Super-Exponential Growth in the Agentic Era

Machine-built capital, construction lead times and the conditions for accelerating growth

Josu San Martin

Working paper, November 2026

Abstract

We split world output into human GDP, agent output and robot output and model the physical capital under them: fab capacity, robots and power, each built first by people and then by machines. Three feedback loops act on that capital. Model improvement makes existing compute and robots more productive; self-replication lets machine-built capital build more of itself; production learning cuts the cost of machine-built capital by a learning rate LR for each doubling of cumulative output. We derive when these loops give exponential growth, accelerating growth or a finite-time singularity, and test how construction lead times, energy and diminishing returns change the answer. Learning on the robot loop, whose returns to scale are linear, gives a finite blow-up date, mid-2037 in the baseline, but only if machine-built capital is delivered instantly. With a one-year construction lead time the singularity disappears and growth keeps accelerating instead: 5% in 2030, 89% in 2035, 910% in 2040. With no learning, self-replication plus model improvement still accelerates growth, but slowly, 67% in 2035 and 400% in 2045. With power still built by people, growth settles near 180% a year, explosive but exponential; trend assumptions alone, with no machine-built capital, reach 30% growth by mid-2034 and settle near 155%. The datacentre capacity path tracked by SemiAnalysis and Patel (2026), 30 to 50 GW a year of additions, is built into the human-built power term; the regime dates barely depend on it. These are growth rates of physical capacity at 2026 prices; GDP records them only if machine output substitutes for human output, and a fixed factor would end the acceleration (Section 7). Every parameter maps to an observable and the code is open.

Keywords: economic growth, artificial intelligence, automation, machine-built capital, Wright's law, delay-differential equations. **JEL:** O33, O41, E22.

1 Introduction

Agents and robots are growing 3× to 7× a year while world GDP grows 3%. This paper asks what happens when those machines start building the physical capital they run on.

The gap is visible in the data. Google's monthly token volume rose sevenfold in a year; the length of task a frontier agent can finish alone doubles about every seven months; humanoid shipments are tripling; the datacentre capacity due in 2027 equals everything built through 2024. Yet all AI revenue is about 0.1% of world GDP, and the IMF still projects 3.1% to 3.4% growth.

Our claim is that the binding constraint on agents is compute, and the binding constraint on compute is construction. Demand for agent work already exceeds inference capacity, so agent output tracks the compute stock. Compute comes from fabs, and fabs and power plants are built by people at a pace that caps compute growth at an exponential rate. The transformative event is not smarter models but machines taking over that construction, because machine-built capital can expand without a capacity limit and gets cheaper as cumulative output grows. What it cannot do is skip the time a fab or a power plant takes to build, and that lead time decides whether there is a singularity.

The paper contributes a physical production model with that transition made explicit: fab capacity, robots and power are separate stocks with a construction lead time; who builds them decides the regime; and the three loops are switched on one at a time, so each one's contribution to the growth rate is visible. The dates are conditional on stated parameters, each tied to an observable, and the code is open.

The baseline is far more aggressive than Acemoglu's (2025) 0.66% TFP gain over a decade or Jones and Tonetti's (2026) 4% output gain by 2040. We present it as the agentic thesis taken literally, and we say which measurements would break it.

2 Related literature

Growth theory agrees on the mechanism and disagrees on the timing: once AI makes labor accumulable like capital, standard models predict accelerating growth, and the dispute is over which bottleneck binds and for how long.

Automation as accumulable labor. Aghion, Jones and Jones (2019) treat AI as the latest stage of automation; automating idea production can give a finite-time singularity unless Baumol tasks hold it back. Trammell and Korinek (2023, rev. 2026) separate self-replicating capital from self-improving capital and expect both. Kremer (1993) and Roodman (2020) show super-exponential growth was the historical norm before the demographic transition; Davidson (2021) argues AI could restart it and puts at least 10% on 30%-a-year growth this century.

Feedback-loop conditions. Erdil and Besiroglu (2023) find no decisive objection to explosive growth. Epoch's GATE couples compute, automation and semi-endogenous growth and gets explosive growth near full automation even with frictions. Davidson, Halperin, Houlden and Korinek (2026) derive a clean condition on labor share, research returns and automation share; their calibration needs only about 13% of tasks automated, with hardware research worth ten times final goods. Cunningham et al. (2026) find today's loops just below self-sustaining. Ord (2026) shows a true singularity needs the loop's generation time to shrink toward zero; otherwise growth is super-exponential without a vertical asymptote.

Physical automation. Davidson and Hadshar (2025) describe an industrial explosion in three stages (AI-directed human labor, robot factories, nanotechnology) that starts after the intelligence explosion and runs slower, with initial robot doubling times of one to four years.

Skeptics. Nordhaus (2021) proposes tests for a singularity and finds it not near. Acemoglu (2025) gets at most 0.66% TFP over ten years from task-level savings. Jones and Tonetti (2026) find tasks are complements, so output is only 4% higher by 2040 and 19% by 2060 even as automation accelerates. Catalini and Hui (2026) name verification as the weak link.

This paper differs in what it counts. GATE and Davidson et al. are compute-centric but model AI as a task share; we model the physical supply chain (fabs, robots, power) as stocks with explicit builders and a lead time. That makes the transition from human-built to machine-built capital an observable, and it separates three loops the growth literature usually bundles: model improvement, self-replication and learning. Ord's generation-time argument then appears as a parameter, the construction lead time τ.

3 Data and stylized facts

Six series anchor the calibration; the machine ones grow one to two orders of magnitude faster than the human ones.

Human GDP. The IMF's October 2026 outlook puts world growth at 3.5% in 2025, 3.1% in 2026 and 3.4% in 2027. World GDP is $124 trillion to $126 trillion in 2026; we use $125 trillion in constant 2026 dollars.

