservative split, replacement spending—the grey areas—accounts for roughly a third of spending during the current boom, and would dominate rapidly if capex growth slowed. The implied compute stock, shown on the right axis, is far smoother than the investment flow that produces it.

Two speeds of adjustment

This pattern—volatile investment, smooth stock—is exactly what growth models predict when one asset depreciates much faster than the other. Investment-specific technological change has long been recognised as a major driver of business cycles (Greenwood, Hercowitz, and Krusell 2000). When combined with heterogeneous capital durability, this generates two dis-
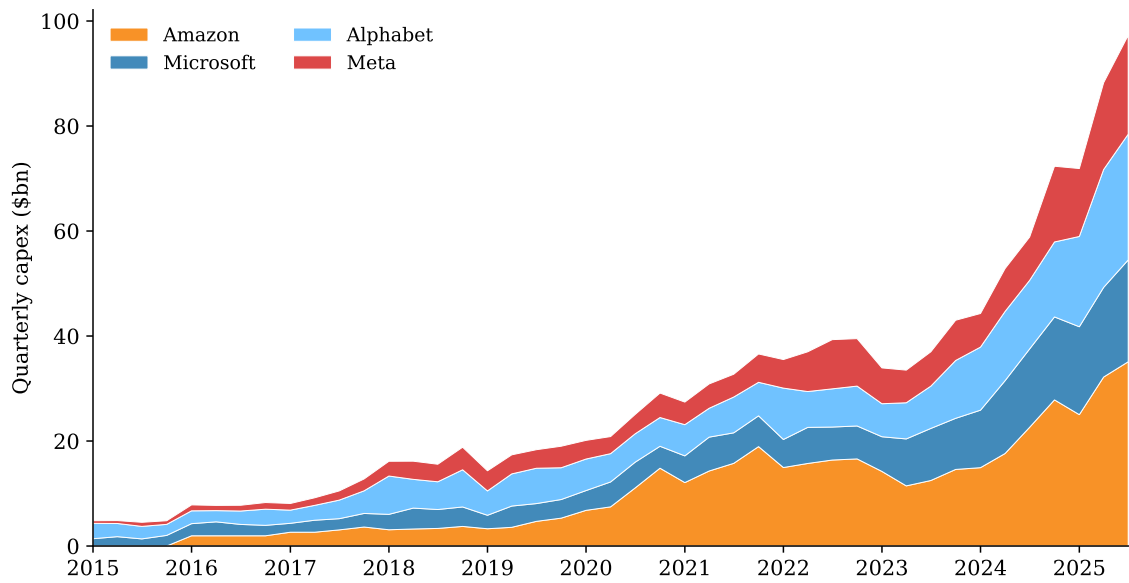
Figure 1: Hyperscaler capital expenditure, 2015–2025. *Source: SEC EDGAR filings (Amazon, Alphabet, Microsoft, Meta Platforms). Quarterly "payments to acquire property, plant and equipment" from cash-flow statements.*
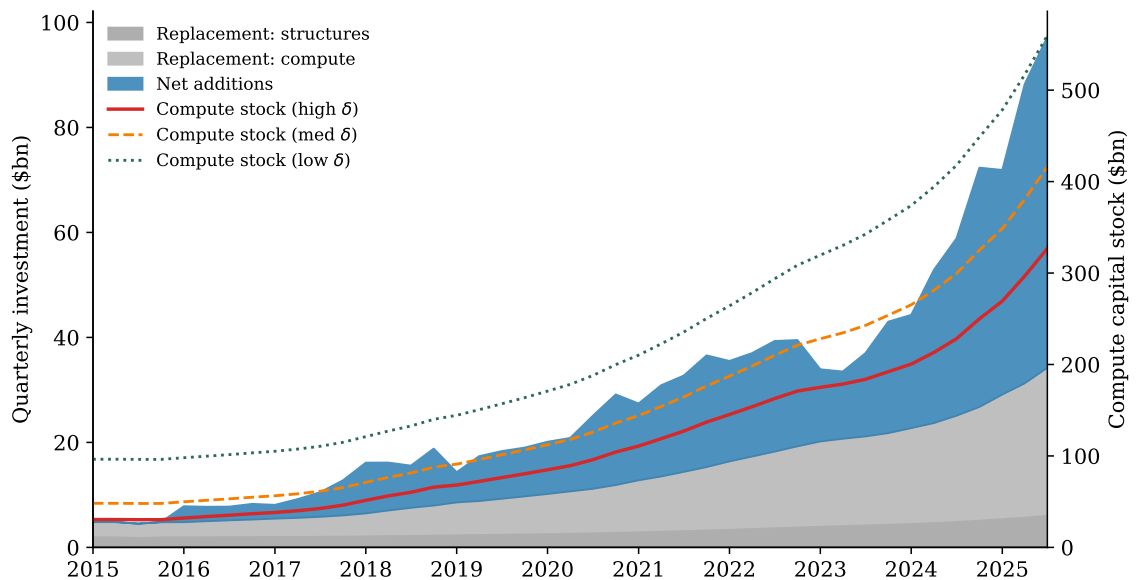

Figure 2: Decomposing AI investment: replacement versus net additions. *Source: Author's illustrative calculations using a two-asset perpetual-inventory method. Assumes 60% of capex is compute (quarterly $\delta = 0.095$) and 40% is structures (quarterly $\delta = 0.0125$). Stock lines show compute capital under alternative depreciation assumptions. The exercise is indicative—actual hardware shares and depreciation rates vary across firms and time.*

tinct speeds of adjustment (Rice 2026). AI capacity reverts to trend with a half-life of about seven quarters, because the high rate of ongoing replacement gives the economy a built-in stabiliser: modest shifts in gross spending translate quickly into stock corrections. Conventional capital, depreciating at 6 per cent annually, adjusts over roughly a decade. The fast mode means that even large swings in measured AI investment can be consistent with an economy that is close to its optimal path.

