NBER WORKING PAPER SERIES

WHAT INVESTMENT DATA IMPLIES ABOUT THE AI TRANSITION

Jessica Wachter
Jonathan Wachter

Working Paper 35290
http://www.nber.org/papers/w35290

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2026

We thank Sylvain Catherine, Gary Gensler, Alex Imas, Max Miller, Stijn Van Nieuwerburgh, Fernando Stein, and seminar participants at Wharton for helpful comments. The information, views, and opinions expressed herein are solely those of the authors and do not necessarily represent the views of Point72, its affiliates, or the National Bureau of Economic Research. Point72 and its affiliates are not responsible for, and did not verify for accuracy, any of the information contained herein.

What Investment Data Implies about the AI Transition
Jessica Wachter and Jonathan Wachter
NBER Working Paper No. 35290
June 2026
JEL No. E22, G12, O33

ABSTRACT

The five largest U.S. technology firms spent $380 billion on capital expenditure in 2025 and are forecast to spend roughly double that in 2026. These firms risk bankruptcy unless expected profits grow commensurately. We embed this observation in a two-sector open-economy model with rare productivity booms. We calibrate the boom size to match the observed increase in investment projected through 2027, implying that a boom raises AI-sector productivity by a factor of roughly 2.7. We then calibrate a two-year window of a 50% annual probability of an increase of the same magnitude, generating a range of scenarios consistent with the wide variety of industry forecasts, along with an elevated permanent probability tied to the valuation of the aggregate market. The implied additional cumulative GDP growth ranges from 5 to 58 percentage points by 2030, with AI shares of the economy ranging from 8% to 39%. Long-term annual growth is in expectation approximately 7% but with substantial risk. With risk aversion of 3, and an elasticity of intertemporal substitution equal to 1, the risk-free rate increases by approximately half a percentage point, and the equity premium rises by approximately 3 percentage points.

Jessica Wachter
University of Pennsylvania
The Wharton School
Department of Finance
and NBER
jwachter@wharton.upenn.edu

Jonathan Wachter
Point 72
wachter@gmail.com

1 Introduction

How much will artificial intelligence contribute to economic growth? Existing answers span an enormous range. Acemoglu (2024) applies a task-based framework and concludes that AI will raise total factor productivity by approximately 0.7 percentage points over ten years. At the other extreme, Amodei (2024, 2026) envisions industry-wide infrastructure spending reaching "multiple trillions a year by 2028 or 2029," with AI compressing decades of scientific progress into years.

Our contribution is to use investment data, combined with revealed preference, and a model for rare productivity booms, to estimate future GDP growth. Assuming the marginal investment is (conservatively) zero net present value implies a near-term increase in productivity of a factor of roughly 2.7 – this is a rare boom of the type studied by Tsai and Wachter (2016). We then use industry forecasts to 2030 to argue that firms are considering additional booms – but that these are far from certain. Finally, firm valuations discipline a long-term rare boom probability that undergoes a structural shift.

This approach complements the growing literature that focuses on task-based analyses, such as Acemoglu (2024), Babina et al. (2024), and Yotzov et al. (2026). The productivity implications from the current investment numbers are significantly greater than those implied by the task-based analyses, even those using more recent data. A second strand of the literature (Korinek and Suh, 2024, Restrepo, 2025, Trammell and Korinek, 2025) focuses on the theoretical question of how wages might respond as human work becomes less necessary. A limiting case is one in which the physical and intelligence sectors are gross complements (Kording and Marinescu, 2025), implying that AI-related growth eventually hits a wall. The extent to which the two sectors are gross complements is itself highly uncertain, and our model captures this possibility through the extreme uncertainty regarding whether further booms arrive or not. Our paper also relates to recent work by Chen (2026), who focuses on the possibility of the singularity as a left-tail event for agents in the economy. Rather than assuming the singularity, we back it out from investment data of value-maximizing firms using a revealed preference argument. Finally, Van Nieuwerburgh (2026) discusses the financing mechanism. Our approach differs from these in that we focus on implications for productivity and the share of AI in the economy, assuming that wages and financing frictionlessly adjust.

Section 2 documents the facts. A small number of firms—Amazon, Alphabet, Microsoft, Meta, and Oracle—account for the vast majority of AI infrastructure spending. Their combined capital expenditure was \$155 billion in 2022, rose to \$226 billion in 2024, and is forecast at \$755 billion for 2026. Each of these figures comes from management disclosures in SEC filings or earnings releases. Section 2 reports these figures company by company and places them in the context of the aggregate economy.

Section 3 develops a two-sector model with rare productivity booms. The AI sector is exposed to the boom; the non-AI sector is not. The initial boom entails an unanticipated jump in productivity, to which optimizing firms respond with a surge in investment. To justify the forecasted 2027 investment, the boom size is roughly a multiple of 2.7 of current productivity. This boom size informs the remaining calibration. With no other changes to productivity, this multiple alone translates into growth of 5 percentage points between now and 2030.

However, industry forecasts, together with historical experience regarding other fundamental technological changes, suggest that the current productivity boom is not the end of the story. The rare boom model of Tsai and Wachter (2016) provides a means of capturing the skewness and optionality inherent in the future trajectory of AI. We model occurrence of future advances as the outcome of a Bernoulli process with maximal uncertainty. We trace out a range of scenarios over 2028–2030, which we refer to as moderate, transformative, and singularity. The moderate scenario corresponds to the 5 percentage points of growth mentioned above, already an order of magnitude above task-based estimates. In the singularity case, growth is approximately 40 percentage points.

We then use the model, applied to current stock market valuations, to obtain a back-of-the-envelope annual probability of a boom after 2030. This probability is approximately 4 percent per annum. This in turn allows us to simulate the very long run, in which, under the moderate scenario, growth accumulates to 30 percentage points by 2050 along the expected path, but with a 90th percentile value of 58%, whereas for the singularity, the expected path corresponds to 200% growth, with a 90th percentile of about 400%. Not surprisingly, this level of uncertainty has strong implications for the riskfree rate and equity premium. Despite the higher expected growth, the equilibrium riskfree rate barely increases, and the equity premium increases by 3 percentage points due to uncertainty. We map out these implications in Section 5. Finally, Section 6 places the AI boom in the context of historical growth episodes. Section 7 concludes.

2 Data

The capital expenditure data that motivate the model come from a small number of publicly traded firms. Because AI infrastructure investment is highly concentrated, it is possible to measure the sector's investment directly from regulatory filings and management forecasts.

2.1 Capital expenditure by firm

Table 1 reports annual capital expenditure for the five firms that account for the majority of AI infrastructure spending: Amazon, Alphabet, Microsoft, Meta, and Oracle. The 2022–2024 figures are drawn from annual reports (10-K filings with the SEC). The 2025–2026 figures reflect quarterly earnings call disclosures and, where available, subsequently filed

10-K data.[1]

Table 1: Capital expenditure of major AI infrastructure firms ($ billions)

	2022	2023	2024	2025	2026 (forecast)	2027 (est.)
Amazon	64	53	83	132	200	285
Alphabet	31	32	53	91	180	255
Microsoft†	24	28	44	65	175	250
Meta‡	31	28	39	72	135	195
Oracle†,§	5	9	7	21	65	105
Total	**155**	**150**	**226**	**381**	**755**	**1,090**

Sources: Consolidated Statements of Cash Flows from each firm's Form 10-K (2022–2025). The 2026 figures are management forecasts disclosed in Form 8-K earnings releases (Amazon, Alphabet, Meta, Oracle) or sell-side compilations of earnings-call commentary (Microsoft, which does not issue annual capex guidance in any SEC filing). The 2027 figures are bottom-up estimates; no firm has issued formal 2027 capex guidance in any SEC-filed document. Each per-firm 2027 figure is anchored to filed forward-looking statements about contracted demand, lease commitments, revenue trajectories, and capacity targets, with full source documentation in Appendix A.2. †Microsoft (fiscal year ending June 30) and Oracle (fiscal year ending May 31) do not file on a calendar-year basis. Microsoft columns 2022–2025 report the fiscal year ending in the calendar year shown. Oracle columns 2022–2025 report the fiscal year ending in the calendar year shown. ‡Meta values for 2023–2025 follow Meta's management-disclosed definition, which adds principal payments on finance leases (financing-activities section) to purchases of property and equipment (investing-activities section). The 2022 value reflects purchases of property and equipment only; the 2022 finance-lease principal was not separately disclosed in the FY2024 10-K. Ranges reflect the low and high ends of management forecasts; midpoints are used for totals. §Oracle's reported capex measures capital at risk on Oracle's balance sheet rather than the total physical AI infrastructure deployed through Oracle data centers; customer-prepayment and customer-supplied-GPU arrangements disclosed in the Q3 FY2026 release are excluded. We use the reported figure throughout; see Appendix A.1.

Table 1 suggests that something unusual occurred in 2024. Aggregate capital expenditure was essentially flat at approximately $150 billion in 2022 and 2023, reflecting baseline spending on cloud computing infrastructure and general corporate investment. This baseline encompasses non-AI uses—retail fulfillment, logistics, office infrastructure, and legacy cloud

[1] For Microsoft, which does not provide formal annual capital expenditure forecasts, the 2026 range reflects a sell-side compilation of filed quarterly actuals together with the leases-not-yet-commenced disclosure in the commitments footnote. Microsoft's figures include capital leases. For all other firms, the figures represent capital expenditure as reported in the statement of cash flows. Appendix A.1 documents the underlying EDGAR filings line by line, including accession numbers, the specific cash-flow-statement line items, and the fiscal-year conventions used for Microsoft (fiscal year ending June 30) and Oracle (fiscal year ending May 31).

services—alongside any pre-boom AI spending. We use the 2022 total of \$155 billion as the pre-boom AI-sector investment in the calibration of Section 3.3; to the extent that some fraction represents non-AI corporate capital, the true pre-boom AI capital stock is smaller and the calibrated boom size $\hat{\xi}$ is correspondingly larger. Appendix A.4 discusses sensitivity to this attribution choice. In 2024, aggregate spending rose 51% to \$226 billion. The acceleration continued in 2025, with a further 69% increase to \$381 billion, and management forecasts for 2026 imply a further 98% increase to approximately \$755 billion.[2] Altogether, capital expenditure has more than tripled between 2024 and 2026. The five firms in Table 1 do not span the universe of AI infrastructure capital spend. A privately-held sixth hyperscaler (xAI, founded mid-2023) and four data-center-owning "neocloud" providers (CoreWeave, Crusoe, IREN, Lambda) add approximately \$17 billion in 2024 and \$38 billion in 2025 of capital expenditure not reflected in Table 1. These additions are fully incremental: each is a data-center owner that sells compute services to the hyperscalers under ASC 606 service-contract accounting, so hyperscaler payments to them flow through hyperscaler operating expense, not through hyperscaler capex.[3]

The 2027 column of Table 1 reports a bottom-up five-firm total of \$1,090B. Unlike the 2026 column, which reflects management forecasts disclosed in earnings releases, no firm has issued formal 2027 capital-expenditure guidance in any SEC-filed document. The 2027 figures sum five per-firm point estimates each anchored to filed forward-looking statements about contracted demand and capacity. Amazon's \$285B is grounded in the OpenAI 2 GW

[2]Management forecasts for 2026 (midpoints of disclosed ranges where relevant) are: Amazon \$200B, Alphabet \$180B, Microsoft \$175B, Meta \$135B, and Oracle \$65B. Amazon, Alphabet, Meta, and Oracle have all disclosed quantitative annual guidance in Form 8-K filings; Microsoft's figure is the midpoint of an analyst range built from quarterly disclosures and earnings-call commentary. Appendix A.2 reports the source documents.

[3]The verbatim accounting language is in CoreWeave's FY2025 10-K Note 2 (revenue recognition): contracts "do not meet the definition of a lease under ASC 842 and are accounted for as service contracts under ASC 606." The same structure applies to the three other Tier 1 neoclouds by business model; xAI owns and operates its own data centers. Appendix A.3 documents firm-by-firm sources.

and Anthropic 5 GW Trainium commitments beginning to ramp in 2027, the 2027 Trainium4 production launch, and the deployment of more than one million NVIDIA GPUs starting in 2026. Alphabet's $255B reflects multi-gigawatt TPU hardware supply agreements with revenue recognition contracted primarily in 2027, disclosed in the Q1 2026 10-Q. Microsoft's $250B is anchored by the leases-not-yet-commenced disclosure, which stepped from $92.7B (June 2025) to $155.1B (December 2025) to $196.6B (March 2026), with commencement spanning fiscal years 2026–2031. Meta's $195B reflects the disclosed intent to continue increasing investment alongside the Superintelligence Labs initiative. Oracle's $105B reflects raised FY2027 total revenue guidance ($90B, from $67B in FY2026) and the $50B February 2026 financing program sized specifically for the FY2027 buildout. Appendix A.2 documents the source materials firm by firm.

2.2 AI investment relative to the aggregate economy

Table 2 places these figures in the context of the U.S. macroeconomy.

Table 2: AI infrastructure investment relative to the U.S. economy

	2022	2024	2025	2026 (forecast)	2027 (est.)
AI infrastructure capex ($B)	155	226	381	755	1,090
U.S. nominal GDP ($T)	25.5	29.2	∼30.5	∼31.5	∼32.4
U.S. gross private fixed investment ($T)	∼4.7	∼5.2	∼5.4	∼5.5	∼5.7
Share of GDP (%)	0.6	0.8	1.2	2.4	3.4
Share of private fixed investment (%)	3.3	4.3	7.1	13.7	19.2

Sources: GDP and gross private fixed investment from the Bureau of Economic Analysis (NIPA Table 1.1.5; FRED series FPIA). 2025–2027 GDP and aggregate investment are author projections assuming 3% nominal growth. The 2027 capex figure is the bottom-up five-firm total from Table 1; Appendix A.2 documents the per-firm sources.

Three measurement caveats apply to Table 2. First, the corporate capital expenditure values are consolidated global totals, while GDP and gross private fixed investment are U.S.

national accounts aggregates; the ratios are therefore approximate. Second, measurement conventions differ across the five firms. Microsoft's reported line is the only one that includes finance-lease additions paid in cash at commencement, and the Microsoft 2026 figure additionally reflects an analyst compilation rather than a filed management point or range. Meta's reported capex follows a management definition that adds principal payments on finance leases to investing-activities purchases of property and equipment; Amazon, Alphabet, and Oracle report cash-flow capital expenditure. Appendix A.1 reconciles the firm-by-firm conventions to verbatim cash-flow-statement values. Residual differences after this reconciliation are small relative to the order-of-magnitude growth in AI infrastructure investment that the calibration is intended to capture. Third, all investment values are nominal. We treat them as real throughout the analysis: cumulative inflation over 2024–2027 is on the order of 8–12%, small relative to the orders of magnitude at stake.

AI infrastructure investment rose from 3.3% of total U.S. private fixed investment in 2022 to a projected 14% in 2026. As a share of GDP, it rose from 0.6% to 2.4%. As a benchmark, the 2025 share of 1.2% sat just below the peak levels of the late-1990s telecommunications investment cycle (approximately 1.5% of GDP); the 2026 forecast implies that AI investment surpasses those peak levels by a wide margin. AI investment has also become a significant driver of aggregate output: it accounted for approximately one-fifth of the 2.2% year-over-year increase in real GDP in the fourth quarter of 2025. Excluding AI spending, corporate equipment investment would have been negative.[4]

3 A Two-Sector Model with Rare Booms

This section develops a two-sector model in which the AI sector is exposed to rare productivity booms and the non-AI sector is not. The production framework builds on Gourio

[4] Pantheon Macroeconomics, U.S. Economic Monitor, February 2026.

(2012) and Gomes et al. (2018), who study firms subject to rare disasters. We depart from those models in two ways. First, we replace disasters with booms. Second—and more consequentially for investment—we break the assumption, introduced by Gourio (2012), that the shock moves productivity and the capital stock proportionally. In the disaster setting, the shock destroys both z and K in proportion, leaving the marginal product of capital unchanged; investment does not respond. Here, the boom raises productivity but leaves the physical capital stock intact, so that the marginal product of capital jumps upward and firms face a large gap between current and optimal capital. This asymmetry is the source of the investment surge. Section 4 adds adjustment costs to produce realistic multi-year dynamics.

3.1 Two sectors

The economy consists of two sectors: AI (a) and non-AI (n). Each sector $j \in \{a, n\}$ produces output using a Cobb-Douglas technology:

$$Y_{jt} = z_{jt}^{1-\alpha} K_{jt}^{\alpha}, \tag{1}$$

where α is the capital share, z_{jt} is sector-specific productivity, and K_{jt} is the sector's capital stock. Capital in each sector evolves according to

$$K_{j,t+1} = (1 - \delta_j) K_{jt} + I_{jt}, \quad j \in \{a, n\} \tag{2}$$

where I_{jt} is gross investment and δ_j is the sector-specific depreciation rate.