Compute. AI computing capacity grew 3.3× a year from 2022 to 2025. Delivered capacity was 20.3 million H100-equivalents at the end of 2025 and about 31.6 million by September 2026, a 2026 pace near 1.8×. The four largest hyperscalers plan about $725 billion of 2026 capex, roughly 0.6% of world GDP; with other builders we put total 2026 AI capital spending at 0.8% of GDP, the value that reproduces the September 2026 compute figure.

Datacentre capacity and capex. SemiAnalysis (2026a) expects 25 GW of net datacentre IT capacity to enter service in 2026 and 51 GW in 2027, more than the whole base built through 2024, with over 60

GW under construction and close to $1 trillion of binding leases signed in 2026 at $30 to $50 billion per GW over 15 to 20 years. Patel (2026) puts AI compute additions at about 30 GW in 2026, 50 in 2027, 70 in 2028 and 90 to 100 in 2029, near 200 GW cumulative by the end of 2028, with OpenAI and Anthropic each passing 5 GW in 2026 and the two labs taking 70 to 80% of additions by 2028. Hyperscaler capex is just over $1 trillion in 2026 and above $2 trillion in 2028, about $11 trillion over 2024 to 2029 with $5 trillion of it debt. Build times are falling: 8 to 12 months for a greenfield datacentre, 9 months at the fastest developers and 91 days for a 210 MW xAI hall, against 12 to 36 months for gas turbines and about five years for grid interconnection (SemiAnalysis, 2025, 2026c).

Agents. Google processed 9.7 trillion tokens a month in 2024, 480 trillion in 2025 and over 3.2 quadrillion in May 2026. Frontier-lab revenue run-rates were about $9 billion (Anthropic) and $20 billion (OpenAI) entering 2026, and $65 billion and $50 billion by the third quarter. We anchor agent output at $50 billion on 1 January 2026 and use the later figures only as a check. Tokens rising 7× against compute rising 2× to 3× implies roughly 3× a year of algorithmic efficiency, consistent with Gundlach et al. (2025). Patel (2026) reports Anthropic's revenue per MW of inference compute rising from about $20 million to $100 million during 2026 against a compute cost of $10 to $15 million per MW-year, and the labs now putting about 60% of compute into training.

Robots. The IFR counts 603,307 industrial installations in 2025 and a stock of 5.08 million, with 655,000 installations forecast for 2026 and 806,000 for 2029. Humanoids shipped 7,000 to 18,000 units in 2025, and about 50,000 are expected in 2026. Bank of America's March 2026 path is 90,000 in 2026, 290,000 in 2027, 500,000 in 2028, 800,000 in 2029 and 1.2 million in 2030; we use it.

Energy. Data centres used about 415 TWh in 2024, 1.5% of world electricity, and the IEA projects about 945 TWh by 2030, a figure the capacity path above would pass in 2027 if run around the clock. SemiAnalysis expects grid headroom to turn negative at the end of 2027, with behind-the-meter gas at $1.5 to $2 billion per GW filling the gap, and counts 20 to 30 GW of the 62 GW of US capacity due in 2028 in states where moratoria are on the ballot (SemiAnalysis, 2026b).

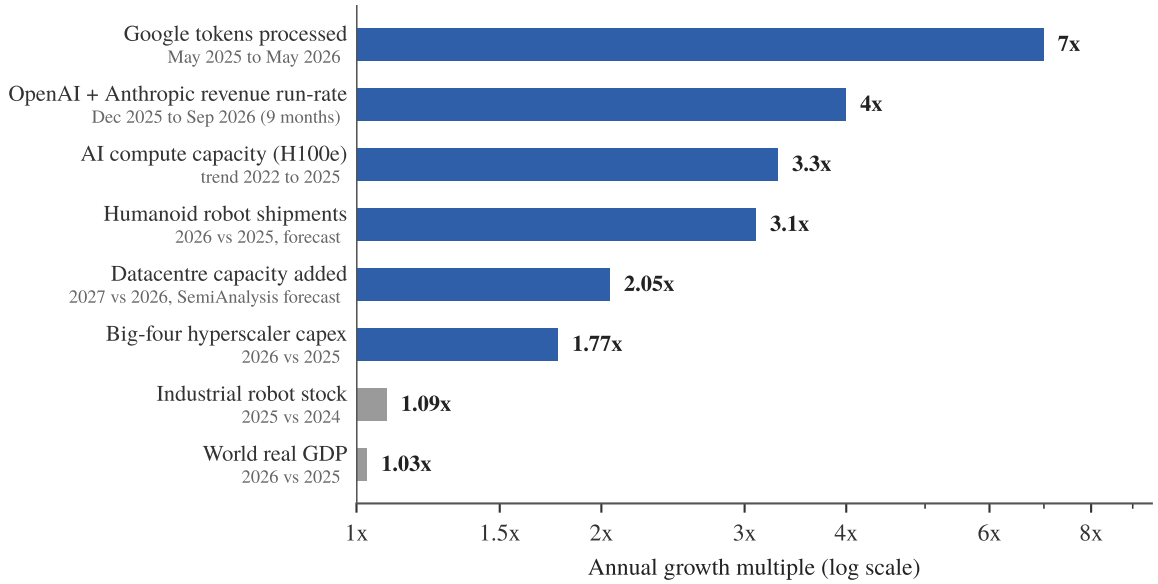


Figure 1: Growth multiples of AI inputs against the legacy economy, 2025 to 2027. Blue: AI and agent inputs; grey: the legacy economy. The compute bar is the 2022 to 2025 trend; delivered capacity grew about 1.8× annualized in 2026. The capacity bar is SemiAnalysis's forecast of 25 GW of net datacentre additions in 2026 and 51 GW in 2027. Sources: IMF, IFR, Epoch AI, Google, Counterpoint, SemiAnalysis, company disclosures.

The industrial robot fleet grows 9% a year; every AI-linked series grows 1.8× to 7×. The model asks what happens when the fast series start producing the slow ones.

4 Model

Three output blocks sit on four physical stocks. Who builds the stocks, and how long building takes, decide the growth regime.

