# What Investment Data Implies about the AI Transition*

Jessica A. Wachter† Jonathan D. Wachter‡

March 24, 2026

**Abstract**

The five largest U.S. technology firms spent $400 billion on capital expenditure in 2025 and are forecast to spend $660 billion in 2026—an increase of over 50%. These firms face 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 the observed increase in investment that has occurred and is projected to continue through 2027, implying that each boom raises AI-sector productivity by a factor of roughly 2.5. 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, and an elevated permanent probability tied to the valuation of the AI sector. The implied additional cumulative GDP growth ranges from 7% to 40% by 2030, with AI shares of the economy ranging from 8% to 30%. Long-term annual growth is in expectation around 8% but with substantial risk. With risk aversion of 3, and an elasticity of intertemporal substitution equal to 1, the riskfree rate increases by 1 percentage point, and the equity premium rises by 4 percentage points.

*Keywords*: Artificial intelligence, Rare booms, Capital investment
*JEL codes*: E22, G12, O33

[1]We thank Sylvain Catherine, Gary Gensler, and Alex Imas 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 or its affiliates. Point72 and its affiliates are not responsible for, and did not verify for accuracy, any of the information contained herein.

[2]Corresponding author. Department of Finance, The Wharton School, University of Pennsylvania, and NBER. Email: jwachter@wharton.upenn.edu.

[3]Point72 Asset Management.

# 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 0.5–0.7 percent over the next decade—roughly one year of normal TFP growth. 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, to inform this debate. A growing literature studies the macroeconomic implications of transformative AI. Acemoglu (2024) uses a task-based framework and finds modest productivity gains; Trammell and Korinek (2025) and Korinek and Suh (2024) model scenarios for the transition to artificial general intelligence (AGI), showing how growth rates, wages, and capital accumulation depend on the pace of automation; Restrepo (2025) distinguishes between bottleneck and supplementary work to characterize the long-run growth implications. Chow et al. (2025) argue that long-term real interest rates contain information about the likelihood of transformative AI and the associated existential risk. Yotzov et al. (2026) present the first large-scale firm-level data on AI adoption, finding that most firms report minimal productivity impact so far. Our approach is complementary: rather than modeling the tasks that AI automates or surveying firms, we use the capital expenditure data, together with a zero-NPV condition, to identify the size of the AI productivity booms. We then trace out scenarios based on this size, and on industry forecasts, and finally we use valuations to capture the permanent probability of ongoing productivity booms.

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 $160 billion in 2022, rose to $250 billion in 2024, and is

forecast at $660–710 billion for 2026. These are not projections by outside analysts; they are management forecasts disclosed to investors. 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, this size is roughly a multiple of 2.5 of current productivity. This boom size informs the remaining calibration. With no other changes to productivity, this multiple alone translates into growth of 6 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. With maximal uncertainty (probability of 1/2) of a boom during the 2028–2030 period, we trace out a range of scenarios between 6 and 40 percentage points by 2030. In a transformative to singularity scenario, AI is a larger share of the economy. We create long-term forecasts assuming that this sector, which eventually becomes the entire economy, experiences an 8% probability per year of a boom of a similar size to the one experienced now. Such a scenario is roughly consistent with current valuations.

Looking forward, a two-year window with elevated boom probability generates three scenarios depending on whether zero, one, or two additional booms occur. In the frictionless base case, all quantities are solved in closed form. The optimality condition implies that cumulative 2024–2027 investment requires annual value creation of at least $917 billion—far in excess of task-based estimates—and that the buildout destroys value unless the boom option value raises expected returns above the user cost of capital. The model identifies three scenarios—moderate, transformative, and singularity—with calibrated output shares

ranging from 5% to over 25% of GDP.

Section 4 derives the implications for GDP growth under each scenario. GDP growth ranges from roughly 5% by the end of the decade in the moderate case to over 40% in the singularity case, where multiple productivity shocks raise both output and investment, which in turn raises output further.

These scenarios, while highly variable, correspond to ranges of values currently viewed as plausible by academics and by decision-makers. What follows the window is far more uncertain. Our calibration using current and near-term projections suggest that large gains to productivity are within the realm of possibility. Matching reasonable long-run price-earnings ratios at standard discount rates, together with booms of this size, leads to a range of possible boom probabilities. We consider GDP trajectories under the resulting probabilities, with the 75th percentile at 300% GDP growth by 2050. This output growth arises from rising AI shares – we assume no spillovers to the rest of the economy – and the AI share at the 75th percentile is above 70 percent.

While previous sections take the discount rate as given, Section 5 derives implications for the equity premium, and for interest rates, taking cash flows as given. Despite predicting expected growth of upwards of 8% past the transition, implications for interest rates are modest, due to precautionary savings. For risk aversion of 3, AI technology adds 4 percentage points to the equity premium when it grows to near the size of the economy. Section 6 places the AI boom in the context of historical rare booms documented by Tsai and Wachter (2016) and computes the implications for GDP growth. 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) |
|---|---|---|---|---|---|
| Amazon | 63 | 54 | 84 | 125 | 200 |
| Alphabet | 32 | 32 | 53 | 75–93 | 175–185 |
| Microsoft | 28 | 41 | 55 | 80 | 116–140 |
| Meta | 32 | 28 | 39 | 65–72 | 115–135 |
| Oracle | 9 | 9 | 12 | 21 | 50 |
| **Total** | **164** | **164** | **243** | **∼400** | **∼660–710** |

*Sources:* Company 10-K filings (2022–2024); quarterly earnings call disclosures (2025–2026). Ranges reflect the low and high ends of management forecasts; midpoints are used for totals.

Table 1 suggests that something unusual occurred in 2024 Aggregate capital expenditure was essentially constant at $164 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 services—alongside any pre-boom AI spending. We use the full $164B 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

[1]For Microsoft, which does not provide formal annual capital expenditure forecasts, the 2026 range reflects sell-side analyst estimates based on quarterly disclosures. Microsoft's figures include capital leases. For all other firms, the figures represent capital expenditure as reported in the statement of cash flows.

size $\hat{\xi}$ is correspondingly larger. Appendix A.1 discusses sensitivity to this attribution choice. In 2024, aggregate spending rose 48% to $243 billion. The acceleration continued in 2025, with a further 65% increase to approximately $400 billion, and management forecasts for 2026 imply yet another 65% increase to $660–710 billion.[2] Altogether, capital expenditure is forecast to roughly triple between 2024 and 2026. These figures are lower bounds: they include only the five largest firms, and independent analyst estimates for 2026 run as high as $770 billion (UBS, February 2026). Including smaller entrants focused specifically on AI compute would imply AI-related capital expenditure likely exceeding $800 billion.

## 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) |
|---|---|---|---|---|
| AI infrastructure capex ($B) | 164 | 243 | 400 | 660–710 |
| U.S. nominal GDP ($T) | 25.5 | 29.2 | ∼30.5 | ∼31.5 |
| U.S. gross private fixed investment ($T) | ∼4.7 | ∼5.2 | ∼5.4 | ∼5.5 |
| Share of GDP (%) | 0.6 | 0.8 | 1.3 | 2.1–2.3 |
| Share of private fixed investment (%) | 3.5 | 4.7 | 7.4 | 12.0–12.9 |

*Sources:* GDP and gross private fixed investment from the Bureau of Economic Analysis (NIPA Table 1.1.5; FRED series FPIA). 2025–2026 GDP and aggregate investment are author projections assuming 3% nominal growth.