Productivity in the AI sector is subject to rare booms. In each period, with probability p_t, a boom occurs: AI productivity jumps by a factor e^{ξ}, where $\xi > 0$ represents the log change in productivity due to the boom. The non-AI sector is unaffected. Formally:

$$z_{a,t+1} = \begin{cases} z_{at}\, e^{\xi} & \text{with probability } p_t, \\ z_{at} & \text{with probability } 1 - p_t, \end{cases} \tag{3}$$

For convenience, we assume neither sector is affected by normal-times volatility or by drift; the effect of these would be small in comparison to rare booms. We will also allow p_t to vary over time; in the calibration below, the variation takes the form of a structural break. It is p_t that determines the probability of the boom that occurs at time $t+1$. Agents choose investment I_{jt}, and hence capital K_{t+1}, with knowledge of z_{at} and p_t (which determines the distribution of productivity at time $t+1$), but without knowledge of $z_{a,t+1}$ and p_{t+1}. In the frictionless setting of Section 3.2, this timing assumption implies that current investment depends on the current one-period probability p_t but not on the anticipated path of p_{t+s} for $s \geq 1$; once adjustment costs are introduced in Section 4, the anticipated path does enter the current investment decision.

3.2 Optimal investment: the frictionless case

Sector a maximizes the present discounted value of earnings, assuming discount rate r_a:

$$V_{at} = \max_{\{I_{as}\}_{s \geq t}} \mathbb{E}_t \sum_{s=t}^{\infty} \frac{1}{(1+r_a)^{s-t}} \left[z_{as}^{1-\alpha} K_{as}^{\alpha} - I_{as} \right], \tag{4}$$

subject to $K_{a,s+1} = (1-\delta_a)K_{as} + I_{as}$ and the productivity process (3). As is standard, we assume firms choose investment I_{at} at time t, determining $K_{a,t+1}$ through the law of motion (2) before $z_{a,t+1}$ is realized.

Sectors face different exogenous required rates of return, r_a and r_n, reflecting differences in risk. Following an "open economy" interpretation, we treat these as exogenous—determined, for instance, by the required return of global investors willing to supply capital elastically at price r_j. In Section 5, we derive implications of rare booms for equilibrium rates of return.

For the firm's problem to have a finite solution, the expected gross growth factor of productivity must be below the gross discount rate:

$$1 + r_a > (1 - p) + p\, e^{\xi}. \tag{5}$$

We verify that this condition holds at the calibrated parameters below.

The first-order condition with respect to $K_{a,t+1}$, or equivalently, $I_{a,t}$, equates the marginal cost of a unit of capital (one, since there are no adjustment costs) to the discounted expected marginal product plus the continuation value of the undepreciated unit:

$$1 = \frac{1}{1 + r_a}\, \mathbb{E}_t \left[\alpha \left(\frac{z_{a,t+1}}{K_{a,t+1}} \right)^{1-\alpha} + (1 - \delta_a) \right],$$

which simplifies to

$$\mathbb{E}_t \left[\alpha \left(\frac{z_{a,t+1}}{K_{a,t+1}} \right)^{1-\alpha} \right] = r_a + \delta_a. \tag{6}$$

The expected marginal product of capital equals the user cost: the required return plus depreciation. Because $K_{a,t+1}$ is determined at t, the expectation is over the boom. Evaluating:

$$\alpha \left(\frac{z_{at}}{K_{a,t+1}} \right)^{1-\alpha} \underbrace{\left[(1 - p_t) + p_t\, e^{(1-\alpha)\xi} \right]}_{\equiv \Omega(p_t)} = r_a + \delta_a, \tag{7}$$

where we define the boom option value:

$$\Omega(p) \equiv (1 - p) + p\, e^{(1-\alpha)\xi}. \tag{8}$$

Note that $\Omega(p) \geq 1$, with equality when $p = 0$. A higher boom probability raises the effective expected productivity of capital, making investment more attractive.

Solving (7) for optimal capital:

$$K^*_{a,t+1} = z_{at} \left(\frac{\alpha\, \Omega(p_t)}{r_a + \delta_a} \right)^{1/(1-\alpha)}. \tag{9}$$

The optimal capital stock is proportional to current productivity z_{at} and increasing in the boom option value $\Omega(p_t)$. The non-AI sector satisfies the analogous condition with $\Omega = 1$ (no boom exposure):

$$K^*_{n,t+1} = z_{nt} \left(\frac{\alpha}{r_n + \delta_n} \right)^{1/(1-\alpha)}. \tag{10}$$

Proposition 3.1 (Capital ratio)**.** *Suppose a boom occurs at time t, so that $z_{a,t} = z_{a,t-1}\, e^{\xi}$, and that the boom probability shifts from p_{t-1} to p_t. Then*

$$\frac{K^*_{a,t+1}}{K^*_{a,t}} = e^{\xi} \left(\frac{\Omega(p_t)}{\Omega(p_{t-1})} \right)^{1/(1-\alpha)}. \tag{11}$$

Proof. From (9), $K^*_{a,t+1} \propto z_{at} \cdot \Omega(p_t)^{1/(1-\alpha)}$. The boom multiplies z_a by e^{ξ}, and the structural break shifts p from p_{t-1} to p_t. The $r_a + \delta_a$ term cancels because it appears in both numerator and denominator. □

Under the open-economy assumption, the capital ratio depends only on the boom size ξ and the shift in p, not on r_a. This means we can identify ξ from investment data without taking a stand on the required return.

How a boom affects sectoral output. Because $K_{a,t+1}$ is chosen at time t, capital is predetermined when the boom is realized at $t + 1$. The output response therefore occurs in two stages.

In the *impact period* $(t + 1)$, capital is fixed at $K^*_{a,t+1}$ but productivity has jumped to

$z_{at}\, e^{\xi}$. Output is

$$Y_{a,t+1} = (z_{at}\, e^{\xi})^{1-\alpha}\, (K^*_{a,t+1})^{\alpha} = e^{(1-\alpha)\xi}\, z_{at}^{1-\alpha}\, (K^*_{a,t+1})^{\alpha}.$$

Only the fraction $(1-\alpha)$ of the productivity gain is reflected in output because the capital share α of the production function operates on the predetermined capital stock. At $\alpha = 1/3$ and $\xi = 0.99$, $e^{(1-\alpha)\xi} = 1.93$: output rises 93%.

In the *adjustment period* $(t+2)$, firms observe the new productivity and set $K^*_{a,t+2} = e^{\xi}\, K^*_{a,t+1}$ (from (9), since $K^* \propto z$). Output becomes

$$Y_{a,t+2} = (z_{at}\, e^{\xi})^{1-\alpha}\, (e^{\xi}\, K^*_{a,t+1})^{\alpha} = e^{\xi}\, z_{at}^{1-\alpha}\, (K^*_{a,t+1})^{\alpha}.$$

The full factor e^{ξ} is now realized: with $\xi = 0.99$, $e^{\xi} = 2.69$. The remaining factor of $e^{\alpha\xi} = 1.39$ comes from capital adjustment.

Thus α governs the split between the immediate and delayed output response. A higher capital share means more of the productivity gain is deferred until capital adjusts. In the frictionless case, output jumps discretely at boom dates (the $(1-\alpha)\xi$ component) and then rises again one period later as capital adjusts (the $\alpha\xi$ component).

3.3 Calibration of the frictionless model

Table 3 summarizes the assumed parameter values. The non-AI depreciation rate δ_n is the standard value in macroeconomic calibrations, based on the ratio of aggregate depreciation to the capital stock in the BEA Fixed Asset Tables. The AI-capital depreciation rate δ_a corresponds to a four-year useful life for server and GPU equipment. Appendix A.4 discusses both calibrations in detail. Appendix A.5 reports the model's quantitative implications under an alternative two-component decomposition of AI capital following Van Nieuwerburgh

(2026); the qualitative conclusions of the paper are unchanged. The required return on AI-sector capital follows the baseline discount rate of Vanguard Group (2025) (range 10–25%).[5]

Table 3: Assumed parameter values

Parameter	Value	Description
α	1/3	Capital share
δ_a	0.25	AI-capital depreciation rate
δ_n	0.07	Non-AI capital depreciation rate
r_a	0.15	Required real return on AI-sector capital
r_n	0.05	Required real return on non-AI capital

Before the boom, we assume both sectors are in steady state, so that investment exactly replaces depreciation: $I_j = \delta_j K_j$. The pre-boom AI capital stock is then $K_{a0} = I_{a,\text{pre}}/\delta_a = 155/0.25 = \$620\text{B}$, where \$155B is the combined capital expenditure of the five largest AI firms in 2022 (Table 1). For the non-AI sector, aggregate U.S. gross private fixed investment was approximately \$4.7 trillion in 2022 (Table 2), giving $K_{n0} = (4{,}700 - 155)/0.07 \approx \$65{,}000\text{B}$. The AI sector's share of the total capital stock is therefore only 0.9%, despite accounting for 3.3% of investment—the wedge arises entirely from the difference in depreciation rates. Appendix A.4 discusses the data sources for both the depreciation rates and the investment figures in detail.

2024–2027 Analogously to the rare-disasters approach of Barro (2006), we simplify the calibration by collapsing a multi-period event into an instant of time in our initial model. While the model will not succeed in matching dynamics, it will transparently convey orders of magnitude.

In the frictionless model, capital adjusts immediately to its optimal level each period (Section 4 incorporates frictions). Applying the first-order condition (7) with $p \approx 0$ (so that

[5] Because the calibrated boom size $\hat{\xi}$ is identified from investment ratios in which $r_a + \delta_a$ cancels (Proposition 3.1), it is invariant to the choice of r_a. At $r_a = 0.10$, the finiteness bound tightens to $p_* < 5.9\%$ and the P/E at $p_* = 0.04$ rises to 34, at the upper edge of the plausible range of 15–35.

$\Omega \approx 1$) both before and after the boom gives $K^*/K_0 = e^\xi$, i.e. the optimal capital stock scales linearly with z. The entire gap between K^* and the inherited capital stock must be closed in the first year. Subsequent investments are then to make up for depreciation. That is, total investment during the 2024–2027 period equals:

$$\bar{I} = K_0\big(\delta_a + e^\xi + 3\delta_a e^\xi - 1\big) \tag{12}$$

The first term makes up for depreciation, the second corresponds to the productivity boom, and the third reflects the need to make up for depreciation in the subsequent 3 years.

Therefore, our revealed-preference estimate of ξ equals

$$e^\xi = \left(\frac{\bar{I}}{K_0} + 1 - \delta_a\right)\frac{1}{1+3\delta_a} \tag{13}$$

Table 1 shows investment unchanged between 2022 and 2023, and a rapid increase thereafter. We calibrate ξ by matching the model's cumulative investment to the observed capital expenditure of \$2,452B over 2024–2027 ($= 226 + 381 + 755 + 1090$; see Table 1). That is, we interpret this investment as the firms' response to a productivity increase that occurred prior to 2024. As (9) shows, an increase in productivity by the factor e^ξ implies a gap between actual and desired capital. Substituting in

$$e^{\hat{\xi}} = \left(\frac{2452}{620} + 0.75\right)\frac{1}{1.75} \approx 2.69, \qquad \hat{\xi} \approx 0.99.$$

This is, approximately, the cumulative increase in productivity required to justify this level of investment, assuming no change in discount rates, and zero economic profits.[6] Were the boom to be smaller than this, the investments would be NPV-negative.

[6] Assuming positive economic profits would increase the size of the boom. Moreover, we argue below that the discount rate if anything, increased during this period.

To capture the range of possibilities within the model, we consider that over a short period of time, p is at the level corresponding to maximum uncertainty, that is, $p = 0.5$.[7] If no booms realize, then there will be no motivation for investment beyond depreciation, however, if one or perhaps two booms realize during this specified window, then the projected investment (or more) is a rational response. We assume:

$$p_t = \begin{cases} p_{\text{pre}} \approx 0 & \text{for } t < 2028, \\ p_{\max} & \text{for } t \in \{2028, 2029\}, \\ p_* & \text{for } t > 2029. \end{cases} \tag{14}$$

Note that the elevated boom probability refers to the productivity boom of the *following* year. Thus $p_{2028} = p_{\max}$ means there is a 50% probability of a boom in 2029, and $p_{2029} = p_{\max}$ means there is a 50% probability of a boom in 2030. Because the elevated-probability window begins in 2028, after the calibration period, 2024–2027 investment is driven entirely by the initial boom. The calibration of ξ is therefore independent of $p_{\max}$ and p_*.

Several comments are in order. First, it is admittedly unrealistic to have a structural break (14) in 2028 unanticipated in 2025. This assumption allows us to transparently calibrate ξ to 2024–2027 investment. If a shift in the probability were anticipated earlier, then this would also drive some of the investments. The effect would be a somewhat lower value of ξ. Second, the assumption $p_{\text{pre}} \approx 0$ is for convenience; it cannot be exactly zero because then a boom could not have occurred. It is realistic to believe it is low as very few foresaw

[7] Existing forecasts of transformative AI exhibit very large uncertainty over the relevant horizon. Metaculus community forecasts place the median date for artificial general intelligence at approximately 2028, with a 25th–75th percentile range from late 2026 to 2031 (Metaculus question 5121, accessed March 2026); this median has fallen from approximately 2060 to 2028 over the past four years. Grace et al. (2024) survey 2,778 AI researchers and find a 50% probability of machines outperforming humans at every task by 2047, with a 10% probability by 2027—a 13-year acceleration relative to the same survey conducted in 2022. The interquartile range across individual responses spans decades. Section 3.4 compares the model paths to specific industry forecasts.

the growth potential of AI in 2024 (and even those who did revised their views upward). Note that it does not affect our prior calculations because we do not take a stance that this value of p changed between 2024 and 2027. Finally, the order-of-magnitude implications of our theory do not depend on a particular choice of interval. The advantage of picking this one is that there are forecasts, however disputed, and that some finite interval allows us to have a probability that captures the full measure of the uncertainty without violating transversality conditions leading to infinite values.

While we separate out the initial investment period from the ongoing uncertainty for convenience, at the core of our calculation is the idea that there is at once a boom that is followed by an increase in probability. This idea follows from Bayesian updating (Wachter and Zhu, 2025), and is also consistent with recency playing a role in expectations formation (Wachter and Kahana, 2024).

Assuming a probability of $p_{\max} = 0.50$ gives maximal uncertainty.[8] With two independent draws at $p_{\max} = 0.50$, there are three outcomes:

Scenario	Additional booms	Probability
Moderate	0	$(1 - 0.5)^2 = 0.25$
Transformative	1	$2 \times 0.5 \times 0.5 = 0.50$
Singularity	2	$0.5^2 = 0.25$

The moderate scenario is that the initial boom was a one-time event and no further breakthroughs occur. The transformative scenario is that exactly one additional boom arrives during the window. The singularity scenario is that two additional booms occur, compounding the productivity gains. In the moderate scenario, the productivity gain from the initial boom is all there is. In the transformative scenario, one additional boom raises productivity

[8] A Bernoulli random variable has maximal variance at $p = 0.50$. This assumption imposes the least informative prior on whether an additional boom occurs in any given year of the window.

by another factor of $e^{0.99} \approx 2.7$. In the singularity scenario, two additional booms raise it by $e^{2\times 0.99} \approx 7.2$ beyond the initial level. During 2029–2030, additional booms may or may not realize. After 2030, the boom probability reverts to p_*, and the economy converges to its new long-run state.

Investment paths during the window. Recall that I_t is chosen at date t after observing z_t and determines K_{t+1} via the accumulation equation (2). When p rises to $p_{\max}$ in 2028, the first affected investment decision is I_{2028}: chosen in 2028 using the option value of $p_{\max}$ but before the 2029 boom outcome is known. This investment determines K_{2029}.

2028 investment: same for all scenarios. With $p_{\max} = 0.50$, the optimal capital stock entering 2029 rises from $K^*(z_1, 0) = \$1{,}669\text{B}$ to $K^*(z_1, p_{\max}) = \$2{,}966\text{B}$, where $z_1 = z_0 e^{\xi}$ is the post-initial-boom productivity. The required investment is $I_{2028} = K^*(z_1, p_{\max}) - (1 - \delta_a)K^*(z_1, 0) = \$1{,}715\text{B}$. This is identical across all four scenarios because the 2029 boom has not yet been observed.

2029 investment: outcomes diverge. The 2029 boom outcome is now known. If no boom occurred (moderate and transformative (b)), z remains at z_1 and investment is just depreciation: $I_{2029} = \delta_a K^*(z_1, p_{\max}) = \$741\text{B}$. If a boom occurred (transformative (a) and singularity), z jumps to $z_2 = z_0 e^{2\xi}$ and the capital stock must adjust: $I_{2029} = K^*(z_2, p_{\max}) - (1 - \delta_a)K^*(z_1, p_{\max}) = \$5{,}758\text{B}$.

2030 investment: outcomes diverge again. Both the 2030 boom outcome and the reversion of p to p_* are reflected. Because $\Omega(p_*)$ is substantially below $\Omega(p_{\max})$, the optimal capital stock falls in scenarios where no new boom occurred. If a late boom raised z, investment remains elevated.

2031: investment replaces depreciation; grows or shrinks depending on the occurrence of a boom. In our calibration, this annual depreciation equals $\delta_a K^*$: \$441B (moderate), \$1,186B (transformative), \$3,192B (singularity).