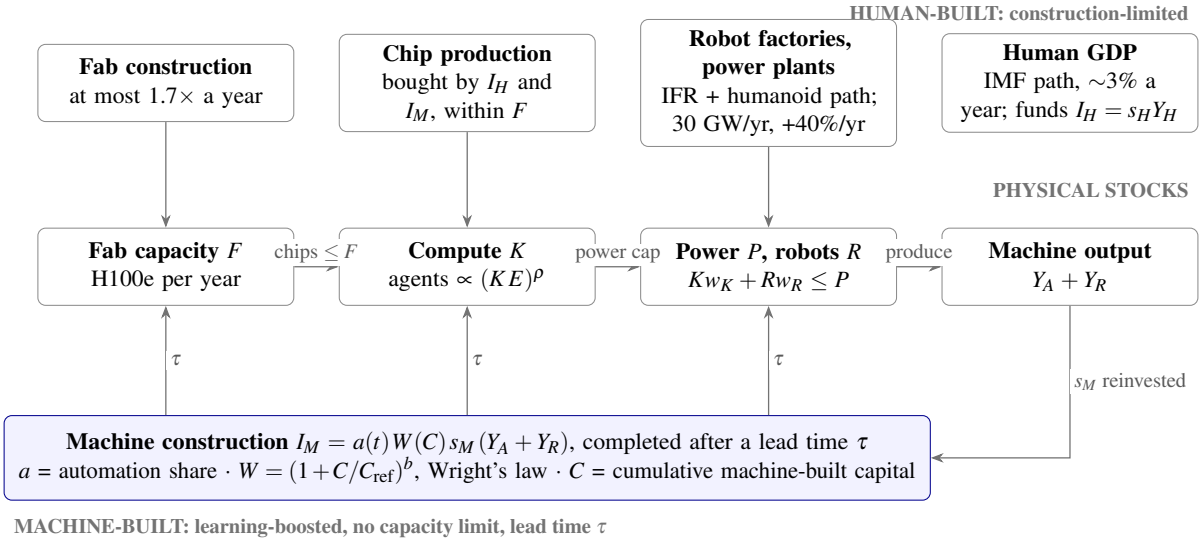


Figure 2: Model structure: three builders, four physical stocks, one learning loop. Who builds the capital, and how long it takes to build, decides the growth regime.

While people build fabs, robot factories and power plants, each stock grows at a construction-limited rate. Once machines take over, a share of machine labor goes into capital with no capacity limit and with learning, but each unit still takes τ years to complete.

4.1 Output index

All quantities are in constant 2026 dollars; $t = 0$ is 1 January 2026. Y is an output index, not GDP: machine output is valued at 2026 prices and added to human GDP, with a displacement variant in which each machine dollar removes a fraction of a human dollar. Section 7 works out what a price-weighted index would record instead.

$$Y = Y_H + Y_A + Y_R, \quad Y_A = Y_{A,0}\left(\frac{KE}{K_0}\right)^{\rho}, \qquad Y_R = v_R R E^{\Psi}, \quad E = e^{g_E t} \tag{1}$$

Human GDP follows the IMF path from a base net of 2026 machine output. Agent output is compute-bound: every H100e is used, effective labor is compute times algorithmic efficiency E, and $\rho < 1$ because the marginal agent is worth less than the last once verification and attention bind. Robot output is the stock times a per-robot value that rises with model quality.

4.2 Construction: human-built, then machine-built

Construction has two sources. Human-built investment I_H is a share s_H of human GDP plus the non-automated share of reinvested machine output; it buys chips and fab capacity in shares $h_K = 71\%$ and $h_F = 29\%$, and it is limited by construction capacity, so fab capacity can grow at most g_F a year. Human-built robots follow the IFR and Bank of America paths, $R_H(t)$, and human-built AI power follows the SemiAnalysis and Patel additions path $A_H(t) = A_0 e^{g_A t}$. Machine-built construction I_M is physical, not financial: a share s_M of machine labor (agent-hours and robot-hours, valued at 2026 prices) is redirected to building capital, and $W(C)$ is how much more capital that labor delivers as experience accumulates.

$$I_H = s_H Y_H + (1-a)\, s_M (Y_A + Y_R), \quad I_M = a(t)\, W(C)\, s_M (Y_A + Y_R),$$
$$W(C) = \left(1 + \frac{C}{C_{\text{ref}}}\right)^b, \quad b = -\log_2(1 - LR), \tag{2}$$
$$\dot{C} = I_M(t-\tau)$$

C is cumulative completed machine-built capital in 2026 dollars, starting at zero; each doubling of C beyond $C_{\text{ref}} = \$100$ billion cuts the labor per unit by LR, as it has for aircraft, solar and batteries. Capital ordered at t is completed at $t+\tau$, so every machine-built flow below uses $I_M(t-\tau)$; the human-built lead times are already inside g_F and A_H. Completed machine-built capital is split s_K, s_F, s_R, s_P = 50, 20, 15, 15% across chips, fabs, robots and power.

$$\dot{F} = \min\left(\tfrac{h_F I_H}{p_F}, (g_F + \delta_F)F\right) + \tfrac{s_F I_M(t-\tau)}{p_F} - \delta_F F,$$
$$\dot{K} = \min\left(\tfrac{h_K I_H + s_K I_M(t-\tau)}{p_C}, F\right) - \delta_K K \tag{3}$$

$$\dot{R} = R_H(t) + \tfrac{s_R I_M(t-\tau)}{p_R} - \delta_R R,$$
$$\dot{P} = A_H(t) + m_P\, \tfrac{s_P I_M(t-\tau)}{p_P}, \quad K w_K(t) + R w_R \le P \tag{4}$$

Chips are produced only within fab capacity, whoever paid for them, so automation raises compute through F rather than around it. Power is rationed when it binds: once $Kw_K + Rw_R$ reaches P, compute grows no faster than P plus the efficiency trend g_w, and robots no faster than P. The flag m_P says whether machines may build power plants. The automation share $a(t)$ is a logistic with midpoint T_a and width w.