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, there are in-

[2]These management forecasts are broadly in line with Goldman Sachs Global Investment Research (2025), who project cumulative spending by the five largest firms of $1.15–1.4 trillion over 2025–2027. Given realized 2025 spending and current 2026 forecasts, the Goldman projection implies 2027 capital expenditure in the range of $90–340 billion—a deceleration from the current pace.

consistent measurement conventions across the five firms: Microsoft's figures include capital leases and incorporate analyst imputations, while the other firms report cash-flow capital expenditure. We use the data as reported; harmonizing these conventions could affect both the levels and the growth rates, though the impact is likely modest relative to the factor-of-four surge that drives the calibration. 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. The depreciation rates $\delta_a$ and $\delta_n$ measure physical wear and technological obsolescence, which are inherently real quantities. The required return $r_a = 0.15$ is a real rate, reflecting the risk premium on AI-sector equity; Vanguard Group (2025) use the same 15% baseline discount rate to value AI-sector earnings, with a range of 10–25%.[3]

AI infrastructure investment rose from 3.5% of total U.S. private fixed investment in 2022 to a projected 12.5–14% in 2026. As a share of GDP, it rose from 0.6% to over 2%. As a benchmark, AI capital expenditure as a share of GDP remains below the peak levels of the late-1990s telecommunications investment cycle (approximately 1.5% of GDP), but the current trajectory implies that it will surpass those levels within a year. 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]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 main effect is to tighten the finiteness bound to $p_* < 6.9\%$, bringing $p_* = 0.06$ close to the bound and pushing the P/E ratio well above the plausible range of 15–35; matching observed valuations would require $p_* \approx 0.02$–$0.05$. Output shares and GDP growth are modestly lower (e.g., the moderate AI share falls from 7.8% to 6.9%), and the asset-pricing effects are correspondingly smaller (at $\gamma = 4$, the singularity equity premium is 1.15% vs. 1.26%). The qualitative conclusions are unaffected.

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

## 2.3 Investment projections beyond 2026

Management forecasts extend only one year. Beyond 2026, physical capacity data provide a guide. Amodei (2026) characterizes the current growth rate of data center power capacity—a physical proxy for AI compute—as approximately a factor of 3 per year. The trajectory from roughly 10–15 gigawatts in 2025 to an estimated 100 gigawatts by 2028 implies an average compound growth factor of approximately 2 per year, as capacity constraints and supply bottlenecks moderate the pace from its current rate. At capital costs on the order of $10–15 billion per gigawatt, this trajectory implies industry-wide capital expenditure of several trillion dollars per year by the end of the decade. Musk (2026) estimates that the world currently deploys 20–25 gigawatts of AI compute and envisions scaling to hundreds of gigawatts within five years, noting that electricity, not capital, is the binding constraint. For context, total average U.S. electricity consumption is approximately half a terawatt.

These projections are uncertain, but they provide a useful range. For the model in Section 3, we consider two scenarios: a moderate scenario in which investment growth decelerates after 2026 and levels off, and a transformative scenario in which 50% annual growth continues through 2029. However unlikely the latter may appear, it roughly corresponds to the outcome that members of the industry highlight as *the most likely*. Table 3 summarizes.

**Table 3:** Investment scenarios ($ billions)

| | 2025 | 2026 | 2027 | 2028 | 2029 |
|---|---|---|---|---|---|
| Moderate | 400 | 660 | 750 | 800 | 800 |
| Transformative | 400 | 660 | 990 | 1,485 | 2,228 |

*Notes:* 2025 and 2026 are actual and forecast, respectively. The moderate scenario assumes 14% growth in 2027, then constant investment. The transformative scenario assumes 50% annual growth from the 2026 base.

# 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 (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 $\alpha$ is the capital share (we set $\alpha = 1/3$ throughout), $z_{jt}$ 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 $\delta_j$ 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 $e^{\xi}$, 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. In this open-economy risk neutral setting, agents' attempts to anticipate changes in $p_t$ will be second-order. Note that 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 $p_{t+1}$.

## 3.2 Optimal investment: the frictionless case

Sector $a$ maximizes the present discounted value of earnings, assuming discount rate $r_{at}$:

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

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 $e^{\xi}$, 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 $\xi$ and the shift in $p$, not on $r_a$. This means we can identify $\xi$ 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 $\alpha$ of the production function operates on the predetermined capital stock. At $\alpha = 1/3$ and $\xi = 0.90$, $e^{(1-\alpha)\xi} = 1.82$: output rises 82%.

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 $e^{\xi}$ is now realized: with $\xi = 0.90$, $e^{\xi} = 2.45$. The remaining factor of $e^{\alpha\xi} = 1.35$ comes from capital adjustment.

Thus $\alpha$ 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

For non-AI capital we set $\delta_n = 0.07$, the standard value in macroeconomic calibrations based on the ratio of aggregate depreciation to the capital stock in the BEA Fixed Asset Tables. For AI capital we set $\delta_a = 0.25$, corresponding to a four-year useful life for server and GPU equipment. Appendix A.1 discusses both calibrations in detail.

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 $K_{a0} = I_{a,\text{pre}}/\delta_a =$

164/0.25 = \$656B, where \$164B 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 $K_{n0} = (4{,}700 - 164)/0.07 \approx$ \$65,000B. The AI sector's share of the total capital stock is therefore only 0.9%, despite accounting for 3.5% of investment—the wedge arises entirely from the difference in depreciation rates. Appendix A.1 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 $\Omega \approx 1$) 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 $\xi$ equals

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

Table 1 shows investment unchanged between 2022 and 2023, and a rapid increase thereafter. We calibrate $\xi$ by matching the model's cumulative investment to the observed capital expenditure of \$2,293B over 2024–2027 ($= 243 + 400 + 660 + 990$; see Table 3). 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 $e^{\xi}$ implies a gap between actual and desired capital. Substituting in

$$e^{\hat{\xi}} = \Big(\frac{2293}{656} + 0.75\Big)\frac{1}{1.75} \approx 2.43, \qquad \hat{\xi} \approx 0.90.$$

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.[5] Were the boom to be smaller than this, the investments would be NPV-negative.

Section 2.3 describes two scenarios for investment beyond 2027 (Table 3). 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, $p = 0.5$. 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 $p_{2029} = p_{\max}$

[5] 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.

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 $\xi$ 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. We make this assumption in order that the calibration of $\xi$ be transparently tied 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 $\xi$. 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 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.[6] With two independent draws at $p_{\max} = 0.50$, there are exactly three outcomes:

[6] 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.

| 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. Appendix A.2 discusses the calibration of $p_{\max}$ and its relation to existing AI forecasts. In the moderate scenario, the productivity gain from the initial boom is all there is. In the transformative scenario, one additional boom raises productivity by another factor of $e^{0.90} \approx 2.5$. In the singularity scenario, two additional booms raise it by $e^{2\times 0.90} \approx 6.0$ 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 $K_{2029}$.

*$I_{2028}$: same for all scenarios.* With $p_{\max} = 0.50$, the optimal capital stock entering 2029 rises from $K^*(z_1, 0) = \$1{,}605\text{B}$ to $K^*(z_1, p_{\max}) = \$2{,}682\text{B}$, where $z_1 = z_0 e^{\xi}$ is the post-initial-boom productivity. The required investment is $I_{2028} = K^*(z_1, p_{\max}) - (1-\delta_a)K^*(z_1, 0) = \$1{,}478\text{B}$. This is identical across all four scenarios because the 2029 boom has not yet been observed.