While this describes a frictionless investment sequence, in reality there are substantial frictions. Ideally, one would micro-found these with a model of adjustment costs, for example, assuming the standard quadratic form (Hayashi, 1982, Jorgenson, 1963). However, such a model would predict that investment after a single shock follows a decelerating path regardless of the adjustment speed, whereas the data show acceleration. An interesting question for further research is how to model adjustment costs in the AI industry where manufacturing lead times and power infrastructure are first-order concerns. Therefore instead we use the considerations above to determine an optimal K^* and assume that from 2029 onward, $K_{t+1} = K_t + (1/\theta_a)(K^* - K_t)$ with $\theta_a = 3$ (one-third of the gap closed per year). Because θ_a does not affect K^*, the calibrated $\hat{\xi}$, output shares, and P/E ratios are unchanged. Figure 1 displays these paths, whereas Figure A.1 shows the corresponding path with no adjustment frictions.

Identifying p_* from aggregate valuations. What happens following 2030 is necessarily more speculative. We identify a long-run boom probability p_* from the aggregate equity-market price-to-dividend ratio, taking the boom size as given. As of May 2026, the S&P 500 trades at a price-to-dividend ratio of approximately 94, equivalent to a dividend yield of 1.06%. Combining the 30-year TIPS yield of 2.7% with a 5% real equity premium gives a real discount rate $r \approx 8\%$, so the Gordon decomposition $D/P = r - g$ implies a long-run real dividend growth rate of approximately 7%. Because dividends are by construction the cash flows that accrue to shareholders—net of all investment, including the expansionary investment that scales capital up after each productivity boom—this identification asks only at what rate aggregate cash flows compound. It does not require taking a stand on the level of V relative to earnings for any sector, nor on any decomposition of consolidated firm value between AI and non-AI business lines.[9] Equating this 7% to the model-implied AI-sector

[9]The investment data used to calibrate ξ are nominal, but the inflation adjustment over 2024–2027 is small relative to the orders of magnitude involved. At $r_a = 0.10$, the calibration of ξ is unaffected (Proposition 3.1);

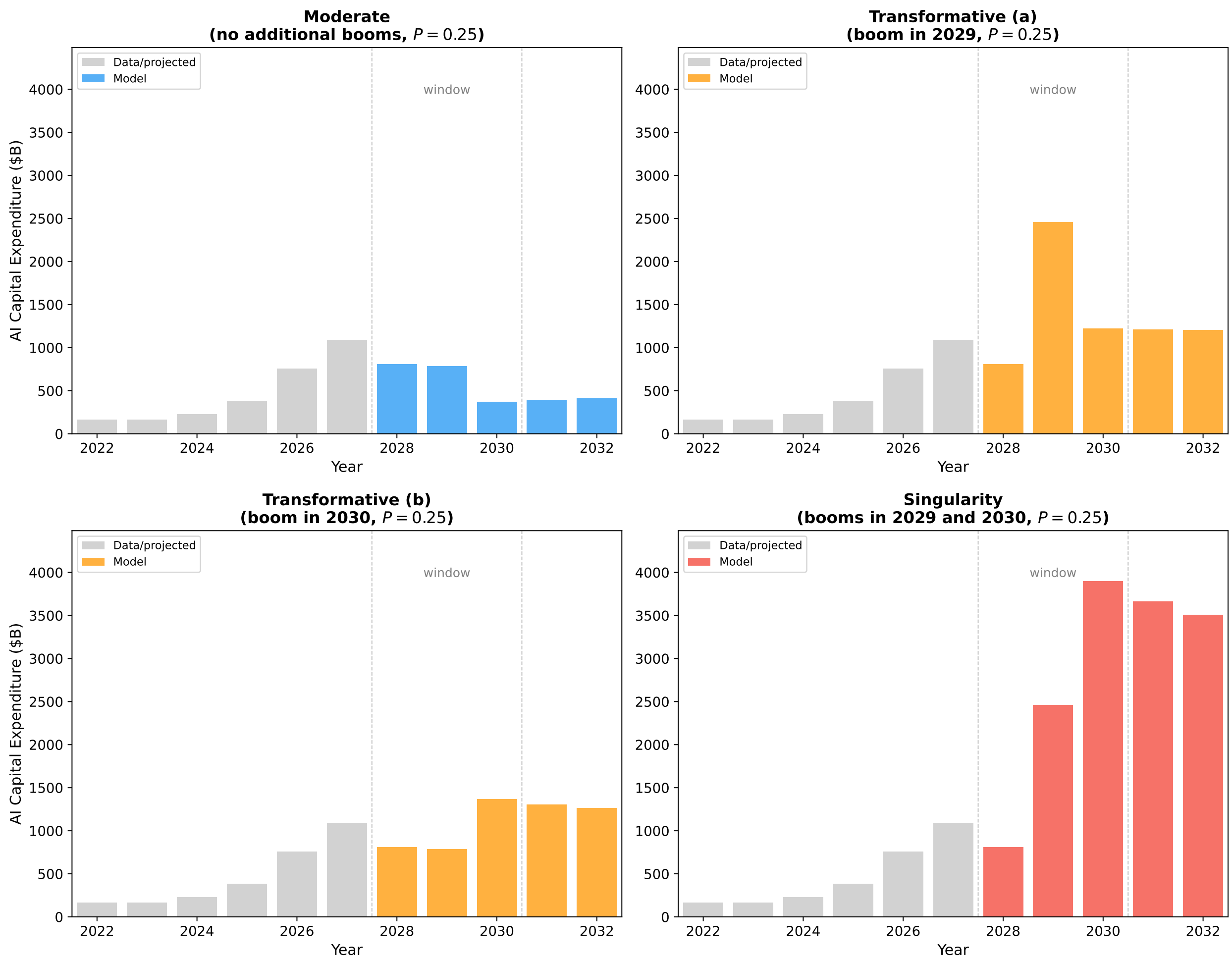

Figure 1: Investment paths by scenario. Gray bars show actual and projected AI capital expenditure through 2027. Colored bars show investment from the window onward, incorporating convex adjustment costs ($\theta_a = 3$). Dashed lines delineate the high-probability window (2028–2030). Investment in 2028 is identical across all four panels because it is chosen before the first window boom is observed. The calculation assumes no additional booms past 2030.

growth $g = p_*(e^{\xi} - 1)$ gives

$$p_* \approx 0.04, \tag{15}$$

which we adopt for the remainder of the paper.

The post-window installed capital stocks K_a^* implied by the optimality condition (9) at $p = p_*$ are \$1,763B in the moderate scenario, \$4,745B in the transformative scenario, and \$12,769B in the singularity scenario. At a working conversion of \$12.5 billion per gigawatt, these stocks correspond to approximately 141 GW, 380 GW, and 1,020 GW, respectively.[10]

3.4 Discussion

The revealed-preference argument identifies the productivity boom that *managers believe* has occurred. It does not establish that the boom has in fact occurred. Two channels could drive a wedge between the implied $\hat{\xi}$ and the true productivity gain. [11]

First, the investment may simply reflect a bubble. Managers are not immune to collective overoptimism, and the history of technology investment is replete with episodes—the late-1990s fiber-optic buildout, for instance—in which firms invested heavily in capacity that was not justified by subsequent demand. How exactly this scenario would play out in the case of AI is not known, but, it is possible, for example, that the productivity-enhancing power of AI may be a mirage. In this case, customers would be unwilling to pay for more AI usage.

the main effect is to tighten the finiteness bound and lower the implied p_*.

[10] Amodei (2026) estimates \$10–15 billion per GW. However, Van Nieuwerburgh (2026) calculates costs as high as \$40 billion. In what follows, we compare forecasts of power demand with those implied by our model. A higher cost per GW would imply ranges of investment greater than what our model would produce. Implications for GDP, however, depend on whether productivity is conceived of in units of GW versus units of dollar investment. Van Nieuwerburgh gives a forecast of 200 GW, between the moderate and transformative scenario, which at a cost of \$40 billion, gives dollar investments between transformative and singularity.

[11] Several technical considerations also raise the boom size. If the investment occurs closer to 2027, as is in fact the case, then it is not correct to include the depreciation terms in (9), thereby raising the boom size. As noted in the discussion following Table 1, calibrating to the full private-sector AI investment universe would also raise the boom size and the installed capital. However, if some of the investment were anticipatory (namely in advance of upward shifts in p_t), then this would lower the boom size, reducing the effect of the scenarios, though the amount of installed capital would be the same.

In terms of the model, this is essentially a mis-estimation of ξ in (3), resulting in negative NPV investments, declining stock valuations, failure to make interest payments on debt, and perhaps bankruptcy. The resulting disinvestment would then lower the GDP growth estimates in Section 4.

Second, the structure of the AI infrastructure market may lead to over-investment. Dasgupta and Stiglitz (1980) show that in R&D races with winner-take-all payoffs, each firm invests as if it may capture a disproportionate share of the market, producing aggregate investment in excess of the cooperative optimum. The current AI buildout has this character: the five firms in Table 1 are simultaneously building infrastructure to serve overlapping markets, and each firm's capital expenditure may reflect an expectation of capturing a larger share than is collectively feasible. If so, aggregate investment overstates the productivity gain that any single consistent set of beliefs would imply.

To this point, suppose however, that each firm were to act as if it will be the sole survivor in 2027, with no strategic interaction at all and with all investment bespoke to that firm's own AI program. The first-order condition then applies separately to each firm's own investment ratio, and the revealed-preference calibration delivers a firm-specific $\hat{\xi}_i$. Carrying out that calibration on the per-firm capex values in Table 1 gives the values in Table 4.

Thus the implied productivity jump is high even when considered on a per-firm basis. That said, if in fact not all firms survive, then installed capital will be lower, reducing GDP growth in what follows.

More broadly, just as over-optimism both about the technology and one's own prospects could lead to over-investment, managerial risk aversion, anchoring to the status quo, or the desire for the quiet life could lead to underinvestment. Our revealed-preference argument would then be conservative. The numbers we are reporting are in fact conservative compared to public statements by some industry executives. For example, Musk (2026b) forecasts "at least, five years from now, a few hundred gigawatts per year of AI in space and rising,"

Table 4: Per-firm bespoke calibration of $\hat{\xi}_i$

Firm	$I_{i,\text{pre}}$ (\$B)	$\bar{I}_i$ (\$B)	$e^{\hat{\xi}_i}$	$\hat{\xi}_i$
Amazon	64	700	2.0	0.69
Meta	31	441	2.5	0.90
Alphabet	31	579	3.1	1.13
Microsoft	24	534	3.6	1.28
Oracle	5	198	6.1	1.81
Aggregate	*155*	*2,452*	*2.7*	*0.99*

Notes: Pre-boom investment $I_{i,\text{pre}}$ is firm-i's 2022 capital expenditure from Table 1. Cumulative 2024–2027 investment $\bar{I}_i$ sums each firm's annual investments from Table 1 (2024–2026) and the bottom-up 2027 estimates in Appendix A.2. $e^{\hat{\xi}_i}$ is computed from the frictionless cumulative-investment formula of Section 3.3, $e^{\hat{\xi}_i} = (\bar{I}_i/K_{i,0} + 1 - \delta_a)/(1 + 3\delta_a)$, under each firm's own pre-boom steady-state assumption $K_{i,0} = I_{i,\text{pre}}/\delta_a$ with $\delta_a = 0.25$.

speaking in early 2026. The singularity scenario implies roughly \$13 trillion of installed capital, which at $\approx$ \$12.5 billion per GW translates to roughly 1,000 GW of total installed capacity. Musk's forecast, assuming $\delta_a = 0.25$, calls for a steady-state stock of four times "a few hundred GW"—on the order of 1,000 GW, the same order of magnitude as the singularity scenario. SpaceX's S-1 (Space Exploration Technologies Corp., 2026), filed in May 2026, sets a goal of launching 100 GW of orbital AI compute capacity per year, with first deployments "as early as 2028."

Amodei (2024) envisions "millions of instances" of AI "smarter than a Nobel Prize winner" running at "10x–100x human speed"—a "country of geniuses in a datacenter"—"as early as 2026," with the timing updated to roughly 2028 in Amodei (2026). Interpreting Amodei's numbers as 300 million more workers, each with a marginal product at the cognitive-worker average of \$175,000, implies an addition to output of \$52 trillion per year. To compare with the model, we divide \$52 trillion by the AI sector's output after 2027. By end-2027 the initial 2024 boom's capital has fully adjusted, so $Y_a^{2027} = e^{\hat{\xi}}\, Y_a^{\text{pre-boom}} \approx 2.7 \times \$870\text{B} \approx \$2.34\text{T}$.[12]

[12] Labor is implicit in the reduced-form z_j of the production function (1): writing $Y_j = z_j^{1-\alpha} K_j^{\alpha}$ as

The implied expansion factor on the AI sector is therefore \$52T/\$2.34T ≈ 22, equivalent to roughly three additional booms of size $\hat{\xi}$ ($\log 22 \approx 3.10$). Since the model's window allows for at most two additional booms—the singularity scenario—Amodei's description sits roughly one boom beyond singularity.

Similarly, Musk (2026a) forecasts in May 2026 that "digital intelligence will exceed the sum of all human intelligence" within five years, and that "the economy is probably twice its current size in 5, maybe 6 years"—a doubling of GDP by 2031, against the singularity scenario's 58% cumulative growth by 2030.

More prosaically, strategic considerations could also point to an underestimate. If firms seek to maintain monopoly power, a possibility envisioned by Korinek and Suh (2024), they would invest less, given a boom size. This would predict a greater boom size for the current level of investment, increasing the effect of the scenarios that follow.

Ultimately, the model's predictions should therefore be read as conditional on the revealed-preference estimate of $\hat{\xi}$, which is itself conditional on the assumption that the firms driving the buildout are, in aggregate, not destroying value.

3.5 Output shares and GDP implications

The AI sector's share of total output depends on ξ, r_a, and the long-run boom probability p_*. Using (1):

$$\frac{Y_a}{Y_a + Y_n} = \frac{z_a^{1-\alpha}(K_a^*)^\alpha}{z_a^{1-\alpha}(K_a^*)^\alpha + z_n^{1-\alpha}(K_n^*)^\alpha}. \tag{16}$$

Pre-boom, the AI sector accounts for approximately 2.5–3.5% of output (depending on r_a). After the initial boom plus the window, Table 5 reports output shares for each scenario.

Section 4 traces the GDP implications of these scenarios during and after the transition.

$Y_j = A_j\, L_j^{1-\alpha}\, K_j^\alpha$ gives $z_j = A_j^{1/(1-\alpha)}\, L_j$, so a labor-augmenting boom that multiplies L_a by a given factor multiplies z_a by the same factor. For simplicity, we assume full capital adjustment so that $Y_a \propto z_a$; without this, it would be even harder to justify the predictions within the model.

Table 5: AI Sector Output Share and GDP Growth: Post-Window Snapshot

	AI share (%)	Cum. GDP growth (%)	Transition avg. (%/yr)	Post-window I_a ($B/yr)
Pre-boom	3.1	—	—	155
Moderate	8.0	5.4	0.8	441
Transformative	19.0	19.7	2.8	1,186
Singularity	38.7	58.2	8.3	3,192

Notes: Parameter values from Table 3; $p_* = 0.04$. AI output shares are post-window values, assuming no further booms occur after 2030. Cumulative GDP growth is the one-time level increase from the pre-boom economy. "Transition avg." is the annualized AI contribution to GDP growth over 2024–2030 (seven years). These figures are a snapshot: as booms continue to arrive at rate p_*, the AI sector's output share and GDP contribution grow over time (Figure 3).

4 GDP Growth

We now translate the calibrated scenarios into GDP paths. At each date t, AI output is $Y_{at} = z_{at}^{1-\alpha} K_{at}^{\alpha}$, where z_{at} reflects the cumulative booms realized in that scenario. Non-AI output is held at its pre-boom level Y_{n0}.[13] Cumulative GDP growth from AI is $(Y_{at} + Y_{n0})/\mathrm{GDP}_0 - 1$, where $\mathrm{GDP}_0 = Y_{a0} + Y_{n0}$ is pre-boom GDP.

GDP during the transition. Figure 2 shows cumulative GDP growth from AI for each scenario during the transition (2024–2032). In the moderate scenario, cumulative GDP growth reaches approximately 5% once capital has fully adjusted to the single pre-window boom. In the transformative scenario, the second boom pushes cumulative growth to about 20%. In the singularity scenario, two additional booms compound to produce cumulative GDP growth approaching 58%. Post-window AI investment is approximately $441B/yr (moderate), $1,186B/yr (transformative), and $3,192B/yr (singularity); the last of these

[13]Holding Y_n fixed isolates the AI contribution to growth. The shares reported below are therefore the AI sector's share of an AI-only-growing economy, not its share of calendar-year GDP after incorporating ordinary non-AI growth; over multi-decade horizons the latter would be smaller, most noticeably in the moderate scenario where the AI contribution accrues slowly.

would exceed half of current U.S. gross private fixed investment.