4.3 Growth regimes

Each machine-built loop has the reduced form $\dot{X} = A(t)\, X(t-\tau)^{\varepsilon}$, where X is the stock that funds its own construction and $A(t)$ grows at most exponentially. The returns exponent ε is what matters. For agents, X = K and $\varepsilon = \rho + b$, because agent output rises as K^{ρ} and learning adds C^b with C tracking K. For robots, X = R and ε = 1 + b, because robot output is linear in R. With $\tau = 0$ there are three regimes. $\varepsilon < 1$ gives exponential growth at rate $g_A/(1-\varepsilon)$. $\varepsilon = 1$ gives a growth rate proportional to $A(t)$: exponential if A is constant, and super-exponential without a finite blow-up if A rises, which is the LR = 0 case, where model improvement E^{ψ} keeps raising robot productivity. $\varepsilon > 1$ gives a finite-time blow-up even with A constant. At the baseline $\rho = 0.65$ and $LR = 15\%$, $b = 0.23$, so the agent loop ($\varepsilon = 0.88$) cannot blow up on its own; the robot loop ($\varepsilon = 1.23$) can, and does.

$$\dot{X} = A(t)\, X(t-\tau)^{\varepsilon}, \quad \varepsilon_A = \rho + b, \quad \varepsilon_R = 1 + b;$$
$$\tau = 0,\ \varepsilon > 1 \ \Rightarrow\ T^* < \infty; \tag{5}$$
$$\tau > 0 \ \Rightarrow\ T^* = \infty$$

With $\tau > 0$ the blow-up disappears for every ε: the stock over any interval of length τ is bounded, so the construction rate over the next interval is bounded, and X is finite at every finite date. Growth can still accelerate without limit, since each generation of capital is larger than the last, but every generation takes at least τ years, which is Ord's (2026) generation-time argument in economic form. Two conditions apply whatever ε is: if a binding input stays human-built ($m_P = 0$), growth is capped at that input's rate and is exponential; and if demand or verification caps what machine output is worth, the cap, not the loop, sets the rate.

5 Calibration

Four parameters decide the answer and none is well measured: the value elasticity ρ, the learning rate LR, the automation midpoint T_a and the construction lead time τ. The table gives each parameter a source or marks it as an assumption, and names the observable that would pin it down.

One parameter is fitted: the human investment share $s_0 = 0.8\%$ of GDP is set so that compute reaches Epoch's 31.6 million H100e in September 2026 (the model gives 31.8 million). The end-2026 agent output of $134 billion is not targeted and sits near lab run-rates of about $150 billion, a weak check because ρ and g_E are free.

Table 1: Calibration. Each parameter has a source or is marked as an assumption, with the observable that would pin it down.

Parameter	Baseline	Slow	Fast	Basis, and what would pin it down
World GDP 2026 and growth	$125T; 3.1%, 3.4%, then 3.1%	same	same	IMF October 2026
Agent output, 1 Jan 2026	$50B	same	same	Lab run-rates entering 2026, grossed up
Compute stock K_0; fab capacity F_0	20.3M H100e; 18M H100e/yr	same	same	Epoch AI deliveries
Human fab-capacity growth, g_F	1.7×/yr	1.5×	2.0×	Assumption; Patel: WFE near $200B in 2027, ~100 EUV tools a year by 2030. Observable in foundry and HBM additions
Human-built AI power additions, A_H	30 GW/yr in 2026, growing 40%/yr	same	same	SemiAnalysis: 25 GW (2026), 51 GW (2027); Patel: 30, 50, 70, 90 to 100 GW for 2026 to 2029; +40%/yr extends that path
Unit costs: chip, fab capacity, robot, power	$37k per H100e; $18k per H100e/yr; $100k per robot; $3B per GW	same	same	Blackwell-class pricing (Patel: compute costs $10 to 15M per MW-year); ~0.5× unit price for capacity; humanoid and industrial list prices; generation plus grid (SemiAnalysis: onsite gas $1.5 to 2B per GW)
Chip price-performance; compute per watt	+37%/yr; +40%/yr (kW per H100e −29%/yr)	same	same	Epoch AI trends
Algorithmic efficiency, g_E	2.5×/yr	2.0×	3.0×	Gundlach et al. (2025); tokens 7× vs compute 2× to 3×
Value elasticity, ρ	0.65	0.5	0.8	Assumption; tokens 7× vs revenue 4× suggests ~0.7, and Anthropic's revenue per MW rose 5× in 2026 (Patel), near 1 at the margin; neither is an identified estimate
Investment shares: s_H (human GDP), s_M (machine output)	0.8% rising to 2.0%; 30%	to 1.2%	to 3.5%	s_0 fitted to Epoch's September 2026 compute; s_max below Patel's capex path ($1T in 2026, over $2T in 2028); s_M an assumption
Allocation of construction: chips, fabs, robots, power	50 / 20 / 15 / 15% (machine-built); 71 / 29 chips, fabs (human-built)	same	same	Assumption
Robot stock R_0; human-built additions	5.08M; IFR path + humanoids 90k, 290k, 500k, 800k, 1.2M (2026 to 2030), then +50%/yr	same	same	IFR; Bank of America (March 2026)
Value per robot-year; capability gain ψ	$25,000; 0.3	0.2	0.4	Assumption
Automation midpoint T_a; width w	2033; 1 yr	2038; 1.5	2031; 1	Assumption; observable as the machine share of fab, datacentre and plant construction
Learning rate on machine-built capital, LR	15%	5%	25%	Wright's law range for aircraft, solar, batteries; observable as unit cost vs cumulative machine output
Construction lead time, τ	1 yr	same	same	SemiAnalysis: datacentres 8 to 12 months, turbines 12 to 36, grid 5 yrs; fabs 2 to 5 yrs. 1 yr is the datacentre figure; varied from 0 to 3 in 6.3
Machines may build power, m_P	yes	yes	yes	Switched off in the human-energy scenario

6 Results

Through 2030 nothing in the baseline depends on automation or on the lead time; after 2033 everything does. The output index is $152 trillion at the end of 2030 in every baseline-family scenario, with $6.2 trillion of machine output. Growth passes 30% a year in mid-2034 and machine output passes human GDP at the start of 2035. What happens next depends on τ: with instant construction the model blows up in mid-2037; with a one-year lead time growth is 89% in 2035, 296% in 2037 and 910% in 2040, and

still rising.