*$I_{2029}$: 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: $I_{2029} = \delta_a K^*(z_1, p_{\max}) = \$671\text{B}$. If a boom occurred (transformative (a) and singularity), $z$ jumps to $z_2 = z_0 e^{2\xi}$ and the capital stock must adjust: $I_{2029} = K^*(z_2, p_{\max}) - (1 - \delta_a)K^*(z_1, p_{\max}) = \$4{,}553\text{B}$.

*$I_{2030}$: $p$ reverts to $p_*$.* 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. With adjustment costs (Section 4), the transition is smooth: investment rises during the window and then gradually declines toward replacement levels. If a late boom raised $z$, investment remains elevated.

*$I_{2031+}$.* The economy converges to its scenario-specific steady state. Long-run annual replacement investment is $\delta_a K^*$: \$431B (moderate), \$1,055B (transformative), \$2,583B (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 $\theta_a$ does not affect $K^*$, the calibrated $\hat{\xi}$, output shares, and P/E ratios are unchanged. Figure 1 displays these paths, whereas Figure 6 shows the corresponding path with no adjustment frictions.

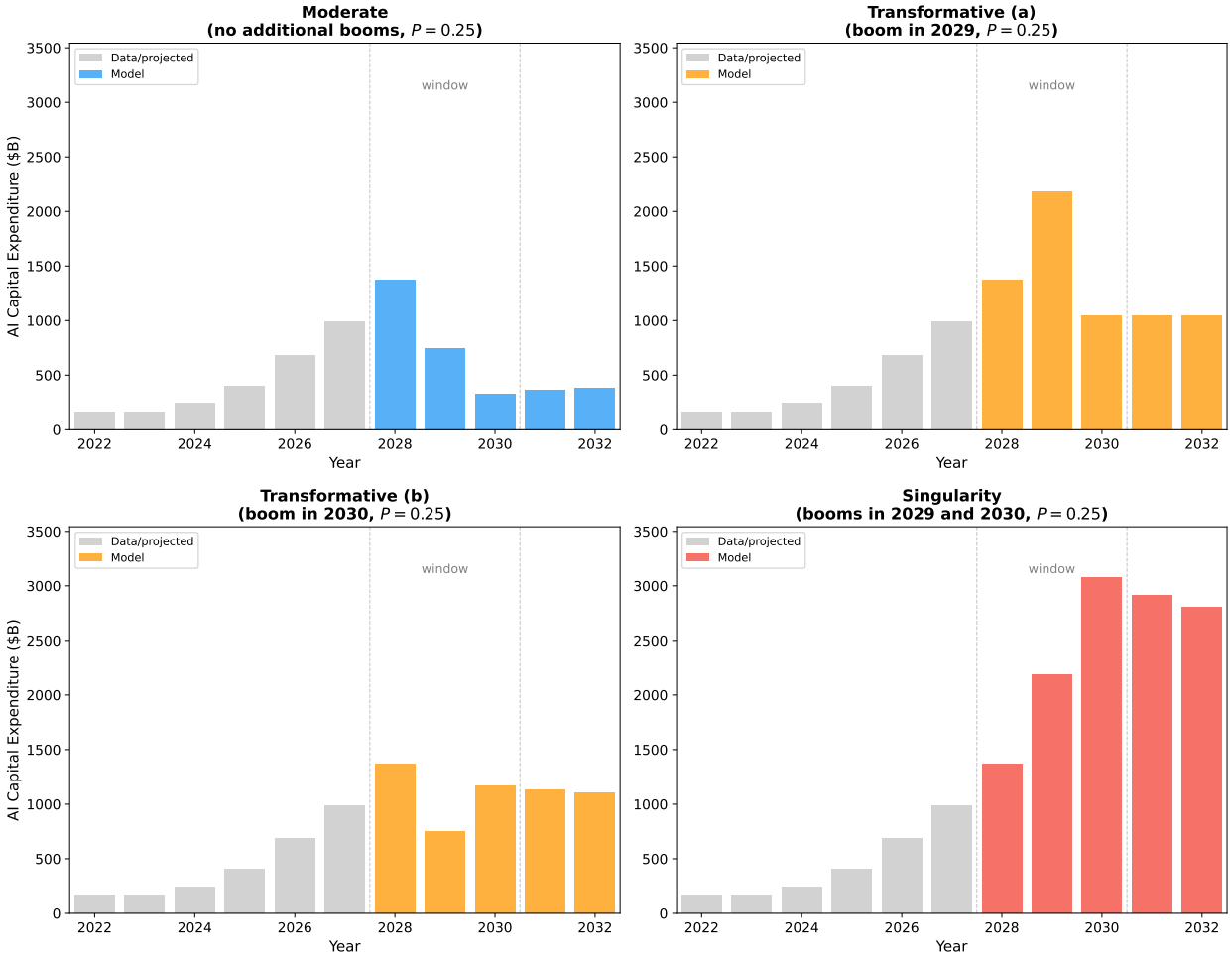



**Figure 1:** Investment paths by scenario. Gray bars show actual and projected AI capital expenditure through 2027. Colored bars show investment from the window onward, incorporating convex adjustment costs ($\theta_a = 3$). Dashed lines delineate the high-probability window (2028–2030). Investment 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.

**Identifying $p_*$ from valuations.** What happens following 2030 is necessarily much more speculative. We attempt to identify a long-run boom probability $p_*$ from ranges of valuations of AI-sector capital, taking the boom size as given. The goal of this exercise is to give a reasonable range for this parameter and derive consequences for growth.

Define earnings as output net of replacement investment:

$$E_{at} \equiv Y_{at} - \delta_a K_{at} = z_{at}^{1-\alpha} K_{at}^{\alpha} - \delta_a K_{at}. \tag{15}$$

In steady state, $I_a = \delta_a K_a$ (investment exactly replaces depreciation), so earnings equal the firm's free cash flow. This is not accounting profit; it is the economic surplus after maintaining the capital stock.

The value of the AI sector is the present discounted value of the earnings stream:

$$V_{at} = \mathbb{E}_t \sum_{s=0}^{\infty} \frac{E_{a,t+s}}{(1+r_a)^s}. \tag{16}$$

To evaluate this, note that the optimal capital stock (9) is $K^*_{a,t+1} = z_{at} \cdot \kappa(p_t)$, where $\kappa(p) \equiv [\alpha\,\Omega(p)/(r_a + \delta_a)]^{1/(1-\alpha)}$. Define *normalized earnings* as the cash flow that obtains once capital has fully adjusted to the current productivity level:

$$\bar{E}_{at} \equiv z_{at}^{1-\alpha}\big[z_{at}\,\kappa(p_*)\big]^{\alpha} - \delta_a\, z_{at}\,\kappa(p_*) = z_{at}\big[\kappa(p_*)^{\alpha} - \delta_a\,\kappa(p_*)\big]. \tag{17}$$