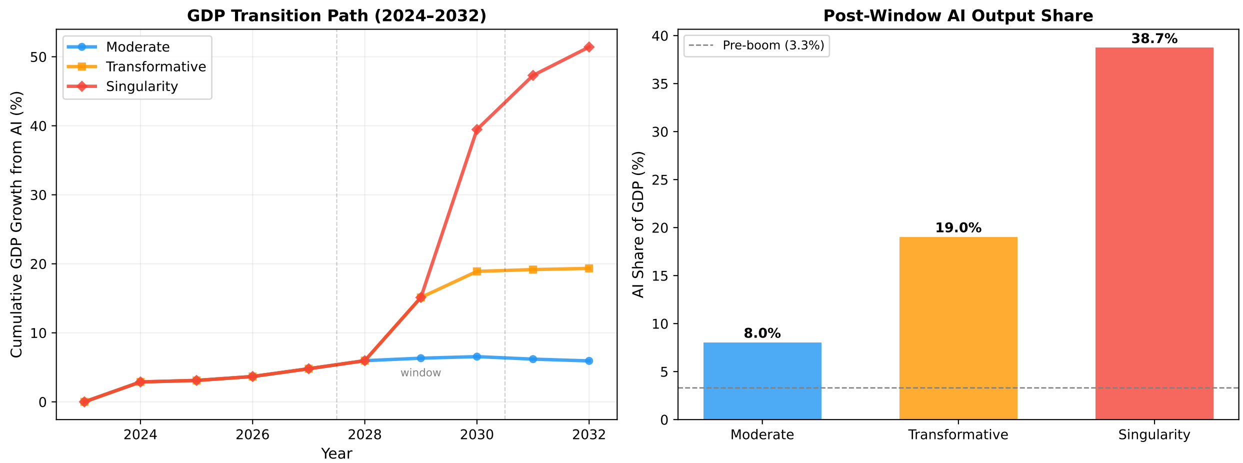

Figure 2: GDP transition paths. Left panel: cumulative GDP growth from AI during the transition (2024–2032), using data-anchored partial adjustment ($\theta_a = 3$). Dashed lines mark the 2028–2030 window. Right panel: post-window AI output share of GDP; the dashed line shows the pre-boom share (3.1%).

Simulated paths (2030–2050). Figure 3 extends the paths to 2050, simulating the effect of ongoing booms. Starting from each scenario's post-window productivity level z, we draw 10,000 independent sample paths in which booms arrive i.i.d. at rate $p_* = 0.04$ per year. Each boom multiplies z by e^{ξ}. Along each path, AI capital adjusts frictionlessly to $K^*_{at} = K^*(z_{at}, p_*)$, the optimal capital stock given current productivity. This frictionless treatment is appropriate for the long-run simulation: adjustment costs govern the speed of the initial transition but are second-order relative to the compounding effect of ongoing booms over a twenty-year horizon.

The compounding effect of post-window booms is first order. In the moderate scenario, expected cumulative GDP growth reaches approximately 30% by 2050, with the AI sector's expected output share rising to roughly 20% (10th–90th percentile ranges: 5–58% for GDP, 8–39% for share). In the singularity scenario, expected cumulative GDP growth reaches 231% by 2050, with the AI sector accounting for 57% of output (10th–90th percentile ranges: 58–

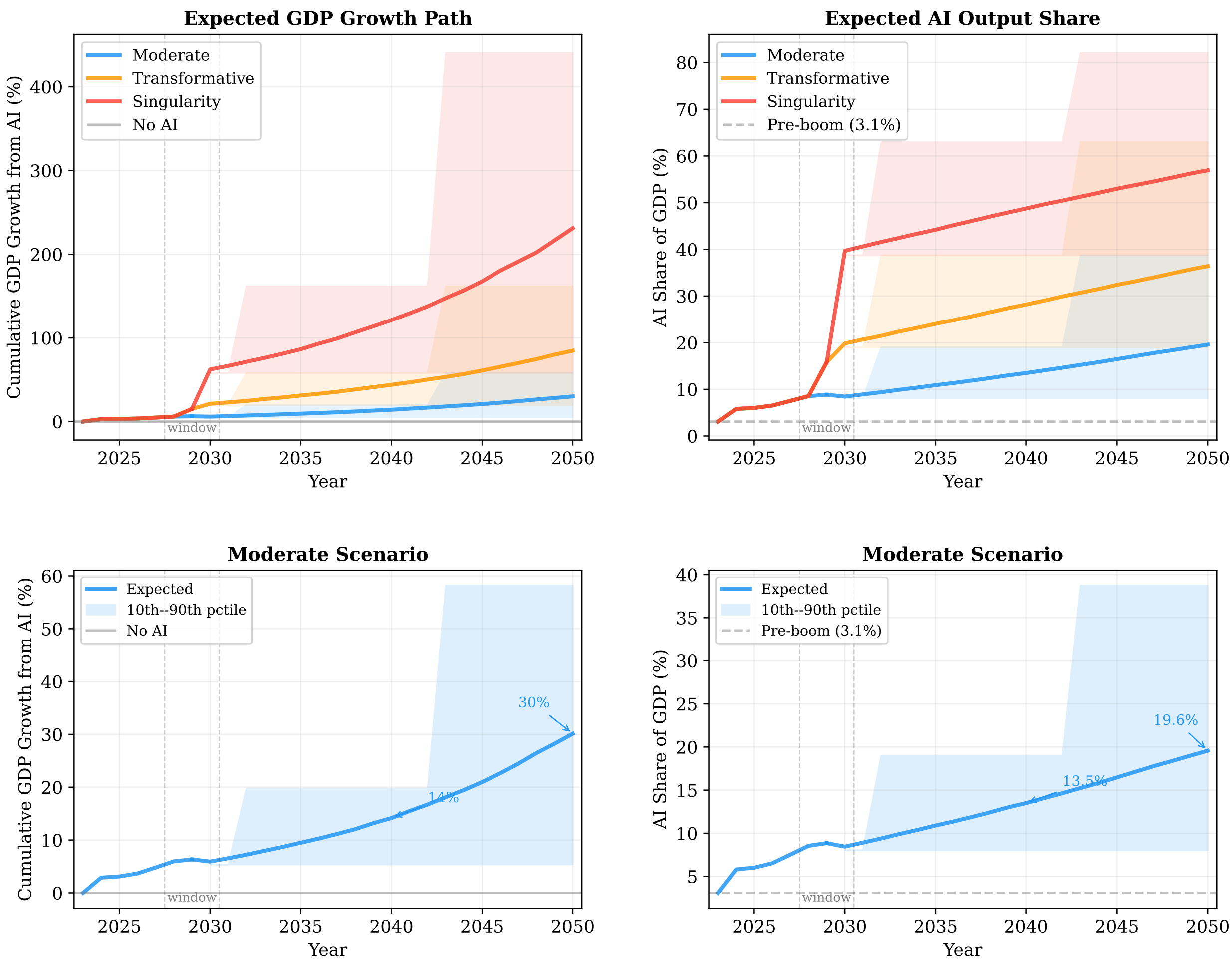

Figure 3: Simulated GDP growth (2024–2050). Top row: all three scenarios. Bottom row: moderate-only snapshot with annotations. Solid lines show the expected (mean) path; shaded bands show the 10th–90th percentile range across 10,000 simulations. The wider band (relative to the more conventional 25th–75th range) accommodates the right skew of the boom distribution: with $p_* = 0.04$ over a twenty-year horizon, the modal outcome is no booms (44% of paths), but rare paths with five or more pull the mean above the median. Gray lines show the no-AI baseline. The transition period (2024–2029) uses adjustment-cost capital paths ($\theta_a = 3$); from 2030 onward, booms arrive i.i.d. at rate $p_* = 0.04$ with frictionless capital adjustment.

440% for GDP, 39–82% for share). The bottom row of the figure provides a detailed view of the moderate scenario, which may be the most policy-relevant benchmark.[14]

Growth rates in the very long run Beyond the simulations of Figure 3, two further considerations apply. First, as the AI sector's output share approaches one, aggregate GDP growth converges to the AI sector's growth rate; the share-weighting that attenuates near-term aggregate effects becomes immaterial. Second, on a balanced growth path with ongoing booms arriving at rate p_*, the firm's first-order condition for capital implies that GDP, productivity, and capital all grow at the same rate.

If the cost of producing AI compute capacity is stationary in the long run, the growth rate of installed capacity coincides with the growth rate of the AI sector itself. This lets us translate Musk's power projections into a model-consistent timeline without taking a stand on the dollar cost per gigawatt. As discussed in Section 3.3, Musk (2026b)'s eventual 1 TW/yr launch flow corresponds to a steady-state installed stock of about 4 TW (at $\delta_a = 0.25$). Applying the model's expected long-run rate $g = p_*(e^{\xi} - 1) = 6.8\%$ as a deterministic compound rate (the back-of-the-envelope benchmark introduced in Section 3.3, which dates the crossing of $E[K_t]$ rather than the typical-path first-passage time), the transformative scenario reaches 4 TW by 2065 (36 years); the moderate scenario (141 GW post-window) reaches 4 TW by 2081 and the singularity scenario (1.0 TW post-window) by 2051. Reaching

[14] For comparison to industry forecasts: SpaceX's May 2026 S-1 (Space Exploration Technologies Corp., 2026) states a quantified AI-related total addressable market of \$26.5 trillion, comprising \$2.4T in AI infrastructure (RAND's 2030 forecast of 235 GW global data-center demand × 70% AI share, marked to a \$3.33/GPU-hour neocloud rental rate), \$760B in consumer subscriptions, \$600B in digital advertising, and \$22.7T in "enterprise applications," which is the Digital Cooperation Organization's estimate of the entire 2026 digital economy. The aggregate is a ceiling rather than a forecast, and mixes gross-output and revenue concepts; it should not be read as an upper bound on *value creation*, which is generically smaller than revenue. Read instead as an upper benchmark on AI-sector annual gross output, the order of magnitude—tens of trillions—is informative: under the calibration, expected $Y_{a,t}$ first crosses \$10T around 2030 in the singularity scenario, 2039 in the transformative scenario, and 2054 in the moderate scenario; \$30T is crossed around 2040, 2055, and 2071, respectively. Even the moderate scenario reaches the lower edge of the S-1's range within the 2050 horizon of Figure 3. The crossing-year computations are produced by `code/y_a_crossings.py`.

4 TW within a decade—as one might read into Musk's mid-2030s phrasing—would require sustained capacity growth above 25% per year, more than three times the model's long-run rate, and is not delivered by any post-window scenario in the calibration.

5 Implications for the equity premium and for interest rates

The open-economy model of Sections 3–4 treats the required return r_a as exogenous. This is the natural formulation for an economy in which global investors supply capital elastically, and it is the most transparent framework for mapping investment decisions into the size of the productivity boom. Moreover, the mapping from risks into interest rates and equity premia is not at present well-understood, as most well-disciplined models fail in various regards (see, e.g., Guo and Wachter (2025)). Nonetheless, the model has strong implications for interest rates and risk premia within the standard framework. We discuss these here.

5.1 Preferences

A representative agent has recursive preferences of the form introduced by Epstein and Zin (1989) with unit elasticity of intertemporal substitution ($\psi = 1$), risk aversion γ, and discount factor β. When $\gamma = 1$, this reduces to time-separable log utility; when $\gamma \neq 1$, the EIS remains unity but the agent is more (or less) risk averse than the log case.

Let W_t denote the cum-dividend value of the consumption claim. At $\psi = 1$, the price–consumption ratio is constant: $P_t/C_t = \beta/(1-\beta)$, so the return on wealth is $R_{W,t+1} = \beta^{-1} C_{t+1}/C_t$. The stochastic discount factor is (see Miller et al., 2025)

$$M_{t+1} = \beta\, \mathbb{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{1-\gamma}\right]^{-1} \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}. \tag{17}$$

The expectation term normalizes the SDF so that $\mathbb{E}_t[M_{t+1}R_{W,t+1}] = 1$. It introduces γ even though the wealth–consumption ratio is constant, and is the source of precautionary savings demand.

Aggregate consumption growth is

$$\Delta c_{t+1} = \mu + \sigma\,\varepsilon_{t+1} + x_{d,t+1} + x_{b,t+1}, \tag{18}$$

where μ and σ are the mean and volatility of normal-times growth, $x_{d,t+1}$ is a rare disaster that equals $\xi_d < 0$ with probability p_d and zero otherwise, and $x_{b,t+1}$ is a rare boom that equals $\xi_b > 0$ with probability p_b and zero otherwise.

Disasters affect the *aggregate* economy. Following Barro (2006) and Wachter (2013), we set $p_d = 0.036$ and $\xi_d = -0.15$; no share-weighting is needed because the disaster is economy-wide. The boom probability p_b is the object that varies across regimes: zero before the AI transition, $p_{\max}$ during the window, and p_* thereafter.

5.2 Assumptions on aggregate cash flows

Equation (18) requires the jump ξ_b in aggregate *consumption*, not output. Mapping the AI-sector productivity boom ξ into a consumption boom involves two considerations that do not arise for disasters, which affect the aggregate economy directly.

First, the boom is sector-specific: when AI productivity jumps by e^{ξ}, non-AI output is unchanged, so the aggregate effect depends on the AI sector's output share s_a.

Second, even if the boom applied to the entire economy, short-run investment dynamics would create a wedge between aggregate output and aggregate consumption. It is for this reason that Gourio (2012) assumes that disasters affect productivity and the capital stock proportionally, and that Miller et al. (2025) assume a shock only to capital and that $\alpha = 1$. These assumptions lead to a production economy that is isomorphic to the endowment economy, greatly simplifying the computation for consumption growth, the stochastic discount factor, and thus investment. Under these assumptions the immediate effect of the boom on consumption would be the same as on productivity and capital.[15]

As discussed in Section 3.2, we do not make either assumption in part because they are not realistic in this setting, and also (relatedly) because we are using the unanticipated shock to productivity to calibrate the investment response. However, a full solution, which would incorporate the gradual growth of the AI sector, as well as complicated short-run consumption dynamics, is difficult, and would not appreciably add to the intuition we are attempting to convey.

We therefore consider two versions of an endowment economy as a means of showing the range of effects the model can produce. We first make the unrealistic but lower-bound assumption that the AI sector remains frozen in time at its post-window level. The effect of post-window productivity booms is then confined to this portion of the economy. We set

$$\xi_b \equiv s_a \cdot (1-\alpha)\,\xi, \tag{19}$$

where s_a is the share of the AI sector under various scenarios described above. This equation

[15]Define total wealth $W_t \equiv (1-\delta)K_t + Y_t$. If C_t equals total consumption, then the law of motion implies that $K_{t+1} = W_t - C_t$ and hence the capital accumulation equation

$$K_{t+2} + C_{t+1} = (1-\delta)K_{t+1} + z_{t+1}^{1-\alpha} K_{t+1}^{\alpha}$$

can be equivalently written, under the special case $\alpha = 1$ (so that $K_{t+1}^{\alpha} = K_{t+1} = W_t - C_t$), as

$$W_{t+1} = (W_t - C_t)\big(1 - \delta + z_{t+1}^{1-\alpha}\big)$$

and thus $1 - \delta + z_{t+1}^{1-\alpha}$ is the return on invested wealth. The assumption of $\psi = 1$ then implies that

$$\frac{C_{t+1}}{C_t} = \frac{W_{t+1}}{W_t} = \frac{1-\delta+z_{t+1}^{1-\alpha}}{1-\delta+z_t^{1-\alpha}} \frac{K_{t+1}}{K_t}.$$

In Miller et al. (2025), for example, z_t is a constant and K_{t+1} is hit by the rare event. Gourio (2012) shows a more general result that states that if capital and productivity fall by the same amount, then all quantities scaled by productivity are independent of the realization of the rare event.

is an approximation even when the AI sector is in fact frozen in time.

The second version takes (19) and sets $s_a = 1$. This also understates the long-run effect on consumption (which will grow at the same rate as output), namely at ξ. However, even this level produces dramatic effects on the riskfree rate and equity premium.

5.3 Asset prices

The risk-free rate and equity premium follow from the Euler equation $\mathbb{E}_t[M_{t+1}R_{t+1}] = 1$. For a riskfree bond with gross return R_f, $R_f^{-1} = \mathbb{E}_t[M_{t+1}]$. The Euler equation implies that the price-dividend ratio equals $1/\beta$, and hence, if we let $R_{W,t+1} = (P_{t+1} + C_{t+1})/P_t$ be the gross return on wealth, $R_{W,t+1} = \beta^{-1} C_{t+1}/C_t$.

Substituting (17) into the Euler equation and defining

$$\Phi_d(u) = (1 - p_d) + p_d\, e^{u\,\xi_d}, \qquad \Phi_b(u) = (1 - p_b) + p_b\, e^{u\,\xi_b}. \tag{20}$$

implies

$$r_f = \rho + \mu + \frac{1 - 2\gamma}{2}\,\sigma^2 + \log\frac{\Phi_d(1-\gamma)}{\Phi_d(-\gamma)} + \log\frac{\Phi_b(1-\gamma)}{\Phi_b(-\gamma)}, \tag{21}$$

where $\rho = -\log\beta$ and $r_f = \log R_f$.