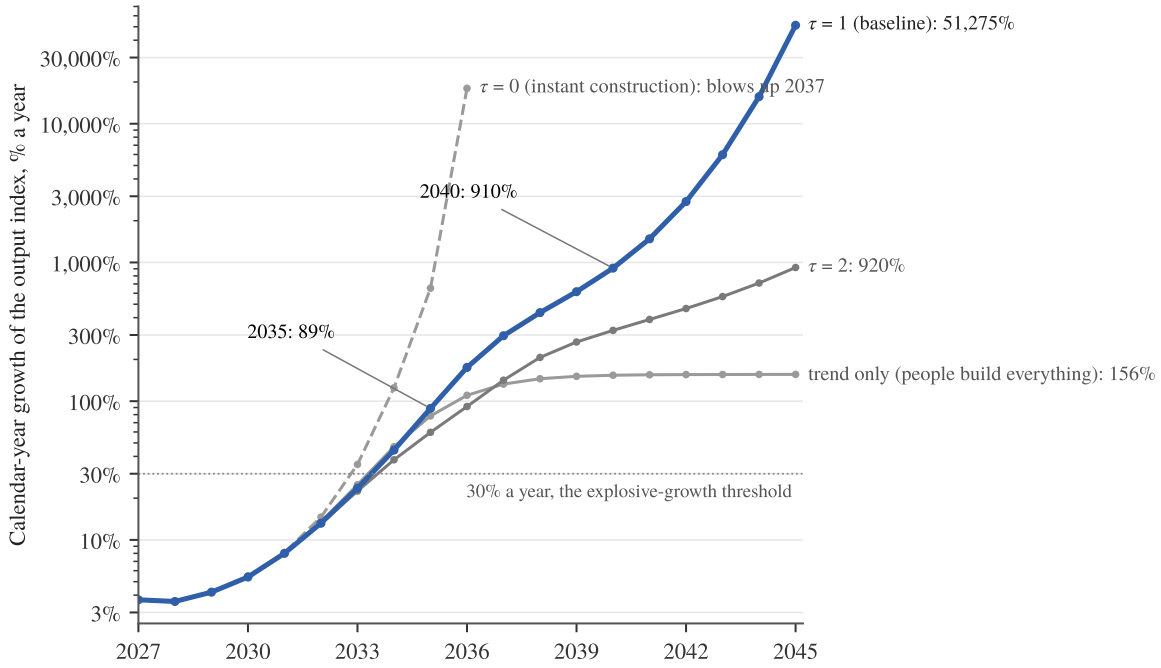


Figure 3: Calendar-year growth of the output index, 2027 to 2045, by construction lead time τ (model v3 simulation, log scale). Every other assumption is shared across the four paths. With $\tau = 0$ the model blows up in 2037; with $\tau = 1$ growth is 89% in 2035 and 910% in 2040 and keeps rising; with $\tau = 2$ it is still under 1,000% in 2045.

The four lines share every trend assumption and differ only in machine-built construction and its lead time. Trend alone reaches 30% in mid-2034 and settles near 155% a year, the exponential rate $\rho(g_E + g_F)$ that human-built fabs allow. Instant machine construction ($\tau = 0$) separates from trend in 2033 and blows up in 2037. With a one-year lead time the curve bends two years later and never goes vertical; with two years, growth is still under 1,000% in 2045.

6.1 End of 2030

Table 2: End of 2030 in the three scenarios (constant 2026 dollars; compute in H100-equivalents). Chip power is the model's demand from chips and robots; capacity is human-built AI datacentre capacity on the SemiAnalysis and Patel path.

Scenario	Output index	Human GDP	Agents	Robots	Machine share	Compute	Robots	Chip power	Capacity
Slow	\$148T	\$146T	\$0.9T	\$0.5T	0.9%	196M H100e	9.9M	50 GW	412 GW
Baseline	\$152T	\$146T	\$5.3T	\$1.0T	4.1%	268M H100e	9.9M	65 GW	414 GW
Fast	\$194T	\$146T	\$45.3T	\$2.4T	24.6%	416M H100e	10.8M	96 GW	442 GW

The last two columns are the paper's sharpest near-term tension. On the SemiAnalysis and Patel path, AI datacentre capacity reaches 414 GW by the end of 2030; the chips the model puts in it, on the Epoch trend with fab capacity growing 1.7× a year and efficiency 1.4×, draw 65 GW, about 570 TWh a year. Either chips grow faster than that (the fast scenario fills a quarter), or capacity runs ahead of chips, as SemiAnalysis expects for 2027 and 2028; section 6.3 shows the regime dates barely depend on which. Agent output per GW of chips rises from \$4 billion in 2026 to \$81 billion in 2030, reaching the \$100 billion per GW that Patel (2026) treats as the end-revenue potential of a gigawatt. The robot stock is 9.9

million against roughly 7.1 million on the IFR's industrial path alone; the difference is the humanoid forecast.

6.2 One loop at a time

Table 3: One loop at a time. Year in which growth first exceeds 30% a year, growth of the output index in 2035, 2040 and 2045, and the regime.