This has a direct parallel. Communications-equipment investment surged from $62 billion to $135 billion annually between 1996 and 2000, then collapsed. The episode is often cited as pure speculative excess. But telecom switching equipment also depreciated rapidly, and much of the late-1990s investment was replacing earlier-generation hardware. The AI cycle has the same structural feature, with shorter hardware lives and correspondingly larger replacement shares.

Why hardware durability matters for the economy

Rapid obsolescence is not just an accounting curiosity. It has first-order consequences for welfare and growth.

Welfare is sensitive to durability. In the calibrated model (Rice 2026), extending AI hardware's economic life from three years to four raises steady-state welfare by 0.36 per cent in consumption-equivalent terms. The intuition: when hardware is short-lived, a large share of resources goes to maintaining the stock rather than expanding it. Anything that extends hardware life—better engineering, software efficiency, or reduced thermal stress—frees resources for consumption or additional capacity.

Investment-specific shocks propagate fast. Shocks to the productivity of investment goods drive a substantial share of business-cycle fluctuations (Greenwood, Hercowitz, and Krusell 2000). When the capital stock turns over quickly, those shocks propagate faster. A breakthrough in chip fabrication has immediate effects on the optimal compute stock, and the economy reaches the new optimum within years, not a decade. The same logic works in reverse: export controls or supply disruptions bite quickly.

Mismeasurement distorts productivity accounting. If statistical agencies understate economic depreciation, they overstate effective capital growth and understate total factor productivity. Brynjolfsson, Rock, and Syverson (2017) argued that AI's productivity effects follow a J-curve. The depreciation mismatch adds a further wrinkle: even tangible hardware investment may be mismeasured if accounting lives diverge from economic lives.

Digital services expand while the labour share holds. In the two-asset framework, AI capacity delivers low-marginal-cost digital services that grow to a substantial share of consumption, yet the aggregate labour share remains close to historical norms. The depreciation rate governs how large the AI stock can be and therefore how far digital services can expand—without the labour-share collapse that some automation narratives predict.

Policy implications

Measurement. Statistical agencies classify AI accelerators with general-purpose equipment, using five-to-six-year depreciation schedules. If true economic depreciation is closer to three years, measured investment overstates capacity growth and distorts productivity attribution. Germany's 2021 reform, setting computer-hardware useful life to one year, is closer to the mark.

Financial stability. Over $200 billion in data-centre debt has been extended. If hardware collateral depreciates at 33 per cent per year, it enters negative equity within two to three years. The Bank of England has opened an inquiry. Lenders should distinguish the long-lived building from the short-lived hardware inside.

Fiscal policy. When replacement flows are large, small changes in the user cost of compute move the steady-state stock substantially and transitions are fast. A two-percentage-point depreciation wedge shifts the optimal AI stock by around a fifth. Compute taxation is high-leverage and fast-acting—a powerful but risky instrument.

A better way to read the capex numbers

The next time a hyperscaler announces a surge—or a cut—in AI spending, ask not what it implies for the *flow* of investment, but what it implies for the *stock* of compute relative to replacement needs. A spending cut that looks dramatic in flow terms may barely register in the stock, because depreciation would have done most of the work anyway. Conversely, the current boom is building a stock that will require enormous ongoing replacement spending just to maintain. The real question is not whether the spending will slow. It is whether it will slow to below replacement, and for how long.

References

Brynjolfsson, E., D. Rock and C. Syverson (2017), "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics", NBER Working Paper 24001.

Goldman Sachs (2025), "Why AI Companies May Invest More than $500 Billion in 2026", Goldman Sachs Research.

Greenwood, J., Z. Hercowitz and P. Krusell (2000), "The Role of Investment-Specific Technological Change in the Business Cycle", *European Economic Review* 44(1): 91–115.

Korinek, A. and L. Lockwood (2025), "Public Finance in the Age of AI: A Primer", in *The Economics of Transformative AI*, University of Chicago Press.

Kshirsagar, M. (2025), "Lifespan of AI Chips: The $300 Billion Question", Princeton CITP Blog.

Rice, J. (2026), "The Macroeconomics of AI Capacity: Insights from a Two-Asset Growth Model", ESRB Working Paper.

Further reading on VoxEU:

Acemoglu, D. and P. Restrepo (2016), "The race between machines and humans: Implications for growth, factor shares and jobs", VoxEU.org, 5 July.

Acemoglu, D. and P. Restrepo (2017), "Robots and jobs: Evidence from the US", VoxEU.org, 10 April.

Bounfour, A., K. Edamura, T. Ishikawa, T. Miyagawa, A. Nonnis and K. Tonogi (2025), "New developments in total factor productivity measurement: Evidence from the productivity J-curve", VoxEU.org, 12 March.

Danielsson, J. (2025), "Artificial intelligence and stability", VoxEU.org, 6 February.

Haskel, J. and S. Westlake (2018), "Productivity and secular stagnation in the intangible economy", VoxEU.org, 23 January.

The views expressed are the author's and do not necessarily reflect those of the European Systemic Risk Board.