Taking logs of $E_t[R_{W,t+1}]$ and subtracting $r_f = \log R_f$ yields the (log) equity premium:

$$\log\mathbb{E}[R_W] - r_f = \gamma\,\sigma^2 + \log\frac{\Phi_d(1)\,\Phi_d(-\gamma)}{\Phi_d(1-\gamma)} + \log\frac{\Phi_b(1)\,\Phi_b(-\gamma)}{\Phi_b(1-\gamma)}. \tag{22}$$

For small p_d and p_b, this premium can be approximated as:

$$\log\mathbb{E}[R_W] - r_f \approx \gamma\,\sigma^2 + p_d(e^{\xi_d} - 1)(1 - e^{-\gamma\xi_d}) + p_b(e^{\xi_b} - 1)(1 - e^{-\gamma\xi_b}).$$

While we use (22) for all calculations below, the approximation highlights the source of risk:

both disasters and booms raise the equity premium.

Table 6 reports the risk-free rate and equity premium across γ for the pre-AI economy and for the two bounding calibrations, with $p_b = p_* = 0.04$ throughout.

Table 6: Equilibrium rates of return ($\psi = 1$)

Panel A: Post-window output shares

	No AI ($p_b = 0$)		Moderate ($\xi_b = 0.053$)		Transformative ($\xi_b = 0.126$)		Singularity ($\xi_b = 0.256$)	
γ	r_f	ERP	r_f	ERP	r_f	ERP	r_f	ERP
1	3.90	0.12	4.11	0.13	4.37	0.18	4.81	0.37
2	3.77	0.25	3.97	0.27	4.19	0.36	4.48	0.70
3	3.63	0.39	3.81	0.42	4.00	0.55	4.18	1.00
4	3.47	0.55	3.65	0.59	3.80	0.75	3.90	1.28
5	3.29	0.72	3.46	0.77	3.58	0.97	3.63	1.55
6	3.10	0.92	3.26	0.97	3.36	1.19	3.36	1.81

Panel B: Absorbing limit $s_a \to 1$

	No AI ($p_b = 0$)		All scenarios ($\xi_b = 0.660$)	
γ	r_f	ERP	r_f	ERP
1	3.90	0.12	5.85	1.84
2	3.77	0.25	4.79	2.90
3	3.63	0.39	4.16	3.53
4	3.47	0.55	3.75	3.94
5	3.29	0.72	3.44	4.25
6	3.10	0.92	3.18	4.51

Notes: $\rho = 0.02$, $\mu = 0.025$, $\sigma = 0.02$, $p_d = 0.036$, $\xi_d = -0.15$. All post-window columns use $p_b = p_* = 0.04$. Panel A uses boom sizes $\xi_b = s_a \cdot (1-\alpha)\xi$ with the post-window AI output shares from Table 5. Panel B uses the absorbing limit $\xi_b = (1-\alpha)\xi = 0.660$, which obtains when the AI sector's output share converges to one after sufficiently many booms. Both panels understate the full effect because the i.i.d. jump framework does not capture the gradual nature of the consumption response (see text). ERP is the equity risk premium on the claim to aggregate consumption. All entries in percent. $P/C = \beta/(1-\beta) = 49.5$ in all regimes, for all γ.

Several features stand out. First, the range between panels is wide, reflecting the difficulty

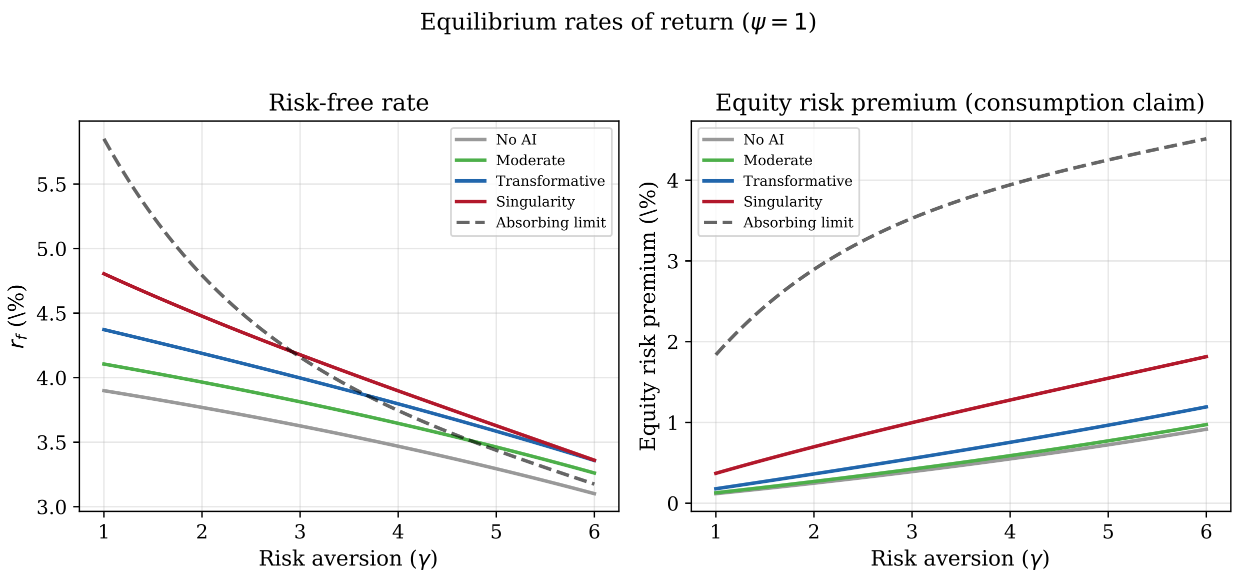

Figure 4: Equilibrium rates of return ($\psi = 1$). Left panel: risk-free rate. Right panel: equity risk premium on the claim to aggregate consumption. Gray line shows the pre-AI economy ($p_b = 0$); colored lines show the three post-window scenarios with $p_b = p_* = 0.04$ and scenario-specific ξ_b (Panel A of Table 6); dashed black line shows the absorbing limit $\xi_b = 0.660$ (Panel B). Other parameters: $\rho = 0.02$, $\mu = 0.025$, $\sigma = 0.02$, $p_d = 0.036$, $\xi_d = -0.15$.

of mapping a sectoral shock to aggregate asset prices without solving the full non-stationary model. At $\gamma = 4$, the equity premium increase ranges from 4 bp (moderate, Panel A) to 339 bp (absorbing limit, Panel B). Panel A treats the AI sector as permanently small; Panel B treats it as if it already dominates the economy. Over time, the economy transitions from Panel A toward Panel B as the AI sector's output share grows; both panels understate the full effect because the gradual consumption catch-up is not captured by the i.i.d. jump framework.

Second, within Panel A the magnitude of the effects depends strongly on the scenario: the moderate scenario barely moves the needle, while the singularity scenario more than doubles the equity premium. Third, the risk-free rate rises in all cases, reflecting higher expected consumption growth: agents who expect to be richer tomorrow save less today. Fourth, the equity premium also rises, because the boom creates a state in which both consumption and equity returns are high. At $\psi = 1$, the SDF is $M_{t+1} \propto \exp(-\gamma\, \Delta c_{t+1})$, so any $\gamma > 0$ penalizes states with high consumption growth. The positive covariance between boom-driven consumption growth and equity returns raises the required return on equity for all levels of risk aversion reported in the table. Fifth, the effect on the risk-free rate is decreasing in γ while the effect on the equity premium is increasing—higher risk aversion strengthens precautionary savings (damping the rise in r_f) and raises the price of risk (amplifying the rise in the equity premium).

5.4 Sovereign default and the observed yield

The general equilibrium risk-free rate in Table 6 is a rate on a default-free claim. Observed government bond yields embed a sovereign default premium. As shown by Miller et al. (2025), higher expected growth lowers the debt-to-GDP ratio on the existing stock of government obligations, reducing the probability of fiscal distress and compressing the default premium.

Higher expected growth from AI would have the same effect: the observed yield may therefore rise by less than the default-free rate—or even fall, if the decline in the default premium exceeds the rise in the risk-free rate.

This channel has quantitative bite for countries with high initial debt-to-GDP ratios. If the AI transition raises trend growth by 2 percentage points, the debt-to-GDP ratio on a 30-year horizon falls by roughly 45% relative to the no-AI baseline (holding primary deficits constant). For the United States, with a debt-to-GDP ratio near 120%, this amounts to a reduction to approximately 65%. The default premium on U.S. Treasuries is small in absolute terms, but at the margin, the growth effect could offset 20–40 basis points of the rise in r_f.

6 Historical Context and Long-Run Implications

The rare-boom framework was developed by Tsai and Wachter (2016) to study cross-sectional asset pricing in a multisector economy with both positive and negative rare events. Several historical episodes provide precedent for the kind of productivity transformation the AI boom represents.

Table 7 reports the largest sustained growth episodes in modern economic history, expressed as cumulative multipliers on GDP per capita.

The AI multipliers in the table are sector-specific productivity gains. To compare with the historical episodes—which are economy-wide—we use the model's aggregate GDP implications. Because the AI sector starts small, the mapping from sector productivity to aggregate GDP depends on the sector's evolving output share. Table 5 reports cumulative aggregate GDP increases of 5% (moderate), 20% (transformative), and 58% (singularity) by the end of the decade, together with long-run AI investment of \$441B, \$1,186B, and \$3,192B, per year respectively.

Table 7: Historical Growth Episodes and the AI Transition

Episode	Period	Years	Multiplier	log(mult.)
Industrial revolutions				
British Ind. Rev. (1st)	1760–1840	80	1.7	0.53
U.S. Railroad era	1850–1910	60	2.8	1.01
2nd Ind. Rev. (U.S.)	1870–1920	50	2.4	0.88
East Asian growth miracles				
Japan postwar	1945–1970	25	10	2.30
South Korea	1960–1990	30	13	2.56
Taiwan	1960–1990	30	10	2.30
Singapore	1965–1990	25	8	2.08
China	1990–2020	30	10	2.30
Technology booms				
U.S. IT boom	1995–2005	10	1.5	0.41
AI window booms only (initial + window booms, $\xi = 0.99$)				
AI moderate (1 boom)	2024–2029	5	2.7	0.99
AI transformative (2 booms)	2024–2029	5	7.2	1.98
AI singularity (3 booms)	2024–2029	5	19.5	2.97
AI sector long-run (post-window expected, $p_* = 0.04$)				
AI sector, 30-year horizon	2030–2060	30	7.2	1.97
AI sector, 50-year horizon	2030–2080	50	26.8	3.29
AI sector, 80-year horizon	2030–2110	80	194	5.27

Notes: Historical multipliers are GDP per capita in purchasing-power-parity terms. Sources: Maddison (2007), Barro (2006), Penn World Table 10.0. British Industrial Revolution figures from Broadberry et al. (2015). U.S. Railroad era GDP per capita from the Maddison Project Database; track mileage from Historical Statistics of the United States, Series Df927. The railroad era overlaps with the Second Industrial Revolution; it is listed separately because the railroad investment cycle is the closest historical analogue to the current AI buildout. The *AI window* multipliers are sector-specific productivity gains from the initial boom plus zero, one, or two additional booms during the 2028–2030 window at $\xi = 0.99$ each (the rounded calibrated value from Section 3.3); they cover the five-year span 2024–2029 only. The *AI long-run* multipliers are the expected (mean) AI-sector productivity gain accruing over T additional years after 2030 under i.i.d. booms at $p_* = 0.04$, computed as Λ^T with $\Lambda = (1 - p_*) + p_* e^{\xi} = 1.068$. Total cumulative AI-sector multiplier from 2024 equals window $\times$ long-run. The AI multipliers apply to the AI sector only, not to aggregate GDP; the share-weighted aggregate effect is smaller during the transition and converges to the AI rate as the AI share approaches one.

The window multipliers in Table 7 cover the five-year period 2024–2029 only. The lower block of the table extends the comparison to longer horizons by reporting the expected post-window AI-sector multiplier $\Lambda^T = (1.068)^T$ under $p_* = 0.04$. Total AI-sector productivity from 2024 onward is the product of the window multiplier and the long-run multiplier; aggregate GDP follows the share-weighted path until the AI share approaches one and aggregate growth converges to $p_*(e^\xi - 1) \approx 6.8\%$ per year.

Two comparisons emerge. Over a five-year window, the AI scenarios (2.7–19.5×) far exceed any historical episode of comparable length; the U.S. IT boom delivered 1.5× over 10 years. Over a 30-year horizon, the AI sector's expected multiplier of 7.2× is comparable to (slightly below) the East Asian miracles, which produced 8×–13× over 25–30 years. Over multigenerational horizons of 50–80 years, the AI sector's expected long-run multipliers (26.8× and 194×) far exceed the Industrial Revolutions (1.7×–2.4×): a sustained 6.8% rate compounds for far longer than any historical productivity trend has been sustained. Because these long-run multipliers depend on the post-window probability p_*, which is less precisely identified than the near-term window booms, the high-horizon numbers should be read as illustrating how the dynamics of the calibrated model extend rather than as forecasts on those timescales.

Within the rare-events framework, speed is captured by the boom probability: the $p_{\max} = 0.50$ window implies far more compressed dynamics than the 4% annual probability that characterizes a boom unfolding over decades. The window delivers, in five years, productivity gains comparable to what the East Asian miracles took 30 years to produce.

These long-run implications invite comparison with the broadest economic transformations in history. The industrial revolutions were general-purpose technology shocks that eventually reshaped every sector of the economy. Because the delineations of sectors are ultimately arbitrary, the growth of the AI sector that we describe in this paper should itself be viewed as incorporating the tendency of general-purpose technologies to spill over, thus following the pattern of electrification—which took roughly 40 years to diffuse from the power sector to manufacturing and services (David and Wright, 2003).

The U.S. railroad era provides the closest historical analogue to the current AI buildout. Railroad track mileage rose from approximately 9,000 miles in 1850 to 164,000 in 1890 and continued growing for another 26 years, peaking at 254,000 miles in 1916.[16] The network did not contract after the construction boom; it kept expanding as successive waves of investment—westward extension, branch lines, interurban rail—filled depreciation and added capacity. The eventual decline to roughly 93,000 miles by 2020 was driven not by a failure of the original technology but by the rise of a competing one: the automobile and the Interstate Highway System.

This trajectory is often cited as evidence that infrastructure booms produce large-scale waste—roughly 63% of peak mileage was eventually abandoned. But the abandonment came decades after the boom, and GDP per capita nearly tripled (2.8) over the railroad era (Table 7). In the language of the model, the productivity gain ξ was large and real; what happened later was a negative shock to the railroad sector's z from a competing technology, not a revelation that the original ξ was illusory. For the AI transition, the railroad precedent suggests that transformative booms routinely involve a capital stock that overshoots its eventual steady state, but after a long time interval and without negating the underlying productivity gain. The relevant question is not whether some fraction of today's GPU capacity will eventually be stranded —depreciation at $\delta_a = 0.25$ already implies that current hardware will be largely replaced within four years—but whether the productivity boom justifies the investment at the time it is made.

[16] Historical Statistics of the United States, Series Df927; Association of American Railroads.

7 Conclusion

The five largest U.S. technology firms are scaling AI capital expenditure at roughly 50% per year, from $155 billion in 2022 to a forecast $755 billion in 2026. This paper embeds this observation in a two-sector model with rare productivity booms to ask what the investment path implies for growth and asset prices.

Calibrating the boom size to match the observed investment ramp implies that each boom raises AI-sector productivity by a factor of roughly 2.7. A two-year window of elevated boom probability generates three scenarios: a moderate outcome (one boom only), a transformative outcome (one further boom), and a singularity scenario (two additional booms). Even in the moderate scenario, AI adds approximately 5 percentage points to cumulative GDP growth over the seven-year transition (Figure 2), on top of the roughly 19% normal-times growth that would otherwise prevail. The transformative and singularity scenarios push the AI contribution to 20% and over 58%, respectively.

In our model, the growth does not arise from economy-wide productivity gains. The non-AI sector continues to grow at its historical rate. What changes is the AI sector's share of the economy, which rises from roughly 3% today to between 8% and 39% depending on the scenario. As this share grows, the AI sector's rapid productivity gains increasingly dominate aggregate GDP growth.

With risk aversion of 3, and an EIS of one, long-term expected annual growth reaches approximately 7%, the risk-free rate increases by approximately half a percentage point, and the equity premium rises by approximately 3 percentage points. The AI sector's earnings load on the boom, and its risk premium rises with risk aversion.

There is a vast range of scenarios, and then substantial long-run uncertainty built on top of those. At its base, however, is a productivity boom that is in investment data, but not yet in productivity data. If the boom fails to materialize, the current buildout will be the largest misallocation of capital in history. On the other hand, those who are certain of that outcome may be making a mistake analogous to that which many forecasters made in 2007, but in reverse. Rare events are hard to imagine until they occur.

References

Acemoglu, D. (2024). The simple macroeconomics of AI. *NBER Working Paper 32487*.

Amodei, D. (2024). Machines of loving grace: How AI could transform the world for the better. Essay, October 2024. Available at https://darioamodei.com/essay/machines-of-loving-grace.

Amodei, D. (2026). We are near the end of the exponential. Interview by Dwarkesh Patel, *Dwarkesh Podcast*, February 13, 2026. Transcript available at https://www.dwarkesh.com/p/dario-amodei-2.