Scenario	What changes	>30%	2035	2040	2045	Regime
1. Trend only	people build everything ($a = 0$)	2034.3	78%	154%	156%	exponential, settles at $\rho(g_E + g_F)$
2. Self-replication alone	$a \to 1$; $LR = 0$, $\psi = 0$	2034.6	61%	223%	337%	exponential, settles at $\rho g_E/(1-\rho)$
3. Replication + model improvement	$LR = 0$	2034.5	67%	239%	397%	accelerating, no blow-up
4. Replication + learning	$\psi = 0$	2034.6	73%	490%	1,109%	accelerating, no blow-up
5. All three: baseline, $\tau = 1$	automation 2033, $LR =$ 15%	2034.4	89%	910%	51,275%	accelerating, no blow-up
6. Baseline, $\tau = 0$	instant construction	2033.9	653%	blow-up mid-2037	—	finite-time singularity
7. Baseline, $\tau = 2$	two-year lead time	2034.6	60%	324%	920%	accelerating, no blow-up
8. Human-built energy	$m_P = 0$	2034.4	89%	161%	178%	exponential, capped by power from 2039
9. Absorption-bound	machine output at most doubles yearly	2035.5	37%	94%	100%	exponential, capped by demand
10. Displacement	each machine dollar removes $0.50 of human output	2035.2	94%	908%	50,767%	as baseline
11. Slow	automation 2038, $LR =$ 5%, $\rho = 0.5$, $g_E = 2\times$	2041.8	6%	23%	78%	accelerating, slowly
12. Fast	automation 2031, $LR =$ 25%, $\rho = 0.8$, $g_E = 3\times$	2031.2	6,820%	run stopped 2039	—	accelerating, no blow-up

Rows 1 to 5 switch the loops on one at a time. Self-replication by itself is exponential: with $\rho < 1$ the agent loop converges to a constant rate, and the robot loop, with linear returns and no model improvement, replicates at a fixed rate. Model improvement makes that rate rise (row 3), learning makes it rise faster (row 4), and together they give 900% a year by 2040 (row 5). Rows 6 and 7 are the paper's main result: the same loops with no lead time blow up in 2037, and with a two-year lead time are still under 1,000% in 2045. Rows 8 and 9 show the two ways the acceleration stops: a binding input that stays human-built, or a cap on what machine output is worth. They still reach 85% and 37% growth by 2035, explosive by Davidson's standard but exponential.

6.3 What moves the dates

The lead time does not touch the near term, because the 30% and 100% dates come from the trend (human-built fabs at 1.7× and efficiency at 2.5× a year), but each year of it pushes the 1,000% date out by three to five years and removes the singularity. The learning rate works the same way: at $\tau = 1$, moving LR from 5% to 25% moves the 100% date by one year (2036.7 to 2035.7) and the 1,000% date by ten (2047.6 to 2038.1); at $LR = 0$ growth stays under 1,000% through 2048. The automation midpoint matters less: 2031 against 2038 moves the 1,000% date by 0.5 years at $LR = 15\%$ and 1.9 years at $LR = 25\%$ (at $\tau = 0$ the blow-up moves 3.2 years, 2034.0 to 2037.2). ρ moves the 100% date

Table 4: Regime dates by construction lead time τ (baseline parameters otherwise). Year in which growth first exceeds each threshold; a finite-time blow-up occurs only at $\tau = 0$.

Lead time τ (years)	Growth $> 30\%$	$> 100\%$	$> 1{,}000\%$	Blow-up
0	2033.9	2034.9	2036.2	2037.3
0.25	2034.1	2035.2	2037.4	none
0.5	2034.2	2035.6	2038.7	none
1 (baseline)	2034.4	2036.2	2041.2	none
2	2034.6	2037.2	2046.3	none
3	2034.6	2038.2	after 2048	none

from 2038.6 ($\rho = 0.5$) to 2033.7 ($\rho = 0.8$), the $\tau = 0$ blow-up from 2038.6 to 2035.8, and 2030 agent output from \$1.7 trillion to \$15 trillion. Algorithmic efficiency sets the near-term dates: 30% growth arrives in 2041 at 1.5× a year, 2037 at 2×, 2034 at 2.5× and 2033 at 3×. The near-term chip path does not: filling the SemiAnalysis and Patel capacity path (fab capacity 2.5× a year, investment 3.5% of GDP) moves the 30% date from 2034.4 to 2034.1, and a 25%-a-year power additions path, the moratoria case, changes nothing, because the power ceiling is slack until the late 2030s.

The dollar levels after growth passes 100% a year are not forecasts. By 2037 the baseline has 3.0 billion robots and draws 3.3 TW of power, about today's world average generation; the model is reporting that an input it does not track, most likely materials, land or human absorption, binds first.

7 Prices, demand and what GDP would record

The growth rates in Section 6 are growth of physical capacity, not of GDP. The output index adds machine output to human output at 2026 prices, so it assumes every unit of machine output is still sold at the price it fetched in 2026. Three objections follow from that, and they decide how the results should be read.

Demand. The model reinvests 30% of machine output (s_M). The other 70% has to be bought. Human income grows about 3% a year, so buyers can absorb a machine output that doubles every year only if its price falls about as fast. Say's law holds, but it holds through prices: the relative price of machine output has to fall towards zero.

The index. National accounts chain-weight quantities by current prices, so a good counts for less as it gets cheap. Suppose human and machine output are demanded with a constant elasticity of substitution σ. The quantity index a statistical office would then report, and the relative price behind it, are

$$G = \left[(1-\omega) \left(\frac{Y_H}{Y_{H,0}} \right)^{\frac{\sigma-1}{\sigma}} + \omega \left(\frac{Y_M}{Y_{M,0}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \qquad \frac{p_M}{p_H} \propto \left(\frac{Y_M}{Y_H} \right)^{-1/\sigma} \tag{6}$$

where Y_M is agent plus robot output and $\omega = 0.14\%$ is the machine share of spending in 2026. The table applies this to the baseline path of Section 6.

The same physical path is recorded as 3% growth or as 900%. Below $\sigma = 1$ the machine share of spending shrinks while machine output explodes, and measured growth stays at the human rate. That is the cost-disease mechanism Aghion, Jones and Jones (2019) apply to AI, and it is what happened with computing, whose price collapse never showed up as explosive GDP (Nordhaus, 2021). Above $\sigma = 1$ the cheap good takes over spending and measured growth converges on the output index a few years late.