Because $\bar{E}_{at}$ is proportional to $z_{at}$, its growth factor is determined entirely by the boom process: if a boom occurs at $t+1$, $\bar{E}_{a,t+1}/\bar{E}_{at} = e^{\xi}$; otherwise the ratio is one. The one-period-ahead expected growth factor is therefore

$$\Lambda \equiv \mathbb{E}_t\left[\frac{\bar{E}_{a,t+1}}{\bar{E}_{at}}\right] = (1-p_*) + p_*\,e^{\xi},$$

and because successive booms are i.i.d., $\mathbb{E}_t[\bar{E}_{a,t+s}] = \bar{E}_{at}\,\Lambda^s$. Substituting into (16) and summing the geometric series:

$$\frac{V_{at}}{\bar{E}_{at}} = \sum_{s=0}^{\infty}\left(\frac{\Lambda}{1+r_a}\right)^s = \frac{1+r_a}{1+r_a-\Lambda} = \frac{1+r_a}{1+r_a-\left[(1-p_*)+p_*\,e^{\xi}\right]}. \tag{18}$$

Equivalently, one can think of each term in the sum as the price of an "earnings strip"—a claim to $\bar{E}_{a,t+s}$ alone—discounted at rate $r_a$ and scaled by the expected growth $\Lambda^s$. The finiteness condition (5) ensures the series converges; at $\hat{\xi} = 0.90$ and $p_* = 0.06$, the condition requires $r_a > 8.8\%$.

We note an important timing subtlety. Because capital is predetermined ($K_{at} = z_{a,t-1}\,\kappa(p_*)$ in steady state), *realized* current-period earnings $E_{at} = Y_{at} - \delta_a K_{at}$ do not jump by exactly $e^{\xi}$ in the period of a boom. Instead, $E_{at}$ reflects a mismatch between the new productivity $z_{at}$ and the old capital stock $K_{at}$. The normalized earnings $\bar{E}_{at}$ remove this one-period lag by evaluating earnings at the capital stock that is optimal for $z_{at}$. For the purpose of long-run valuation—where capital is at its target in every period except the impact period—the distinction is immaterial, and $V/\bar{E}$ provides the correct long-run price–earnings ratio.

This expression is valid in steady state, where $I_a = \delta_a K_a$ and no adjustment costs are incurred. Outside steady state—e.g., during the transition when investment spikes and adjustment costs are positive—earnings do not grow at rate $\Lambda$, so the formula does not apply directly. The table should therefore be interpreted as a long-run valuation anchor, not an exact pricing formula for the transition period.

Table 4 reports this ratio across $r_a$ and $p_*$. Here $r_a$ serves as the required return on AI-sector equity—the project discount rate for the earnings stream.Because $p_*(e^{\xi}-1)$ approaches $r_a$ as $p_*$ rises toward the finiteness bound $r_a/(e^{\xi}-1)$, the ratio is convex and steeply increasing: small changes in $p_*$ near the bound have large effects on valuations.

The P/E range of 15–35 shown in Figure 2 is informed by a sum-of-the-parts decom-

**Table 4:** Price-to-Earnings Ratio of AI Sector ($V/\bar{E}$)

| | $r_a = 10\%$ | $r_a = 12\%$ | $r_a = 15\%$ | $r_a = 20\%$ |
|---|---|---|---|---|
| $p_* = 0.02$ | 15.5 | 12.3 | 9.5 | 7.0 |
| $p_* = 0.04$ | 26.1 | 18.0 | 12.5 | 8.4 |
| $p_* = 0.06$ | 83.6 | 33.8 | 18.2 | 10.6 |
| $p_* = 0.08$ | — | 265.8 | 33.6 | 14.2 |

*Notes:* $\xi = 0.90$. $\bar{E}$ denotes normalized earnings (17), evaluated at the capital stock that is optimal for the current productivity level. An entry of "—" indicates that the finiteness condition is violated. The finiteness bound is $p_* < r_a/(e^{\xi} - 1)$: approximately 6.9% at $r_a = 10\%$, 8.3% at 12%, 10.4% at 15%, and 13.8% at 20%.

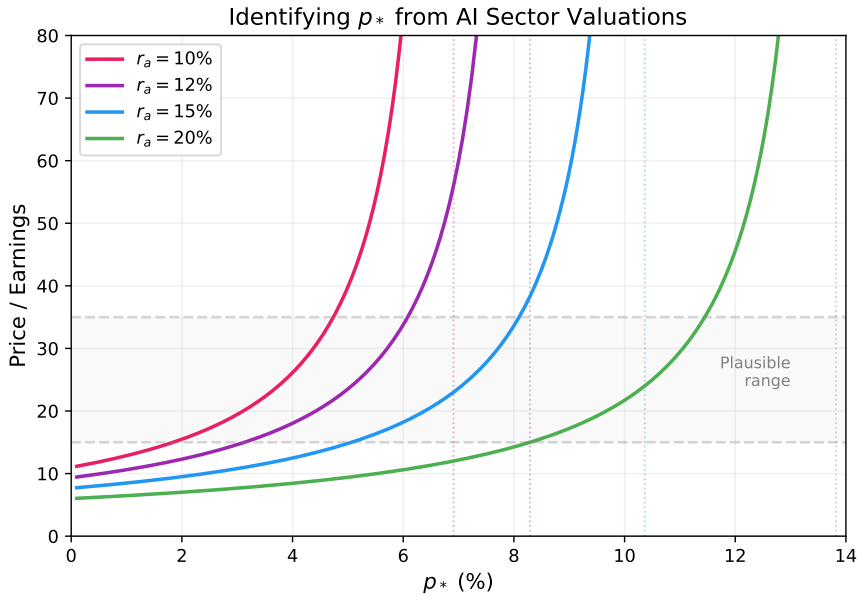



**Figure 2:** Identifying $p_*$ from AI sector valuations. Each curve plots the price-to-earnings ratio $(1+r_a)/(1+r_a-\Lambda)$ as a function of $p_*$ for a given required return $r_a$. Dotted vertical lines mark the finiteness bound $p_* = r_a/(e^{\xi} - 1)$. The shaded band indicates a plausible P/E range of 15–35.

position: we strip out non-AI enterprise value from the consolidated market capitalizations of the five firms using segment-level revenue and operating-income disclosures, and divide the residual AI-attributed equity value by the corresponding earnings $Y_a - \delta_a K_a$. This procedure is approximate—the firms do not report AI-segment financials—but it provides data-disciplined bounds rather than a free parameter. If the required return on AI capital is $r_a = 15\%$ and the observed P/E ratio for the AI sector is in the range 12–18, the implied $p_*$ is 4–6%. At $r_a = 12\%$, the same valuation range implies $p_*$ of 2–4%.[7] We take $p_* = 0.06$ as a baseline, corresponding to a 6% annual probability of a major AI productivity breakthrough in normal times.

**Implied value creation.** The user-cost condition (6) has a direct quantitative implication. Over 2024–2027, cumulative AI investment totals $\bar{I} = \$2{,}293\text{B}$ (Table 3). Each dollar of capital must earn at least the user cost $r_a + \delta_a$ per year; otherwise the investment is NPV-negative. This is not a model-specific result—it follows from any setting in which firms equate the marginal product of capital to the user cost. The annual gross value the installed capital must generate approximately equals

$$(r_a + \delta_a) \times \bar{I} = 0.40 \times 2{,}293 = \$917\text{B per year},$$

roughly 3% of GDP. The moderate scenario in the model—one boom followed by no further booms—formalizes this calculation and delivers a comparable figure: the AI sector's annual output is $Y_a = \alpha^{-1}(r_a + \delta_a)K_a^* \approx \$1{,}700\text{B}$, of which the capital share $\alpha Y_a \approx \$560\text{B}$ represents the return to capital.