Babina, T., Fedyk, A., He, A., and Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121(3):823–866.

Broadberry, S., Campbell, B. M. S., Klein, A., Overton, M., and van Leeuwen, B. (2015). British economic growth, 1270–1870. *Cambridge University Press*.

Chen, A. Y. (2026). Hedging the singularity. *Working Paper, Federal Reserve Bank of New York*. Available at https://arxiv.org/abs/2604.16997.

Cooley, T. F. and Prescott, E. C. (1995). Economic growth and business cycles. In Cooley, T. F., editor, *Frontiers of Business Cycle Research*, pages 1–38. Princeton University Press.

Dasgupta, P. and Stiglitz, J. (1980). Industrial structure and the nature of innovative activity. *Economic Journal*, 90(358):266–293.

David, P. A. and Wright, G. (2003). General purpose technologies and surges in productivity: Historical reflections on the future of the ICT revolution. In David, P. A. and Thomas, M., editors, *The Economic Future in Historical Perspective*. Oxford University Press.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Gomes, J. F., Grotteria, M., and Wachter, J. A. (2018). Foreseen risks. *Journal of Economic Theory*, 178:181–218.

Gomme, P. and Rupert, P. (2007). Theory, measurement and calibration of macroeconomic models. *Journal of Monetary Economics*, 54(2):460–497.

Gourio, F. (2012). Disaster risk and business cycles. *American Economic Review*, 102(6):2734–2766.

Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., and Brauner, J. (2024). Thousands of AI authors on the future of AI. *arXiv preprint arXiv:2401.02843*.

Guo, H. and Wachter, J. A. (2025). “Superstitious” investors. *Review of Asset Pricing Studies*, 15(1):1–45.

Hayashi, F. (1982). Tobin’s marginal q and average q: A neoclassical interpretation. *Econometrica*, 50(1):213–224.

Hulten, C. R. and Wykoff, F. C. (1981). The measurement of economic depreciation. In Hulten, C. R., editor, *Depreciation, Inflation, and the Taxation of Income from Capital*, pages 81–125. Urban Institute Press, Washington, DC.

Jorgenson, D. W. (1963). Capital theory and investment behavior. *American Economic Review*, 53(2):247–259.

Kapoor, S. and Narayanan, A. (2025). Lifespan of AI chips: The $300 billion question. Center for Information Technology Policy, Princeton University, Blog post, October 2025.

Kording, K. and Marinescu, I. (2025). (artificial) intelligence saturation and the future of work. Working paper, Brookings Institution, Center on Regulation and Markets. Available at https://www.brookings.edu/articles/artificial-intelligence-saturation-and-the-future-of-work/.

Korinek, A. and Suh, D. (2024). Scenarios for the transition to AGI. *NBER Working Paper 32255*.

Maddison, A. (2007). *Contours of the World Economy 1–2030 AD: Essays in Macro-Economic History*. Oxford University Press.

Miller, M., Paron, J. D., and Wachter, J. A. (2025). Sovereign default and the decline in interest rates. *Review of Financial Studies*. Forthcoming.

Musk, E. (2026a). Breaking: Elon Musk's first interview since jury rejected claim against Altman's OpenAI. Interview by Randall Lane, *Forbes*, May 18, 2026. Video, approximately 13:30–15:00.

Musk, E. (2026b). Terawatt of GPUs in space. Interview by Dwarkesh Patel, *Dwarkesh Podcast*, February 5, 2026. Transcript available at https://www.dwarkesh.com/p/elon-musk.

Restrepo, P. (2025). We won't be missed: Work and growth in the AGI world. *NBER Working Paper 35315*.

Space Exploration Technologies Corp. (2026). Form S-1 registration statement. U.S. Securities and Exchange Commission, filed May 20, 2026, accession 0001628280-

26-036936. Available at https://www.sec.gov/Archives/edgar/data/1181412/000162828026036936/spaceexplorationtechnologi.htm.

Trammell, P. and Korinek, A. (2025). Economic growth under transformative AI. *Working Paper*.

Tsai, J. and Wachter, J. A. (2016). Rare booms and disasters in a multisector endowment economy. *Review of Financial Studies*, 29(5):1113–1169.

Van Nieuwerburgh, S. (2026). Financing the AI buildout. *Journal of Economic Perspectives*. Forthcoming; draft of March 20, 2026.

Vanguard Group (2025). AI exuberance: Economic upside, stock market downside. Vanguard Research. Available at https://corporate.vanguard.com/content/corporatesite/us/en/corp/vemo/ai-exuberance-economic-upside-stock-market-downside.html.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance*, 68(3):987–1035.

Wachter, J. A. and Kahana, M. J. (2024). A retrieved-context theory of financial decisions. *The Quarterly Journal of Economics*, 139(2):1095–1147.

Wachter, J. A. and Zhu, Y. (2025). Learning with rare disasters. *Quantitative Economics*, 16(4):1181–1221.

Yotzov, I., Barrero, J. M., Bloom, N., et al. (2026). Firm data on AI. *NBER Working Paper 34836*.

A Further details on data

A.1 Capital expenditure data sources

This subsection documents the EDGAR filings from which the capital-expenditure figures in Table 1 are drawn. The goal is twofold: to specify the exact line item extracted from each firm's cash-flow statement, and to make explicit the convention used to map filings to calendar years for Microsoft (fiscal year ending June 30) and Oracle (fiscal year ending May 31). Table A.1 reports the verbatim figures.

Table A.1: Capital expenditure: cash-flow-statement values from EDGAR ($ millions)

	2022	2023	2024	2025
Calendar-year filers (fiscal year ending December 31)				
Amazon: "Purchases of property and equipment"	63,645	52,729	82,999	131,819
less: proceeds from P&E sales and incentives	5,324	4,596	5,341	3,499
equals: net cash purchases of P&E	58,321	48,133	77,658	128,320
Alphabet: "Purchases of property and equipment"	31,485	32,251	52,535	91,447
Meta: "Purchases of property and equipment"	31,186	27,045	37,256	69,691
plus: principal payments on finance leases	n/a†	1,058	1,969	2,524
equals: capex per management definition	n/a†	28,103	39,225	72,215
Microsoft: fiscal year ending June 30 (FYT = July $T-1$ – June T)				
Microsoft: "Additions to property and equipment"‡	23,886	28,107	44,477	64,551
Oracle: fiscal year ending May 31 (FYT = June $T-1$ – May T)				
Oracle: "Capital expenditures"	4,511	8,695	6,866	21,215

Notes: Figures are taken verbatim from the Consolidated Statements of Cash Flows in each firm's Form 10-K. For the calendar-year filers, the column heading is the calendar year. For Microsoft and Oracle, the column heading is the fiscal year (e.g., the "2025" column for Microsoft reports FY2025 = July 2024–June 2025). † The decomposition of Meta's capex (purchases of P&E plus principal payments on finance leases) was first disclosed in the 2023 10-K; the 2022 finance-lease principal value is not separately reported in the FY2024 10-K cash-flow statement. ‡ Microsoft's reported line includes finance-lease additions paid in cash, unlike the corresponding lines for Amazon, Alphabet, and Meta.

Amazon (CIK 0001018724). The relevant cash-flow line is labeled *Purchases of property and equipment.* A separate line (*Proceeds from property and equipment sales and incentives*) records small recoveries which we report on a second row of Table A.1 for completeness. Equipment acquired under finance leases is reported as a non-cash supplemental disclosure and is excluded from the cash capex line. Sources: FY2022 figures from the FY2024 10-K (accession 0001018724-25-000004, filed February 6, 2025); FY2023–FY2025 figures from the FY2025 10-K (accession 0001018724-26-000004, filed February 6, 2026). The 2026 management forecast of "about $200 billion" was first disclosed in the Q4 2025 earnings release (Form 8-K, February 5, 2026, Exhibit 99.1).

Alphabet (CIK 0001652044). The cash-flow line is labeled *Purchases of property and equipment*. Alphabet does not report finance-lease additions in this line and does not net proceeds from sales against gross purchases at this position. Sources: FY2022 figures from the FY2024 10-K (accession 0001652044-25-000014, filed January 30, 2025); FY2023–FY2025 figures from the FY2025 10-K (accession 0001652044-26-000018, filed February 5, 2026). The 2026 guidance of $175–$185 billion was disclosed in the Q4 2025 earnings release (Form 8-K, February 4, 2026, Exhibit 99.1, attributed to CFO Anat Ashkenazi).

Microsoft (CIK 0000789019). The cash-flow line is labeled *Additions to property and equipment* and—unlike the analogous lines for the four other firms—includes finance-lease additions paid in cash. Microsoft's fiscal year ends June 30, so the figure reported for "FY2022" covers July 2021–June 2022, "FY2023" covers July 2022–June 2023, and so on. The figures in Table 1 of the main text are anchored to the fiscal year ending in each calendar year: the "2022" column reports Microsoft's FY2022 ($23.9B), the "2023" column reports FY2023 ($28.1B), and similarly for later years. Sources: FY2022 figures from the FY2024 10-K (accession 0000950170-24-087843, filed July 30, 2024); FY2023–FY2025 figures from the FY2025 10-K (accession 0000950170-25-100235, filed July 30, 2025). Microsoft does not issue annual capex guidance in any SEC filing; all forward dollar figures appear only on the live earnings call (see Microsoft's standard language in each 8-K Exhibit 99.1: "Microsoft will provide forward-looking guidance in connection with this quarterly earnings announcement on its earnings conference call and webcast"). The 2026 range in Table 1 reflects sell-side compilations of call commentary, not an SEC-filed quantitative guide.

Meta (CIK 0001326801). The cash-flow line is labeled *Purchases of property and equipment*. Meta separately reports *Principal payments on finance leases* in the financing-activities section, and its management commentary defines "capital expenditures" to be the sum of the two—i.e., investment-activity P&E purchases plus financing-activity lease principal. We report both the cash-flow-statement value and the management-definition total in Table A.1. Sources: FY2022 figures from the FY2024 10-K (accession 0001326801-25-000017, filed January 29, 2025); FY2023–FY2025 figures from the FY2025 10-K (accession 0001628280-26-003942, filed January 29, 2026). The 2026 guidance of \$125–\$145 billion (raised from a January 2026 initial range of \$115–\$135 billion) was disclosed in the Q1 2026 earnings release (Form 8-K, April 29, 2026, Exhibit 99.1) and again in the Q1 2026 10-Q MD&A.

Oracle (CIK 0001341439). The cash-flow line is labeled *Capital expenditures*. Oracle's fiscal year ends May 31, so the figure reported for "FY2022" covers June 2021–May 2022, and so on. Like Microsoft, Oracle is anchored in Table 1 of the main text to the fiscal year ending in each calendar year: the "2022" column reports Oracle's FY2022 (\$4.5B), the "2023" column reports FY2023 (\$8.7B), and similarly for later years. Sources: FY2022 figures from the FY2024 10-K (accession 0000950170-24-075605, filed June 20, 2024); FY2023–FY2025 figures from the FY2025 10-K (accession 0000950170-25-087926, filed June 20, 2025). Oracle disclosed an FY2026 (June 2025–May 2026) capex target of approximately \$50 billion at the Q1 FY2026 earnings release (Form 8-K, September 9, 2025, Exhibit 99.1) and reaffirmed it on the Q3 FY2026 call. Two reporting peculiarities affect interpretation. First, Oracle's reported capex excludes infrastructure acquired through customer-prepayment or customer-supplied-GPU arrangements; the Q3 FY2026 release states that "most of the equipment needed is either funded upfront via customer prepayments so Oracle can purchase the GPUs, or the customer buys the GPUs and supplies them to Oracle." Section 2 of the main text treats the reported figure as the relevant magnitude for the calibration; the true infrastructure footprint flowing through Oracle data centers is likely somewhat larger. Second, Oracle delivers data-center capacity to Amazon, Alphabet, and Microsoft under its MultiCloud partnership, raising a potential double-count concern with the other four firms at the data-center-shell layer (though not at the server/GPU layer, which sits in each hyperscaler's separate capex).

Non-GAAP nature of the capex concept. The Consolidated Statement of Cash Flows itself is mandated by GAAP (ASC 230), and a line near "purchases of property and equipment" is required in the investing-activities section. But "capital expenditure" is a managerial concept that firms use in MD&A and earnings guidance, not a GAAP-defined line item. Firms accordingly differ in the label they use, in whether proceeds from sales are netted against gross purchases, in whether finance-lease additions (typically non-cash and disclosed elsewhere) are included, and in whether and how forward guidance is delivered. Table 1 of the main text reflects these conventions firm by firm: Microsoft's reported line includes cash paid for finance leases, Meta's management-disclosed capex adds finance-lease principal from the financing-activities section to investing-activities P&E purchases, and Oracle reports the gross figure under the label "Capital expenditures." These conventions are documented above and in Table A.1.

A.2 Capital-expenditure forecasts for 2026 and bottom-up estimates for 2027

This appendix documents the firm-by-firm source materials underlying the "2026 (forecast)" and "2027 (est.)" columns of Table 1. For 2026, every figure except Microsoft's traces to an explicit management disclosure in a Form 8-K (Exhibit 99.1), 10-Q, or 10-K filed with the SEC. Microsoft alone among the five firms does not issue annual capital-expenditure guidance in any filed document; its 2026 midpoint is a sell-side compilation from filed quarterly actuals plus the leases-not-yet-commenced disclosure in the commitments footnote. For 2027, no firm has issued formal annual guidance. Each 2027 estimate is anchored to multi-year commitments already in filings — leases not yet commenced, contracted revenue with capex prerequisites, and signed multi-year customer agreements — together with firm-specific forward-looking statements about contracted demand and capacity in 10-Qs and 8-Ks.

2026 management forecasts (five-firm aggregate: $755B)

Amazon: $200B (point). The Q4 2025 earnings release (Form 8-K, February 5, 2026), Exhibit 99.1, states verbatim:

> *"With such strong demand for our existing offerings and seminal opportunities like AI, chips, robotics, and low earth orbit satellites, we expect to invest about $200 billion in capital expenditures across Amazon in 2026, and anticipate strong long-term return on invested capital."*

attributed to Andy Jassy, President and CEO. This is a direct full-company point forecast at the same accounting scope as Table 1 (Amazon's cash-flow "purchases of property and equipment" line). The figure was reaffirmed in commentary on the Q1 2026 earnings call (April 29, 2026); Q1 2026 gross capex of $44.2B annualizes consistently with a $200B full-year run-rate given Amazon's historical Q4 back-loading, and trailing-twelve-month net purchases of property and equipment through Q1 2026 reached \$147.3B (+67% year-over-year), well on the \$200B trajectory.

Alphabet: \$175–\$185B range; \$180B midpoint. The Q4 2025 earnings release (Form 8-K, February 4, 2026), Exhibit 99.1, states, attributed to CFO Anat Ashkenazi:

> *"Our 2026 CapEx investments are anticipated to be in the range of \$175 to \$185 billion."*

Q1 2026 actual gross capex was \$35.7B (a +107% year-over-year increase), representing 19.8% of the \$180B midpoint — consistent with the typical back-loading of data-center construction commissioning costs into late-year quarters. The Q1 2026 10-Q MD&A reaffirms the forward stance: "In 2026, we expect to significantly increase, relative to 2025, our investment in our technical infrastructure, including servers and network equipment and data centers."

Microsoft: \$155–\$195B range; \$175B midpoint (sell-side compilation). Microsoft is the only firm in Table 1 that does not issue formal annual capital-expenditure guidance in any filed document. Each Microsoft 8-K Exhibit 99.1 contains the same boilerplate:

> *"Microsoft will provide forward-looking guidance in connection with this quarterly earnings announcement on its earnings conference call and webcast."*

Forward dollar figures are therefore delivered verbally on quarterly calls and never appear in a subsequently filed document.

To arrive at \$175B, we consider the filed fiscal-year-2026 quarterly trajectory — \$19.4B (1Q FY2026), \$29.9B (2Q FY2026), \$30.9B (3Q FY2026) reported in successive 10-Qs, summing to \$80.146B for the nine months ended March 31, 2026. This is roughly a 70%

increase relative to the comparable nine months in FY2025. We extrapolate these filed numbers to $35–$40B for Q4 FY2026, and $90–$125B to H1-FY2027.

Calendar 2026 spans the second half of FY2026 (January–June 2026, “H2 FY2026”) plus the first half of FY2027 (July–December 2026, “H1 FY2027”). The range is $155–$195B with $175B at the midpoint.

Microsoft’s reported cash-capex line already includes finance-lease additions paid in cash (Table A.1), so the $103.9B nine-month increase in leases-not-yet-commenced ($92.7B at June 30, 2025 to $196.6B at March 31, 2026) is not additive to the trajectory; it is a forward indicator that the commencement rate already implicit in the quarterly trajectory will continue or accelerate over coming quarters. The upper end of the range reflects continued acceleration; the lower end reflects moderate continuation.