Which σ. Earlier information technology complemented labour, which points to σ below one. The claim of this paper is that agents and robots do the tasks workers do, which is substitution and points to σ well above one for the output they replace. The paper does not estimate σ, and its title holds for GDP

Table 5: What a price-weighted quantity index records for the baseline path, by the elasticity of substitution σ between human and machine output. The last row is the fixed-price output index used in the rest of the paper.

Elasticity of substitution σ	Growth recorded in 2035	Growth recorded in 2040	Machine share of spending, end of 2040
0.5	3.1%	3.1%	under 0.1%
1	3.3%	3.5%	0.1%
1.5	4.6%	53%	23%
2	8%	416%	81%
3	18%	832%	98%
Output index (fixed 2026 prices)	89%	910%	100%

only if σ exceeds one. The observable is the machine share of spending: it rises if σ is above one and falls if it is below.

A fixed factor. If some essential input is in fixed supply, such as land, ores or energy sites, its price rises without limit as machines multiply, and each extra machine is worth less. In the model that is a returns exponent ε that falls below one at scale, which moves the economy from accelerating growth back to exponential growth (Section 4.3). The model has no such input. Its paths should not be read literally once machine output is thousands of times human output, a ratio the baseline passes during 2040.

8 Discussion

The robust result is a pair of conditions, not a date. Machine-built capital with learning on a loop with linear returns, here robots, gives a finite-time singularity only if construction is instantaneous; any positive lead time turns it into accelerating growth with no vertical asymptote. And any binding input that stays human-built, or any cap on what machine output is worth, turns accelerating growth back into exponential growth. Four things decide where the baseline sits.

The output index is not GDP. Machine output is valued at 2026 prices and added to human GDP. Section 7 shows that a price-weighted index records anything from 3% to 900% growth for the same path. Nominal revenue will grow more slowly still, because the price of a fixed level of AI performance is falling about 13× a year, and national accounts deflate quality change in digital services poorly. Displacement changes little: with half of each machine dollar displacing a human dollar, human output falls to $34 trillion by 2035 and the growth path does not move. The practical dashboard is closer to Nordhaus's (2021) tests, a rising capital share and falling relative prices of compute-intensive services, than to headline GDP.

Weak links. The additive structure lets machine output grow without human complements. If tasks are complements, as Jones and Tonetti (2026) estimate, or if human verification caps what agent output is worth (Catalini and Hui, 2026), the absorption-bound scenario is the right one: growth of 37% in 2035 and no singularity. ρ captures some of this but not a hard ceiling; a nested CES version with σ below one is the most useful extension.

Energy. On the SemiAnalysis and Patel path, human-built AI capacity reaches 414 GW by the end of 2030 and 1.5 TW by 2033, and the model's chips use a sixth of it, so the power ceiling is slack until the late 2030s, when the loop draws 3 TW. If machines never build power plants, that path binds near 2039 and growth settles around 180% a year, explosive but exponential. Until then the acceleration lives in the energy sector's learning curve and build time: 12 to 36 months for turbines and about five years for interconnection today, and 20 to 30 GW of 2028 US capacity exposed to moratoria.

Financing. Patel (2026) puts AI capex at about $11 trillion over 2024 to 2029, $5 trillion of it debt, and 100 GW a year at current prices at $5 trillion a year, $7 to $10 trillion with power and buildings,

near a tenth of world GDP by 2030. The model's human investment share peaks at 2% of GDP, below that path, and the fast scenario's 3.5% is nearer; because compute is fab-bound, the extra money would mostly bid up chips. Growth of 89% a year in 2035 implies real interest rates of the same order, which is the sovereign-debt channel raised in the same interview: capacity financed at 5 to 8% for a 3% world meets rates set by a 90% one.

Distribution. Human GDP stays on its IMF path by construction, so the human share of output falls below half at the start of 2035 while its level rises. Patel (2026) expects two labs to take 70 to 80% of compute additions by 2028, so the machine share is at first the share of two firms. Whether wages follow depends on Korinek and Suh's (2024) race between automation and capital accumulation; the model is silent on it.

What would falsify the baseline. Each key parameter has a measurement that would move it.

1. Compute shortage: GPU lead times and rental prices stay elevated if agents saturate compute. If they fall while token growth continues, demand, not compute, binds, and the absorption-bound scenario applies.
2. Automation share $a(t)$: the fraction of fab, datacentre and power-plant construction hours done by robots and agents. The baseline needs 12% by 2030.
3. Lead time τ: the time from order to first output for a machine-built fab or power plant. The baseline assumes one year against two to five today; at two years the 1,000% date moves from 2041 to 2046.
4. Learning rate LR: unit cost of machine-built fab capacity, robots and generation against cumulative machine-built output. At $LR = 5\%$ growth stays under 1,000% a year until 2047; at 25% it passes 1,000% in 2038.
5. Value elasticity ρ: frontier-lab revenue growth against token growth, re-estimated each quarter. The 2030 agent-output range of $1.7 to $15 trillion is this parameter.
6. Energy: whether human-built capacity keeps to the 30, 50, 70, 95 GW path for 2026 to 2029. The moratoria SemiAnalysis tracks are the near-term test; the model's dates do not depend on it.

9 Conclusion

Agents are compute-bound, compute is construction-bound, and as long as people do the construction, growth is exponential: fast in the baseline, near 80% a year by 2035 and 155% by 2040 on trend assumptions alone, but exponential. The threshold for something else is not general intelligence. It is the point at which machine capital reproduces the physical infrastructure it depends on, with costs that fall as the installed base grows.

That gives a testable claim, and a correction to the obvious version of it. If machine-built fabs, robots and power have a positive learning rate and no binding input stays human-built, growth accelerates without limit: 89% in 2035 and 910% in 2040 in the baseline. It reaches infinity in finite time only if that capital is built instantly; with a one-year lead time there is no singularity, and with two years growth is still below 1,000% in 2045. If there is no learning, growth still accelerates, but slowly; if energy stays human-built or demand caps what machine output is worth, growth stays exponential at 100% to 180% a year. The numbers to measure are the learning rate on self-built capital and the time it takes to build. Whether GDP records the acceleration depends on a third, the elasticity of substitution between machine and human output.