If the investment were NPV-negative, firms would be destroying value. At the scale of

[7] The investment data used to calibrate $\xi$ 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 $\xi$ is unaffected (Proposition 3.1); the main effect is to tighten the finiteness bound and lower the implied $p_*$.

hundreds of billions per year, sustained value destruction would quickly become apparent. The value creation implied by the optimality condition is therefore a revealed-preference lower bound on what these firms expect the AI capital stock to produce. Section 3.4 discusses alternative interpretations under which the investment may overstate the true productivity gain.

## 3.4 Alternative interpretation: overoptimism and coordination failure

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

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. If the current AI buildout is of this character, the revealed-preference estimate of $\hat{\xi}$ overstates the true productivity gain, and the implied growth and asset-pricing effects are correspondingly smaller.

Second, even absent pure overoptimism, managers may be wrong about the specific technological trajectory. The current investment surge is predicated on the assumption that scaling laws—the empirical regularity that model performance improves as a power law in compute, data, and parameters—will continue to hold. If scaling encounters diminishing returns at current or near-current levels, the realized productivity gain will fall short of what the investment path implies. The history of general-purpose technologies offers examples of both outcomes: electrification ultimately exceeded early expectations, while nuclear power fell well short. Which pattern AI follows is, at this stage, an empirical question that the

model cannot answer.

Third, even if the aggregate productivity gain is correctly anticipated, 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 co-operative 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.

The calibration of $\hat{\xi}$ is partially insulated from the third concern. Because $\hat{\xi}$ is identified from the *ratio* of post-boom to pre-boom investment (Proposition 3.1), and $r_a + \delta_a$ cancels from this ratio, strategic motives that inflate investment symmetrically before and after the boom leave $\hat{\xi}$ unchanged. The first two concerns—that the productivity boom is smaller than managers believe—are more fundamental: if the true $\xi$ is lower, then the output shares, GDP growth paths, and asset-pricing implications derived below are all correspondingly smaller. 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 $\xi$, $r_a$, and the long-run boom probability $p_*$. Using (1):

$$\frac{Y_a}{Y_a + Y_n} = \frac{z_a^{1-\alpha}(K_a^*)^\alpha}{z_a^{1-\alpha}(K_a^*)^\alpha + z_n^{1-\alpha}(K_n^*)^\alpha}. \tag{19}$$

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.

**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 | 3.3 | — | — | 164 |
| Moderate | 7.8 | 4.9 | 0.8 | 431 |
| Transformative | 17.2 | 16.8 | 2.6 | 1,055 |
| Singularity | 33.6 | 45.7 | 6.5 | 2,583 |

*Notes:* $\alpha = 1/3$, $\delta_a = 0.25$, $\delta_n = 0.07$, $r_n = 0.05$, $r_a = 0.15$, $p_* = 0.06$. 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). Post-window AI investment is independent of $r_a$; at $r_a = 0.10$, output shares are modestly lower (e.g., 6.9% vs. 7.8% for the moderate scenario). 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 4).

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

# 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 $Y_{n0}$. Cumulative GDP growth from AI is $(Y_{at} + Y_{n0})/\text{GDP}_0 - 1$, where $\text{GDP}_0 = Y_{a0} + Y_{n0}$ is pre-boom GDP.

**The window tree (2028–2030).** In 2028, the window opens: $p_{2028} = p_{\max}$, meaning a boom may occur at 2029. Investment $I_{2028}$ determines $K_{2029}$, but the 2029 boom has not yet been observed. The firm targets the *expected* optimal capital stock, averaging over the two

possible 2029 outcomes:

$$K^*_{2029} = (1 - p_{\max})\, K^*(z_1, p_{\max}) + p_{\max}\, K^*(z_1 e^{\xi}, p_{\max}).$$

Here $K^*(z, p_{\max})$ is the frictionless target for productivity $z$ when the *next* boom probability is still $p_{\max}$ (the window extends to 2030). The probability weight $p_{\max}$ outside $K^*(\cdot)$ is the chance of a 2029 boom; the $p_{\max}$ inside captures the option value of the 2030 boom. In 2029, the boom outcome is observed and the firm re-targets accordingly. From 2030 onward, the window has closed ($p \to p_*$), and there are four deterministic convergence paths—one for each scenario—each approaching $K^*(z, p_*)$.

**GDP during the transition.** Figure 3 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 17%. In the singularity scenario, two additional booms compound to produce cumulative GDP growth exceeding 40%. Post-window AI investment is approximately \$430B/yr (moderate), \$1,050B/yr (transformative), and \$2,580B/yr (singularity); the last of these would exceed half of current U.S. gross private fixed investment.

**Simulated paths (2030–2050).** Figure 4 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 $p_* = 0.06$ per year. Each boom multiplies $z$ by $e^{\xi}$. 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

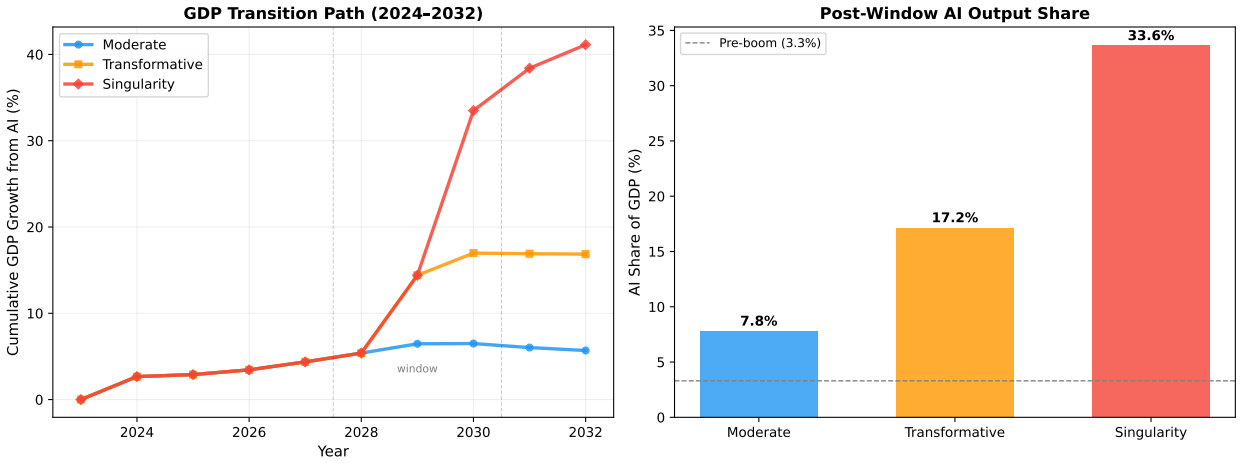



**Figure 3:** GDP transition paths. Left panel: cumulative GDP growth from AI during the transition (2024–2032), using data-anchored partial adjustment ($\theta_a = 3$). Dashed lines mark the 2028–2030 window. Right panel: post-window AI output share of GDP; the dashed line shows the pre-boom share (3.3%).

twenty-year horizon.

The compounding effect of post-window booms is first order. In the moderate scenario, expected cumulative GDP growth reaches approximately 17% by 2040 and 43% by 2050, with the AI sector's expected output share rising to roughly 24%. In the singularity scenario, expected cumulative GDP growth exceeds 117% by 2040 and 276% by 2050, with the AI sector accounting for nearly 58% of output. The bottom row of the figure provides a detailed view of the moderate scenario, which may be the most policy-relevant benchmark.