Meta: $125–$145B range; $135B midpoint. The Q1 2026 earnings release (Form 8-K, April 29, 2026), Exhibit 99.1, states verbatim:

> *“We anticipate 2026 capital expenditures, including principal payments on finance leases, to be in the range of $125–145 billion, increased from our prior range of $115–135 billion.”*

Meta’s disclosed figure is the “capex plus finance-lease principal” composite documented in Table A.1 and is therefore comparable, after that composition, with the corresponding Microsoft line. Q1 2026 actuals were $19.84B ($19.00B in purchases of property and equipment plus $0.84B in finance-lease principal), a +45% year-over-year increase versus Q1 2025; the $125–145B full-year range therefore requires Q2–Q4 averaging approximately $38B per quarter, consistent with the Q4-heavy capex profile that has characterized Meta’s recent build cycle. Meta funded the buildout in part through $29.9B of net proceeds from fixed-rate senior unsecured notes issued in November 2025 (FY2025 10-K).

Oracle: \$65B. Oracle's fiscal year ends May 31, so calendar 2026 is approximated by the third and fourth quarters of fiscal 2026 plus the first and second quarters of fiscal 2027 (December 2025 through November 2026, one month offset from calendar 2026 proper). The \$65B figure in Table 1 blends the disclosed FY2026 (June 2025 to May 2026) management-guided capex of \$50B with the beginning of the FY2027 ramp. The Q3 FY2026 earnings release reaffirms the guidance: "For fiscal year 2026, we expect revenue of \$67 billion and capital expenditures of \$50 billion." Nine-month FY2026 actual capex was \$39.17B (+223% YoY versus the comparable nine months of FY2025), with Q3 FY2026 alone at \$18.63B (+305% YoY), consistent with the full-year \$50B trajectory. Oracle's \$50B financing program announced in February 2026 (senior notes, mandatory-convertible preferred, and an at-the-market equity facility, disclosed in the corresponding 8-K exhibits) was sized specifically to fund the FY2027 buildout; \$30B was raised within days of announcement.

Oracle's reported capex line excludes customer-funded GPU spend under its large-scale AI contracts. The Q3 FY2026 release states verbatim:

> *"Most of the increase in RPO in Q3 related to large scale AI contracts where Oracle does not expect to have to raise any incremental funds to support these contracts as most of the equipment needed is either funded upfront via customer prepayments so Oracle can purchase the GPUs, or the customer buys the GPUs and supplies them to Oracle."*

The reported capex line therefore understates the gross compute footprint deployed through Oracle data centers, by an estimated factor of 1.3–2.0× at peak ramp. Table 1 uses the reported figure; the gross-compute implications are discussed in Appendix A.3.

2027 bottom-up estimates (five-firm aggregate: $1,090B)

No firm has issued formal full-year 2027 capital-expenditure guidance in a filed document. The 2027 (est.) column of Table 1 reports bottom-up estimates that, for each firm, are bounded below by mechanical commitments visible in filings and bounded above by recent quarterly run-rates.

Amazon: $285B. Amazon's cash capex grew $53B (FY2023) → $83B (FY2024, +57%) → $132B (FY2025, +59%) → $200B (FY2026 guided, +52%). The $285B 2027 estimate (range $280–$310B) is supported by four convergent forward anchors in the Q1 2026 earnings release (Form 8-K, April 29, 2026), Exhibit 99.1:

- An OpenAI commitment to consume approximately 2 GW of Trainium capacity through AWS infrastructure "[which] begins ramping in 2027."
- An Anthropic commitment "to secure up to five gigawatts (GW) of current and future generations of Amazon's Trainium chips to train and power their advanced AI models."
- Trainium4 production launch in 2027, "with 6 times the FP4 compute performance, 4 times more memory bandwidth, and 2 times more high memory bandwidth capacity than Trainium3."
- Deployment of more than one million NVIDIA GPUs starting in 2026 (Q4 2025 release), with revenue recognition extending through 2027 and beyond.

At the integrated-campus build cost of $10–15B per gigawatt, the 2 GW contracted OpenAI Trainium build alone implies a $20–30B envelope, with the Anthropic 5 GW ramp, Trainium4 production capacity, and NVIDIA GPU deployment program layered on top. The $285B 2027 figure is therefore a continuation of the multi-year trajectory, not a break from it.

Alphabet: $255B. Alphabet's cash capex grew $32B (FY2023) → $53B (FY2024, +66%) → $91B (FY2025, +72%) → $180B (FY2026 midpoint, +97%). Q1 2026 actual capex of $35.7B (+107% year-over-year) is 19.8% of the $180B 2026 midpoint, consistent with the typical back-loading of data-center commissioning into late-year quarters. Alphabet has not issued explicit 2027 guidance; the Q1 2026 10-Q MD&A states that 2026 investment will "significantly increase" relative to 2025. The $255B 2027 point estimate corresponds to roughly +42% growth over the 2026 midpoint and is anchored by four multi-year commitments disclosed in filings:

- Multi-gigawatt TPU hardware supply agreements with customers, disclosed in the Q1 2026 10-Q MD&A: "Google Cloud has entered into a limited number of agreements to supply multiple gigawatts of TPU hardware to customers who require or provide on-premises infrastructure for specialized, high-scale workloads.... We expect to begin recognizing revenues from these agreements later in 2026, with the significant majority to be recognized in 2027." The host disclosure is the $460B Google Cloud backlog cited in the Q1 2026 8-K Exhibit 99.1.
- Future lease payments on leases not yet commenced, primarily for data centers, totaling $75.6B at March 31, 2026 (Q1 2026 10-Q, Note 4), up $17B in a single quarter from $58.5B at December 31, 2025. These leases commence between 2026 and 2031 with non-cancelable terms primarily between one and 25 years, and convert to "additions to property and equipment" on commencement.
- A $33.3B forward-backstop framework, disclosed for the first time in the Q1 2026 10-Q, to "support the build-out of data center and energy supply infrastructure," against which a $15.3B specific-provider backstop was finalized in April 2026. Combined with $37.4B of pre-existing committed backstops (financial guarantees plus credit derivatives), total off-balance-sheet support for third-party infrastructure exceeds $70B at

quarter-end.

- A \$9.9B power-purchase agreement covering 2027–2047, executed January 2026, expected to be accounted for as a lease (FY2025 10-K).

Total purchase commitments at March 31, 2026 stood at \$332.4B, "primarily for technical infrastructure" (Q1 2026 10-Q). Cash-flow capex of \$255B in 2027 represents the visible-cash portion of a substantially larger committed infrastructure program.

Microsoft: \$250B. Filed nine-month FY2026 additions to property and equipment totaled \$80.146B, a +69% increase versus the comparable nine months of FY2025. The most-recent reported quarter, Q3 FY2026 (January–March 2026), was \$30.876B alone, a +84% year-over-year increase indicating that quarterly capex remained accelerating at the end of the available data. Calendar-2026 capex of \$175B (midpoint of \$155–\$195B) extrapolates these quarterly actuals through the remaining fiscal-year quarters and rebases to a calendar basis. The \$250B 2027 figure is bounded below by five mechanical commitments visible in the filings:

- Leases primarily for data centers that had not yet commenced stepped from \$92.7B (June 30, 2025) to \$155.1B (December 31, 2025) to \$196.6B (March 31, 2026), with commencement spanning fiscal years 2026–2031. The \$103.9B of fresh data-center lease commitments added over nine months convert to "additions to property and equipment" on commencement and support the multi-year capex trajectory through FY2031 even with no further commitments.
- Servers, network equipment, and software on the balance sheet grew from \$132.8B (June 30, 2025) to \$190.9B (March 31, 2026) — a +44% gross-stock increase in nine months reflecting the AI-compute layer.

- Accounts payable for purchases of property and equipment rose from \$6.9B (June 30, 2025) to \$22.6B (March 31, 2026), a \$15.7B in-flight invoice pipeline implying near-term cash conversion.

- Commercial remaining performance obligation reached \$627B at March 31, 2026, up +99% year-over-year — equivalent to roughly three years of Microsoft Cloud revenue contracted forward. AI-business annual revenue run-rate reached \$37B, up +123% YoY.

- Microsoft's strategic partnership with OpenAI gives Microsoft "a right of first refusal on OpenAI's new capacity needs" (FY2025 10-K), tying Azure capex to the largest single source of AI compute demand.

Both the lease-commencement pipeline and the asset-stock trajectory imply meaningful step-up in calendar-2027 cash capex relative to calendar 2026, with \$250B sitting in the conservative half of plausible outcomes.

Meta: \$195B. Meta's reported capex (including principal payments on finance leases) grew \$28B (FY2023) → \$39B (FY2024, +40%) → \$72B (FY2025, +84%) → \$135B (FY2026 midpoint, +87%). The \$195B 2027 estimate is supported by four firm-specific indicators:

- The Q1 2026 earnings release attributes part of the \$10B upward revision to the 2026 capex guidance range to "higher component pricing this year and, to a lesser extent, additional data center costs to support *future year capacity*" (Form 8-K, April 29, 2026, Exhibit 99.1). "Future year capacity" is the in-filing 2027 anchor: a portion of the 2026 raise is explicitly pre-funding 2027 capacity.

- The Q1 2026 10-Q MD&A states: "We have made significant investments in AI initiatives . . . and expect to continue to increase these investments. . . . [W]e have significantly increased our infrastructure investments in connection with our AI initiatives . . . and expect our investments to continue to increase."

- Q1 2026 actual capex of \$19.84B (+45% YoY versus Q1 2025) confirms the underlying growth trajectory. Meta back-loads heavily to Q4, so the implied Q4 2026 quarterly exit-rate of approximately \$40–\$45B per quarter (required to hit the \$135B full-year midpoint given Q1's \$19.84B start) annualizes flat to roughly \$160–\$180B for 2027 before any further growth from that exit rate.
- In November 2025, Meta issued \$29.9B of net proceeds from fixed-rate senior unsecured notes (FY2025 10-K), funding the multi-year AI buildout.

The \$195B estimate is therefore supported by Meta's own filed trajectory and the exit-rate arithmetic. A multi-year, multi-billion CoreWeave capacity contract was widely reported in March 2026; under ASC 606 service-contract accounting it does not enter Meta's capex line and therefore does not support the \$195B figure directly. It does indicate Meta's contracted demand for 2027 compute beyond what Meta builds on its own balance sheet, and the corresponding capex is reported separately in Appendix A.3.

Oracle: \$105B. Oracle's 2027 figure is built from a different bridge. The Q3 FY2026 earnings release (March 10, 2026, Form 8-K) raised FY2027 total revenue guidance to \$90B, an increase from approximately \$67B in FY2026:

> *"For fiscal year 2027, we are raising total revenue guidance to \$90 billion."*

The Q1 FY2026 release (Form 8-K, September 9, 2025) disclosed a five-year Oracle Cloud Infrastructure revenue trajectory — FY2026 \$18B, FY2027 \$32B, FY2028 \$73B, FY2029 \$114B, FY2030 \$144B — requiring a major capital-expenditure step-up beginning in FY2027. Holding the FY2026 capex-to-revenue ratio (\$50B/\$67B = 0.75) constant against the \$90B

FY2027 revenue guidance gives FY2027 capex of approximately \$68B. FY2028 OCI revenue of \$73B (from \$32B in FY2027, a 2.3× step) implies a comparable capital-intensity step into FY2028; applying the same revenue-to-capex ratio to projected FY2028 total revenue of roughly \$115–\$125B (FY2028 OCI plus continued growth in non-OCI lines) gives FY2028 capex of approximately \$85–\$95B. Calendar 2027 (December 2026–November 2027) spans the second half of FY2027 and the first half of FY2028; averaging the two semesters gives approximately \$105B.

The trajectory of recent actuals supports the ramp: nine-month FY2026 capex was \$39.17B (+223% YoY versus the comparable nine months of FY2025) and Q3 FY2026 alone was \$18.63B (+305% YoY). Remaining performance obligation reached \$553B at Q3 FY2026, up +325% year-over-year, with Q3 alone adding \$29B — substantially representing large-scale AI contracts. The \$105B figure understates the gross compute deployed through Oracle data centers, for the customer-funding reasons described in the 2026 section above.

Reconciliation. The five-firm 2026 forecasts sum to \$200 + \$180 + \$175 + \$135 + \$65 = \$755B, matching the "2026 (forecast)" column of Table 1. The five-firm 2027 bottom-up estimates sum to \$285 + \$255 + \$250 + \$195 + \$105 = \$1,090B, the value reported in the "2027 (est.)" column of Table 1. Both totals reflect the paper's five-firm scope. Including the sixth hyperscaler (xAI, the AI segment of SpaceX following the February 2026 acquisition), at an estimated \$40B in 2026 (anchored to the \$7.7B Q1 2026 AI-segment capex disclosed in SpaceX's S-1) and \$60B in 2027, raises the totals to \$795B and \$1,150B, respectively; the xAI evidence is detailed in Appendix A.3. Bottom-up reconstructions that incorporate US-domiciled neocloud operators (CoreWeave, Crusoe, IREN, Lambda, and the partial-overlap names Applied Digital and Hut 8) place 2027 US AI infrastructure capex at approximately \$1,220B, with the neocloud increment fully incremental to the five-firm total under ASC 606 service-contract accounting. The \$1,090B five-firm figure in Table 1 is therefore a conservative measure of the relevant total.

A.3 Capital expenditure outside the five-firm scope

The five firms in Table 1 account for the bulk of US AI infrastructure capex through 2025, but they are not the entire universe. This subsection documents two additional categories of spending that the main-text table omits: a sixth hyperscaler (xAI) and four "neocloud" service providers whose revenue is recognized under ASC 606 service-contract accounting (CoreWeave, Crusoe, IREN, Lambda). A third category—"Tier 2" partial-overlap firms (Applied Digital, Hut 8) whose revenue includes ASC 842 operating-lease arrangements with the hyperscalers—is deferred to the Internet Appendix, where the lease-double-count question can be treated in detail without cluttering the main-text reconciliation.

Why the Tier 1 neoclouds are fully incremental to the five-firm aggregate. CoreWeave's FY2025 10-K (Note 2, Revenue Recognition) states verbatim:

> "[The Company's] arrangements do not meet the definition of a lease under ASC 842 and are accounted for as service contracts under ASC 606."

In consequence, a hyperscaler purchasing AI compute from CoreWeave (or any Tier 1 peer with the same accounting structure) records the payment as cost of revenue, not as a finance-lease right-of-use asset. CoreWeave's underlying property and equipment remains on CoreWeave's balance sheet. Adding CoreWeave's reported capex to the five-firm aggregate therefore introduces no double-count.[17] xAI is in a different position: it owns and operates its own data centers, is not a service provider to the hyperscalers, and is treated as

[17] An alternative model—under which the hyperscaler would consolidate the leased compute as its own asset—would require the lease to convey control of an identified asset under ASC 842. The 10-K language above rules out this classification. Crusoe, IREN, and Lambda follow the same service-contract model. Among the public filers in this group, IREN explicitly discloses ASC 606 treatment for its Microsoft and NVIDIA multi-year contracts.

a sixth hyperscaler that is simply not publicly traded. In all five cases below the spending is incremental to Table 1.

Historical capex by firm, 2022–2025. Table A.2 reports annual capital expenditure for the five additional firms. Two (CoreWeave and IREN) file with the SEC directly and have audited cash-flow disclosures. xAI, originally a private standalone firm, was acquired by SpaceX in February 2026 and is reported in SpaceX's May 2026 S-1 as the AI segment of the combined entity (Space Exploration Technologies Corp., 2026); the 2023–2025 figures shown for xAI are the audited AI-segment capital-expenditure line from the Consolidated Statements of Cash Flows. Crusoe and Lambda remain private with no audited filings, and defensibility for those two rows is correspondingly weaker. Before 2024, all five firms were either pre-AI in business model (IREN was a bitcoin miner; Crusoe was a flare-gas/crypto firm), pre-founding (xAI was founded in mid-2023), or very small (CoreWeave's 2023 capex was \$3.4B economic).

Comparison to the five-firm aggregate. The 2022–2023 contribution from this expanded universe is negligible relative to the \$155B and \$150B five-firm aggregates. In 2024, the \$17B expanded-universe contribution is approximately 7.5% of the five-firm \$226B; in 2025, the \$41B contribution is approximately 11% of the five-firm \$381B. The expanded universe grows substantially in 2026 (forecast contribution approximately \$95B, dominated by xAI and CoreWeave) and is expected to continue growing into 2027. For the calibration in Section 3.3, omitting the expanded universe leaves the pre-boom AI capital stock K_{a0} essentially unchanged (the 2022 contribution is negligible) and biases the post-boom investment $\bar{I}$ downward. Under the calibration formula $e^{\hat{\xi}} = (\bar{I}/K_0 + 1 - \delta_a)/(1 + 3\delta_a)$, $\hat{\xi}$ is increasing in $\bar{I}$, so the five-firm calibration yields a downward-biased $\hat{\xi}$ relative to one that incorporates the expanded universe.

Table A.2: Capital expenditure of additional firms outside the five-firm scope ($ billions)

	2022	2023	2024	2025	Source basis
CoreWeave†	—	3.4	9.6	21.5	10-K (cash + OEM-financed)
Crusoe	—	—	∼1.0	∼4.0	private; press anchors
IREN‡	—	—	0.5	1.6	10-K (fiscal year ending June 30)
Lambda	—	—	<0.5	∼1.0	private; press anchors
xAI§§	—	0.5	5.6	12.7	SpaceX S-1 (AI segment, audited)
Subtotal	**0**	**∼4**	**∼17**	**∼41**	

Notes: "—" indicates that the firm was either pre-AI in business model, pre-founding, or that AI-attributable capital expenditure was negligible. †CoreWeave's "economic" capex equals cash purchases of property and equipment plus liabilities related to property-and-equipment additions financed by original equipment manufacturers (chiefly NVIDIA). The OEM-financed component grew 12.5× in 2025 as suppliers provided extended-term financing on GPU deliveries (CoreWeave 10-K for fiscal year 2025, filed March 2, 2026, accession 0001769628-26-000010, Consolidated Statements of Cash Flows). ‡IREN's fiscal year ends June 30; the values reported are calendar-year approximations stitched from quarterly cash-flow disclosures. §§xAI was acquired by SpaceX in February 2026 and is reported as the AI segment of the combined entity in SpaceX's S-1, filed May 20, 2026 (Space Exploration Technologies Corp., 2026); the 2023–2025 figures are the audited AI-segment capital-expenditure line in the Consolidated Statements of Cash Flows. Crusoe and Lambda remain private with no audited filings; the values shown for these two firms are best estimates from primary press sources and venture-capital databases, with substantially weaker defensibility than the public-filer values.

Sources by firm. For **CoreWeave**, FY2023–FY2025 capex is from the FY2025 10-K cash-flow statement (filed March 2, 2026, accession 0001769628-26-000010), with the OEM-financed layer from "Liabilities related to property and equipment additions, including OEM-financed additions" in the same statement. For **IREN**, FY2024 and FY2025 fiscal-year capex (ending June 30) is from the FY2025 10-K (filed August 28, 2025); calendar 2024 and 2025 are stitched from Q3 FY26 quarterly disclosures (10-Q filed May 8, 2026). For **xAI**, the 2023–2025 figures are the AI-segment line of the Consolidated Statements of Cash Flows in SpaceX's S-1, filed May 20, 2026, accession 0001628280-26-036936 (Space Exploration Technologies Corp., 2026). The S-1 reports the AI segment on a pro-forma basis as if the February 2026 xAI acquisition had occurred at the beginning of the historical-period presented, so the 2023–2025 figures correspond to the standalone xAI entity (which also encompasses the X platform following the March 2025 X/xAI merger described in the S-1). The S-1 also reports Q1 2026 AI-segment capex of $7,723 million, which is the empirical anchor for the 2026 forecast cited below.[18] **Crusoe** and **Lambda** are private and have not filed audited financial statements. Crusoe's 2024 and 2025 estimates are press-anchored to its revenue trajectory ($276 million in 2024, approximately $998 million in 2025) and its disclosed share of the Stargate Phase 2 joint venture ($15 billion total commitment, with $11.6 billion of debt and equity raised through 2026). Lambda's estimates rely on its November 2025 Series E disclosure ($1.5 billion at over $2.3 billion cumulative equity), with modest pre-2025 spend reflecting an earlier stage of capacity build-out.

Deferred: Tier 2 partial-overlap firms. Two additional public filers—Applied Digital and Hut 8—are data-center owners with operating-lease customer relationships that fall under ASC 842 rather than ASC 606. When a hyperscaler tenant in such an arrangement

[18]The same filing reports Q1 2026 capex of $1,052 million for SpaceX's Space segment (launch and Starship development) and $1,332 million for the Connectivity segment (Starlink). Neither figure is included in our AI infrastructure aggregate.

classifies its lease as a finance lease, the underlying property is capitalized as a right-of-use asset on the hyperscaler's balance sheet and therefore appears in the hyperscaler's "additions to property and equipment" line. In that case, simply adding the lessor's capex to the hyperscaler aggregate would double-count. Applied Digital's FY2026 third-quarter 10-Q discloses that the majority of its HPC hosting revenue is recognized under ASC 842 as operating leases; Hut 8's Beacon Point contract is structured as a 15-year triple-net lease that would typically receive finance-lease treatment by the tenant. Resolving the double-count requires firm-by-firm review of each lease's classification and the tenant's identity, which is more detail than the Table 1 reconciliation requires. The Internet Appendix carries this analysis with a portfolio-level haircut for the partial overlap, and reports the corresponding adjusted aggregates and sensitivity calculations.

A.4 Depreciation rates

This subsection discusses the calibration of the sector-specific depreciation rates $\delta_n = 0.07$ and $\delta_a = 0.25$.

A.4.1 Non-AI capital: $\delta_n = 0.07$

The depreciation rate for non-AI capital follows standard practice in macroeconomic calibration. The Bureau of Economic Analysis (BEA) constructs estimates of the net capital stock and consumption of fixed capital (depreciation) for the U.S. economy using geometric depreciation rates derived from the studies of used-asset prices by Hulten and Wykoff (1981). The ratio of aggregate depreciation to the aggregate net stock of private fixed assets in the BEA Fixed Asset Tables has averaged approximately 6–8% per year over the past two decades, with the precise figure depending on the asset mix. The value $\delta_n = 0.07$ is the midpoint of this range and is consistent with values used in the real business cycle literature (e.g., Cooley and Prescott, 1995, Gomme and Rupert, 2007).

A.4.2 AI capital: $\delta_a = 0.25$

AI capital consists primarily of GPUs, accelerators, and specialized data center equipment. Unlike traditional structures and machines, this equipment faces rapid technological obsolescence in addition to physical wear. The depreciation rate $\delta_a = 0.25$ corresponds to a four-year useful life under straight-line depreciation.

This choice is informed by the accounting policies disclosed in SEC filings by the firms that own the majority of this equipment. Prior to 2022, the standard useful-life assumption for server and networking equipment at major technology firms was three to four years. Between 2022 and 2024, several firms extended these estimates:

- Microsoft extended the estimated useful life of server and network equipment from four years to six years, effective fiscal year 2023. The company attributed the change to investments in software that increased the efficiency of operating server equipment, and estimated that the change would increase fiscal year 2023 operating income by approximately \$3.7 billion (Microsoft Corporation, 10-K, fiscal year 2022).

- Alphabet extended the estimated useful life of servers from four years to six years and certain network equipment from five years to six years, effective January 2023. The change reduced 2023 depreciation expense by \$3.9 billion and increased net income by \$3.0 billion (Alphabet Inc., 10-K, fiscal year 2023).

- Meta Platforms extended server useful lives in three stages: from four years to 4.5 years, then to five years, and most recently to 5.5 years effective fiscal year 2025. The latest extension is expected to reduce 2025 depreciation expense by approximately \$2.9 billion (Meta Platforms Inc., 10-K, fiscal year 2024).

- Amazon increased the estimated useful life of its servers in three stages: from three to four years in January 2020, from four to five years in January 2022, and from five to six years in early 2024. However, in February 2025 Amazon reversed course, shortening the useful life of a subset of servers and networking equipment from six years back to five years, citing "the increased pace of technology development, particularly in the area of artificial intelligence and machine learning." The reversal increased quarterly depreciation expense by $217 million and was accompanied by a $920 million accelerated depreciation charge in Q4 2024 for early-retired equipment (Amazon.com Inc., 10-Q, Q1 2025).

The extensions have meaningful consequences for reported earnings: the combined reduction in annual depreciation expense across these four firms exceeds $10 billion. Whether the extensions reflect genuine increases in equipment longevity or optimistic accounting remains an open question. Amazon's partial reversal suggests that actual useful lives for AI-intensive workloads may be shorter than the extended estimates imply. For further discussion, see Kapoor and Narayanan (2025).

Our baseline of $\delta_a = 0.25$ (four years) is conservative relative to the pre-extension industry norm of three to four years and more conservative still relative to estimates that AI accelerators running at high utilization face effective useful lives of two to three years. The qualitative results of the model are not sensitive to moderate variation in δ_a; what matters is that δ_a is substantially larger than δ_n, so that maintaining a given capital stock requires a much higher ratio of investment to capital in the AI sector.

Van Nieuwerburgh (2026) offers a complementary perspective by decomposing the cost of an integrated AI campus into two components. Approximately one-third of total cost is attributable to the data-center facility and supporting power infrastructure (about $11 billion per gigawatt for the facility and $2 billion per gigawatt for substations, transmission, and grid interconnection), and approximately two-thirds is attributable to IT equipment, including GPUs, networking, and storage (approximately \$28 billion per gigawatt at his estimates). Van Nieuwerburgh (2026) further observes that the two components face different forces of obsolescence: the facility and power components are long-lived physical assets, whereas AI hardware "has a shorter economic life and faces substantial technological obsolescence risk, as it depends more directly on the evolution of model architectures and demand for compute." A calibration that adopts this decomposition, with a low depreciation rate for the facility and power component and a moderate rate for the IT component, yields a weighted-average δ_a below our baseline of 0.25. Appendix A.5 below reports the model's quantitative implications under such an alternative calibration.

A.4.3 Pre-boom capital stocks

The pre-boom capital stocks are inferred from the steady-state condition $I_j = \delta_j K_j$. For the AI sector, the 2022 capital expenditure of the five largest firms (\$155B; see Table 1) serves as the pre-boom investment level, giving $K_{a0} = 155/0.25 = \$620\text{B}$. We use 2022 because it is the last full year before the structural break: combined capital expenditure was essentially flat between 2022 (\$155B) and 2023 (\$150B), indicating that the boom had not yet affected investment decisions.

For the non-AI sector, we use aggregate U.S. gross private fixed investment from BEA NIPA Table 1.1.5, which was approximately \$4.7 trillion in 2022 (FRED series GPDI, less change in private inventories). Subtracting the AI share gives non-AI investment of $4{,}700 - 155 = \$4{,}545\text{B}$. The non-AI capital stock is then $K_{n0} = 4{,}545/0.07 \approx \$65{,}000\text{B}$.

The resulting AI share of the total capital stock is $620/(620+65{,}000) = 0.9\%$—well below its 3.3% share of investment. This wedge is a direct consequence of the high depreciation rate: at $\delta_a = 0.25$, a given level of investment supports a much smaller capital stock than at $\delta_n = 0.07$.

A.5 Alternative depreciation calibration

This subsection recomputes the main calibration under Van Nieuwerburgh (2026)'s two-component decomposition of integrated AI campus cost. Approximately one-third of total cost is the data-center facility and supporting power infrastructure—a long-lived physical asset—and approximately two-thirds is IT equipment subject to rapid technological obsolescence. Assigning $\delta^F = 0.05$ (twenty-year useful life, consistent with BEA rates for industrial structures) and $\delta^{IT} = 0.20$ (five-year useful life, the lower end of Van Nieuwerburgh (2026)'s suggested range), the weighted-average AI depreciation rate is

$$\delta_a^{IA} \;=\; \tfrac{1}{3}\cdot 0.05 \,+\, \tfrac{2}{3}\cdot 0.20 \;=\; 0.15. \tag{23}$$

At the upper end of the IT range ($\delta^{IT} = 0.17$, six-year useful life), the weighted average falls further to $\delta_a^{IA} \approx 0.13$. All other parameters retain their baseline values. We retain $\hat{\xi} = 0.99$ as a structural feature of the productivity process; the sensitivity to re-identifying $\hat{\xi}$ from the cumulative-investment formula under the lower δ_a is discussed at the end of the subsection.

Pre-boom stocks and the identification of p_*. The lower depreciation rate implies a larger pre-boom AI capital stock at the same 2022 capital expenditure of \$155B: $K_{a0}^{IA} = 155/0.15 \approx$ \$1,033B, versus \$620B at baseline. Pre-boom AI output rises commensurately to $Y_{a0}^{IA} \approx$ \$930B and the AI share of pre-boom GDP rises from 3.1% to 3.8%.

We identify p_* here using the same aggregate Gordon-growth decomposition as the main text (Section 3.3): $p_* = g/(e^{\xi} - 1)$ with $g = r - D/P \approx 7\%$. This identification does not depend on δ_a, so the alternative calibration adopts the same numerical value as the main paper, $p_*^{IA} \approx 0.041 \approx 0.04$. The model-implied normalized earnings $\bar{E}_a$ rise by approximately 32% under the lower depreciation rate, so the same AI-sector market value now maps to a roughly 24% lower observed price-earnings ratio; that adjustment shows up in observed P/E space rather than in p_*.

Post-window output shares and GDP growth. Table A.3 reports post-window quantities. AI output shares rise by 2–6 percentage points across scenarios. Cumulative GDP growth rises commensurately. Post-window replacement investment is nearly unchanged because $I_a^* = \delta_a K_a^* \propto \delta_a/(r_a+\delta_a)^{1/(1-\alpha)}$, and at $\alpha = 1/3$ the ratio $\delta_a/(r_a+\delta_a)^{3/2}$ varies by only about 8% between the two calibrations.

Table A.3: Post-window output shares and GDP growth under the alternative depreciation calibration

	AI share (%)	Cum. GDP growth (%)	Transition avg. (%/yr)	Post-window I_a ($B/yr)
Baseline ($\delta_a = 0.25$, $p_* = 0.041$)				
Pre-boom	3.1	—	—	155
Moderate	8.0	5.4	0.77	441
Transformative	19.0	19.7	2.81	1,188
Singularity	38.7	58.2	8.31	3,197
Alternative ($\delta_a = 0.15$, $p_*^{IA} = 0.041$)				
Pre-boom	3.8	—	—	155
Moderate	9.8	6.7	0.95	441
Transformative	22.7	24.4	3.49	1,188
Singularity	44.1	72.2	10.31	3,197

Notes: Construction parallels Table 5 of the main paper. The alternative uses $\delta_a = 0.15$ and the main-text $p_*^{IA} = 0.041$; all other parameters as in the main paper. The baseline row reports the main paper's headline scenarios re-evaluated at $p_* = 0.041$ to keep both calibrations on the same footing. Cumulative GDP growth is the one-time aggregate level increase from the pre-boom economy to the post-window steady state. Transition average is the annualized contribution to GDP growth over the seven-year 2024–2030 window. Pre-boom investment is anchored to the 2022 capital expenditure observation of $155B under both calibrations.

The transition investment paths and GDP transition figures are qualitatively the same as Figures 1 and 2 of the main paper, with peaks and post-window levels modestly higher reflecting the larger underlying capital stock and target.

Asset pricing. Because $p_b = p_*^{IA} = 0.041$ is unchanged from the baseline, the asset-pricing differences under the alternative depreciation rate come entirely through the s_a channel of the bounding-case mapping $\xi_b = s_a(1-\alpha)\xi$. The scenario-specific boom sizes are larger under the alternative ($\xi_b^{IA} = 0.065,\ 0.150,\ 0.291$ for moderate, transformative, and singularity, versus baseline $0.053,\ 0.126,\ 0.256$), producing modestly higher equity premia. At $\gamma = 4$, the moderate-scenario equity premium rises from 0.59% to 0.61%, the transformative premium from 0.76% to 0.84%, and the singularity premium from 1.30% to 1.48%; risk-free rates move by at most a few basis points. In the absorbing limit $\xi_b = (1-\alpha)\xi = 0.660$, both the boom probability and the boom size are independent of δ_a, so equity premia and risk-free rates are exactly invariant to the depreciation calibration: at $\gamma = 4$, the absorbing-limit equity premium is 4.03% and the risk-free rate is 3.75% under both calibrations.

Re-identifying $\hat{\xi}$ as a further sensitivity. Under the lower depreciation rate, mechanically re-identifying $\hat{\xi}$ from the frictionless cumulative-investment formula,

$$e^{\hat{\xi}^{IA}} = \frac{\bar{I}/K_0^{IA} + 1 - \delta_a^{IA}}{1 + 3\,\delta_a^{IA}} = \frac{2452/1033 + 0.85}{1.45} \approx 2.22,$$

gives $\hat{\xi}^{IA} \approx 0.80$, below the structural value $\hat{\xi} = 0.99$. Under this smaller boom size, post-window AI shares fall to roughly 8.3%, 16.7%, and 30.9% (moderate, transformative, singularity) and cumulative GDP growth falls correspondingly. We do not adopt this as the primary alternative because $\hat{\xi}$ is interpreted in the model as a real productivity shift; treating it as an accounting-implied residual would conflate the structural shock with short-run depreciation conventions.

Summary. Under Van Nieuwerburgh (2026)'s alternative depreciation calibration, the main paper's qualitative conclusions survive intact. AI output shares are larger by 2–6 percentage points across scenarios and cumulative GDP growth is larger commensurately.

The long-run boom probability p_* is identified from the same aggregate Gordon decomposition as the main text and is unchanged at 0.041. Asset-pricing moments shift modestly upward in the bounding-case mapping (because ξ_b scales with the larger s_a) and are exactly invariant in the absorbing limit.

B Frictionless investment paths

Figure A.1 shows the investment paths from the frictionless base case, in which capital adjusts immediately to K^* each period. Two artifacts of the frictionless assumption are visible: a one-period investment spike far above the observed data, and negative investment when p reverts to p_* without a new boom. Compare with the smoothed paths in Figure 1.

Figure A.1: Frictionless investment paths by scenario. Gray bars show actual and projected AI capital expenditure through 2027. Colored bars show the model-implied frictionless investment. Dashed lines delineate the high-probability window (2028–2030).