# 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

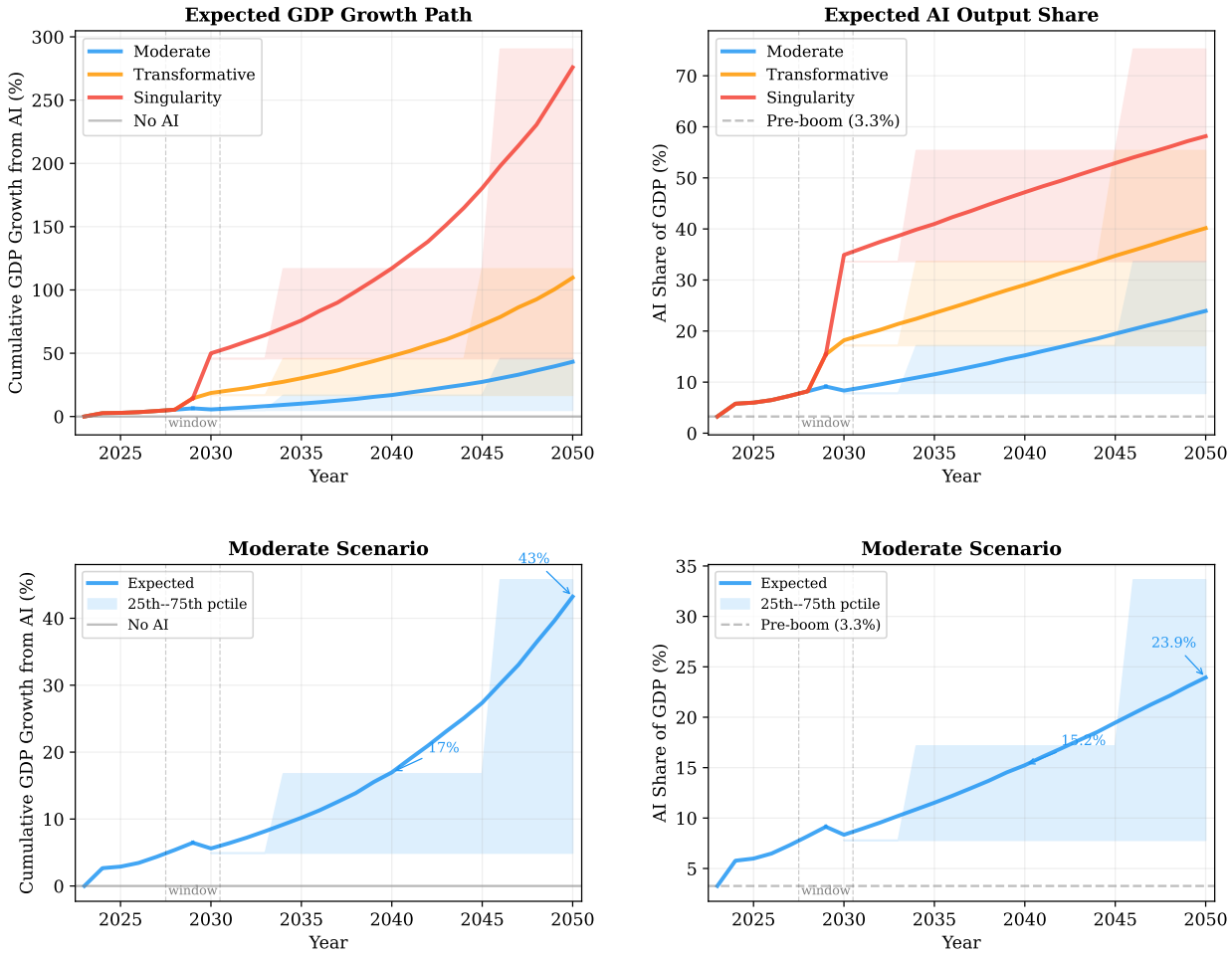



**Figure 4:** Expected GDP growth (2024–2050). Top row: all three scenarios. Bottom row: moderate-only snapshot with annotations. Solid lines show the expected path; shaded bands show the 25th–75th percentile range across 10,000 simulations. Gray lines show the no-AI baseline. The transition period (2024–2030) uses adjustment-cost capital paths ($\theta_a = 3$); from 2031 onward, booms arrive i.i.d. at rate $p_* = 0.06$ with frictionless capital adjustment.

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 $\gamma$, and discount factor $\beta$. 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)

$$M_{t+1} = \beta \, \mathbb{E}_t \left[ \left( \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \right]^{-1} \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma}. \tag{20}$$

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

Aggregate consumption growth is

$$\Delta c_{t+1} = \mu + \sigma \, \varepsilon_{t+1} + x_{d,t+1} + x_{b,t+1}, \tag{21}$$

where $\mu$ and $\sigma$ 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 (21) requires the jump $\xi_b$ in aggregate *consumption*, not output. Mapping the AI-sector productivity boom $\xi$ 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 $e^{\xi}$, 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.[8]

[8]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

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

can be equivalently written 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

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

$$\xi_b \equiv s_a \cdot (1-\alpha)\,\xi, \tag{22}$$

where $s_a$ is the share of the AI sector under various scenarios described above. This equation is an approximation even when the AI sector is in fact frozen in time.

The second version takes (22) 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 $\xi$. 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 established 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, $R_{W,t+1} = \beta^{-1}\, C_{t+1}/C_t$.

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.

Substituting (20) into the Euler equation and defining

$$\Phi_d(u) = (1 - p_d) + p_d\, e^{u\,\xi_d}, \qquad \Phi_b(u) = (1 - p_b) + p_b\, e^{u\,\xi_b}. \tag{23}$$

implies

$$r_f = \rho + \mu + \frac{1 - 2\gamma}{2}\,\sigma^2 + \log\frac{\Phi_d(1-\gamma)}{\Phi_d(-\gamma)} + \log\frac{\Phi_b(1-\gamma)}{\Phi_b(-\gamma)}, \tag{24}$$

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

Evaluating $E_t[R_{W,t+1}]$ directly and subtracting $r_f$ yields the equity premium:

$$\mathbb{E}[R_W] - r_f = \gamma\,\sigma^2 + \log\frac{\Phi_d(1)\,\Phi_d(-\gamma)}{\Phi_d(1-\gamma)} + \log\frac{\Phi_b(1)\,\Phi_b(-\gamma)}{\Phi_b(1-\gamma)}. \tag{25}$$

For small $p_d$ and $p_b$, the equity premium can be approximated as:

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

While we use (25) for all calculations below, the approximation highlights the source of risk. Both disasters and booms raise the equity premium: disasters because they reduce consumption when marginal utility is high; booms because, with $\gamma > 1$, the agent demands compensation for the consumption risk associated with the upside jump.

Table 6 reports the risk-free rate and equity premium across $\gamma$ for the pre-AI economy and for the two bounding calibrations, with $p_b = p_* = 0.06$ throughout.

Several features stand out. First, the range between panels is wide, reflecting the difficulty 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 5 bp (moderate, Panel A) to 430 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

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

*Panel A: Post-window output shares*

| | No AI ($p_b = 0$) | | Moderate ($\xi_b = 0.047$) | | Transformative ($\xi_b = 0.103$) | | Singularity ($\xi_b = 0.202$) | |
|---|---|---|---|---|---|---|---|---|
| $\gamma$ | $r_f$ | ERP | $r_f$ | ERP | $r_f$ | ERP | $r_f$ | ERP |
| 1 | 3.90 | 0.12 | 4.17 | 0.13 | 4.49 | 0.18 | 5.00 | 0.35 |
| 2 | 3.77 | 0.25 | 4.03 | 0.27 | 4.30 | 0.36 | 4.68 | 0.67 |
| 3 | 3.63 | 0.39 | 3.88 | 0.43 | 4.11 | 0.56 | 4.38 | 0.97 |
| 4 | 3.47 | 0.55 | 3.71 | 0.60 | 3.91 | 0.76 | 4.09 | 1.26 |
| 5 | 3.29 | 0.72 | 3.52 | 0.78 | 3.69 | 0.98 | 3.80 | 1.55 |
| 6 | 3.10 | 0.92 | 3.32 | 0.98 | 3.46 | 1.21 | 3.52 | 1.83 |

*Panel B: Absorbing limit $s_a \to 1$*

| | No AI ($p_b = 0$) | | All scenarios ($\xi_b = 0.597$) | |
|---|---|---|---|---|
| $\gamma$ | $r_f$ | ERP | $r_f$ | ERP |
| 1 | 3.90 | 0.12 | 6.63 | 2.17 |
| 2 | 3.77 | 0.25 | 5.31 | 3.49 |
| 3 | 3.63 | 0.39 | 4.48 | 4.31 |
| 4 | 3.47 | 0.55 | 3.94 | 4.85 |
| 5 | 3.29 | 0.72 | 3.56 | 5.24 |
| 6 | 3.10 | 0.92 | 3.25 | 5.55 |

*Notes:* $\rho = 0.02$, $\mu = 0.025$, $\sigma = 0.02$, $p_d = 0.036$, $\xi_d = -0.15$. All post-window columns use $p_b = p_* = 0.06$. 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 $\xi_b = (1-\alpha)\xi = 0.597$, 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 $\gamma$.

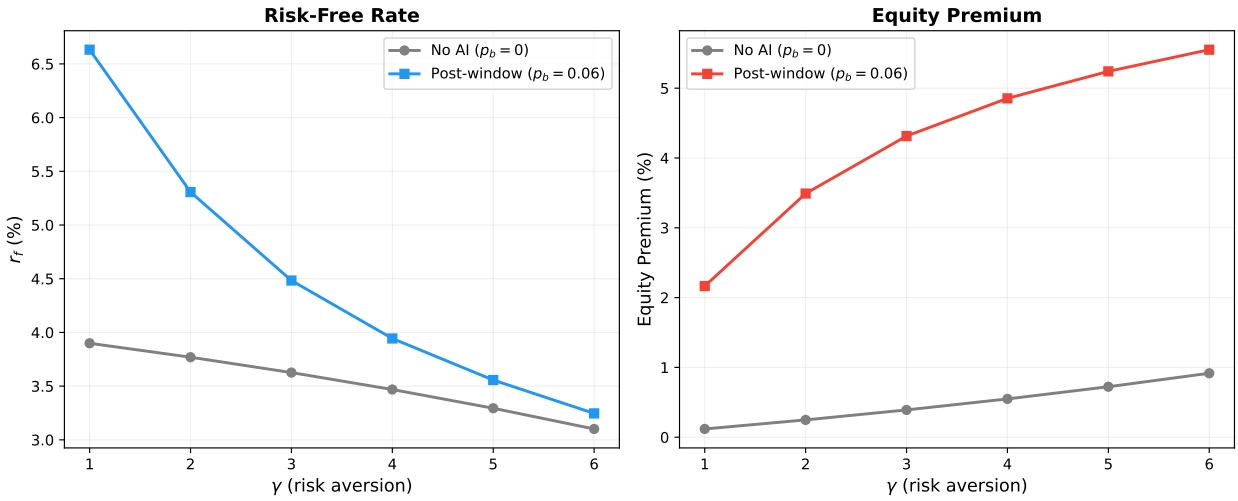



**Figure 5:** Equilibrium rates of return ($\psi = 1$). Left panel: risk-free rate. Right panel: equity risk premium on the claim to aggregate consumption. Gray line shows the pre-AI economy ($p_b = 0$); colored lines show the three post-window scenarios with $p_b = p_* = 0.06$ and scenario-specific $\xi_b$ (Panel A of Table 6); dashed black line shows the absorbing limit $\xi_b = 0.597$ (Panel B). Other parameters: $\rho = 0.02$, $\mu = 0.025$, $\sigma = 0.02$, $p_d = 0.036$, $\xi_d = -0.15$.

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 roughly 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 $\gamma$ 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.

**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 (moderate) | — | — | 2.5 | 0.90 |
| AI (transformative) | — | — | 6.0 | 1.79 |
| AI (singularity) | — | — | 14.7 | 2.69 |

*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. AI multipliers are cumulative productivity gains from the initial boom plus zero, one, or two additional booms at $\xi = 0.90$ each (the rounded calibrated value from Section 3.3). The AI multipliers apply to the AI sector only, not to aggregate GDP.

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), 17% (transformative), and 46% (singularity) by

the end of the decade, together with long-run AI investment of $431B, $1,055B, and $2,583B, per year respectively.

These cumulative figures reflect only the near-term transition through the end of the decade. In the longer run, the AI sector's share of the economy continues to grow as further booms arrive at rate $p_*$. As the AI share rises toward one, aggregate GDP growth converges to the AI sector's expected growth rate of approximately 8% per year (Section 5). Over a multi-decade horizon, the cumulative effects would therefore be far larger than the near-term numbers suggest. Because these long-run implications depend on $p_*$, which is less precisely identified than the near-term transition, we focus the historical comparison on the better-disciplined short-term numbers.

Even on this conservative basis, the AI boom stands out. The First Industrial Revolution raised British GDP per capita by roughly 70% (1.7) over 80 years; the Second Industrial Revolution produced a 140% increase (2.4) over 50 years. The near-term aggregate GDP effects of the AI scenarios (5–46%) are smaller in cumulative magnitude, but they unfold over fewer than ten years rather than half a century or more. The singularity scenario (46%) achieves in roughly five years about two-thirds of what the First Industrial Revolution delivered in 80.

The East Asian growth miracles represent the most dramatic sustained growth episodes in the historical record. For example, GDP per capita increased by a factor of 10 over 30 years. In the rare-events framework, speed is captured by the boom probability: $p_{\max} = 0.50$ implies far more compressed dynamics than the 3–5% annual probability that would characterize a boom unfolding over decades.

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. The AI scenarios in this paper are sector-specific, but the history of general-purpose technologies suggests that cross-sector spillovers

are the norm rather than the exception. If AI follows the pattern of electrification—which took roughly 40 years to diffuse from the power sector to manufacturing and services (David and Wright, 2003)—the aggregate effects could ultimately dwarf the sector-specific calculations presented here.

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.[9] 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 $\xi$ 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 $\xi$ 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.

[9]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 $164 billion in 2022 to a forecast $660–710 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 2.5. 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, cumulative GDP growth from AI reaches approximately 6% during the transition (Figure 3), more than doubling normal-times GDP growth over this period. In the transformative and singularity scenarios, cumulative GDP growth rises to 17% and over 40%, respectively.

Crucially, 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 30% depending on the scenario. As this share grows, the AI sector's rapid productivity gains increasingly dominate aggregate GDP growth—a compositional effect, not a broad-based productivity miracle.

With risk aversion of 3, and an EIS of one expected annual growth reaches approximately 8%, the risk-free rate increases by about 1 percentage point, and the equity premium rises by 4 percentage points. The AI sector's earnings load on the boom, and its risk premium rises with risk aversion, while higher expected growth compresses sovereign default premia.

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.

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*.

Chow, T., Halperin, B., and Mazlish, J. Z. (2025). Transformative AI, existential risk, and real interest rates. *Working Paper*.

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.

Goldman Sachs Global Investment Research (2025). AI capital expenditure outlook.

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.

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. (2026). 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*.

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.

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 Depreciation rates

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

### A.1.1 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.1.2 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 $\delta_a$; what matters is that $\delta_a$ is substantially larger than $\delta_n$, so that maintaining a given capital stock requires a much higher ratio of investment to capital in the AI sector.

### A.1.3 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 ($164B; see Table 1) serves as the pre-boom investment level, giving $K_{a0} = 164/0.25 = \$656\text{B}$. We use 2022 because it is the last full year before the structural break: combined capital expenditure was essentially unchanged between 2022 ($164B) and 2023 ($164B), 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 $4{,}700 - 164 = \$4{,}536\text{B}$. The non-AI capital stock is then $K_{n0} = 4{,}536/0.07 \approx \$65{,}000\text{B}$.

The resulting AI share of the total capital stock is $656/(656+65{,}000) = 1.0\%$—well below its 3.5% 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.2 AI expectations data

The model assigns a per-year boom probability of $p_{\max} = 0.50$ during the 2028–2030 window. This is not an estimate of the probability that a specific AI milestone (such as AGI) will be achieved; rather, it is a deliberately agnostic choice that maximizes uncertainty over whether additional booms occur. A Bernoulli random variable has maximal variance at $p = 0.50$, so this choice imposes the least informative prior on the boom outcome in any given year. The three scenarios then follow mechanically from two independent draws: no additional booms (probability 0.25), one additional boom (0.50), or two additional booms (0.25).

Existing forecasts of transformative AI are consistent with very large uncertainty, which is the key input to this calibration. Grace et al. (2024) survey 2,778 AI researchers and report 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 of individual responses spans decades, reflecting deep disagreement. Metaculus community forecasts, aggregating approximately 1,700 forecasters, place the median date for AGI at approximately 2028, with a 25th–75th percentile range from late 2026 to 2031.[10]

Two features of these forecasts are relevant. First, the range of expert opinion is enormous: reasonable forecasters assign anywhere from under 10% to over 80% probability to transformative AI arriving within five years. Our $p_{\max} = 0.50$ does not take a stand within this range; it simply acknowledges the uncertainty. Second, the forecasts have been mov-

[10] Metaculus, "Date of Artificial General Intelligence," https://www.metaculus.com/questions/5121/. Accessed March 2026. The Metaculus definition of AGI includes robotic capabilities, which may lag purely cognitive benchmarks.

ing rapidly—the Metaculus median has fallen from roughly 2060 to 2028 in the span of four years—suggesting that any point estimate is fragile. A maximum-entropy calibration is therefore more robust than one anchored to a specific survey or prediction market.

# B Frictionless investment paths

Figure 6 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.

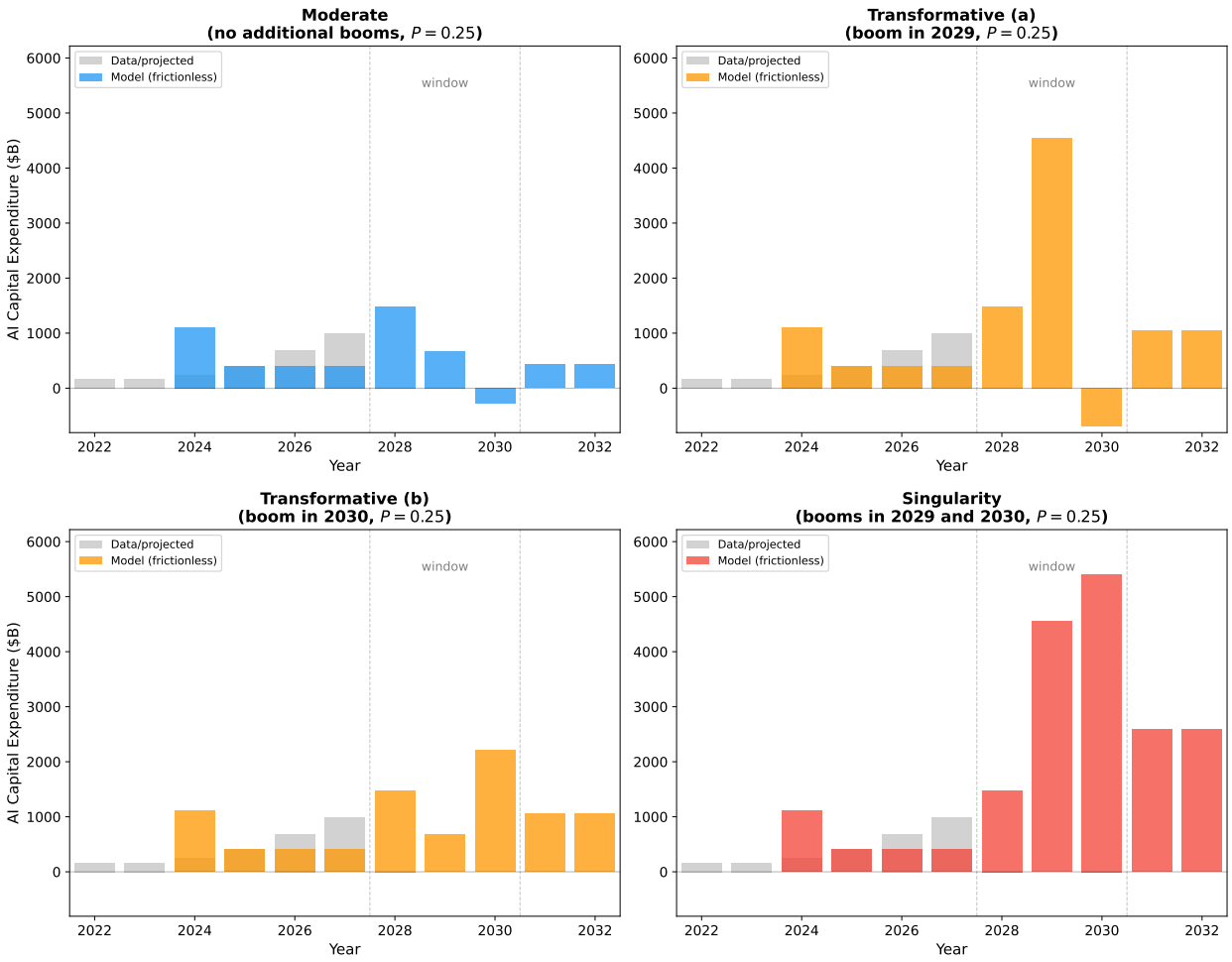



**Figure 6:** Frictionless investment paths by scenario. Gray bars show actual and projected AI capital expenditure through 2027. Colored bars show the model-implied frictionless investment. Dashed lines delineate the high-probability window (2028–2030).