References

Literature

Acemoglu, D. (2025). The simple macroeconomics of AI. *Economic Policy* 40(121): 13–58. https://www.nber.org/papers/w32487

Aghion, P., B. F. Jones and C. I. Jones (2019). Artificial intelligence and economic growth. In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press. https://www.nber.org/papers/w23928

Catalini, C. and X. Hui (2026). Some simple economics of AGI. arXiv:2602.20946. https://arxiv.org/abs/2602.20946

Cunningham, T., L. Althoff, B. Halperin, B. Jabarian, A. Koh, A. Ramani, P. Trammell, P. Whitfill and C. Wu (2026). The economics of recursive self-improvement. arXiv:2609.15802. https://arxiv.org/abs/2609.15802

Davidson, T. (2021). Could advanced AI drive explosive economic growth? Open Philanthropy. https://coefficientgiving.org/research/could-advanced-ai-drive-explosive-economic-growth/

Davidson, T. and R. Hadshar (2025). The industrial explosion. Forethought. https://forethought.org/research/the-industrial-explosion

Davidson, T., B. Halperin, T. Houlden and A. Korinek (2026). When does automating AI research produce explosive growth? Feedback loops in innovation networks. NBER Working Paper 35155. Summary: VoxEU. https://www.nber.org/papers/w35155 https://cepr.org/voxeu/columns/ai-feedback-loops-and-conditions-explosive-growth

Epoch AI (2025). GATE: An integrated assessment model for AI automation. arXiv:2503.04941; see also AI and explosive growth redux. https://arxiv.org/abs/2503.04941

Erdil, E. and T. Besiroglu (2023). Explosive growth from AI automation: A review of the arguments. arXiv:2309.11690. https://arxiv.org/abs/2309.11690

Gundlach, H., J. Lynch, M. Mertens and N. Thompson (2025). The price of progress: Algorithmic efficiency and the falling cost of AI inference. arXiv:2511.23455. https://arxiv.org/abs/2511.23455

Jones, C. I. and C. Tonetti (2026). Past automation and future A.I.: How weak links tame the growth explosion. Working paper, Stanford GSB. https://web.stanford.edu/~chadj/JonesTonetti_Automation.pdf

Korinek, A. and D. Suh (2024). Scenarios for the transition to AGI. NBER Working Paper 32255. https://www.nber.org/papers/w32255

Kremer, M. (1993). Population growth and technological change: One million B.C. to 1990. *Quarterly Journal of Economics* 108(3): 681–716. https://gwern.net/doc/economics/1993-kremer.pdf

Kwa, T. et al. (2025). Measuring AI ability to complete long tasks. METR, arXiv:2503.14499. https://arxiv.org/abs/2503.14499

Nordhaus, W. D. (2021). Are we approaching an economic singularity? Information technology and the future of economic growth. *AEJ: Macroeconomics* 13(1): 299–332. https://www.aeaweb.org/doi/10.1257/mac.20170105

Ord, T. (2026). The dynamics of intelligence explosions. arXiv:2608.14426. https://arxiv.org/abs/2608.14426

Roodman, D. (2020). Modeling the human trajectory. Open Philanthropy. https://www.openphilanthropy.org/blog/modeling-human-trajectory

Trammell, P. and A. Korinek (2023, revised 2026). Economic growth under transformative AI. NBER Working Paper 31815. https://www.nber.org/papers/w31815

Data sources

Bank of America Institute (2026). Humanoid robot shipment forecast, March 2026.

Counterpoint Research (2026). Global humanoid robot shipments soar nearly 300% YoY in H1 2026. https://counterpointresearch.com/en/insights/global-humanoid-robot-shipments-soar-nearly-300-percent-yoy-in-h1-2026

Epoch AI (2026). Global AI computing capacity is doubling every 7 months; data centre explorer coverage and delivered capacity; trends. https://epoch.ai/data-insights/ai-chip-production https://alphasignal.ai/news/epoch-maps-44-of-global-ai-compute-across-86-data-centers https://epoch.ai/trends

Google (2026). I/O 2026 keynote figures; API tokens per minute, Q3 2026. https://blog.google/innovation-and-ai/sundar-pichai-io-2026/?ref=t01.li

https://fourweekmba.com/ai-googles-api-nears-46-trillion-tokens-a-day-up-over-4x/

IEA (2025). Energy and AI, data-centre electricity 2024 and 2030. https://www.artificialintelligence-news.com/news/iea-opportunities-and-challenges-ai-for-global-energy/

IFR (2026). World Robotics 2026: Industrial robots, executive summary. https://ifr.org/img/worldrobotics/Executive_Summary_WR_2026_-_Industrial_Robots.pdf

IMF (2026). World Economic Outlook, October 2026, via Trading Economics; world GDP levels via Bursa and India Macro Indicators. https://tradingeconomics.com/world/full-year-gdp-growth/news/591413 https://www.bursa.ro/the-worlds-largest-economies-in-2026-us-china-and-germany-maintain-their-positions-in-the-top-3-45483752 https://indiamacroindicators.co.in/resources/blogs/world-gdp-breakdown-2026-by-country-region-economic-group

Lab revenues: Anthropic run-rate; OpenAI run-rate. https://mezha.net/eng/bukvy/bdf67aa1_anthropic_surpasses_-65/ https://valueaddvc.com/blog/openai-revenue-2026-arr-path-to-profitability

NextMSC (2026). Humanoid robots: announced vs. shipped units. https://www.nextmsc.com/industry-deep-dive/humanoid-robots-announced-vs-shipped-units

OfficeChai (2026). Hyperscaler capex 2026. https://officechai.com/ai/ai-capex-spend-at-top-4-hyperscalers-to-touch-715-billion-in-2026/

Patel, D. (2026). Two labs will soon control most of the world's workforce. Interview by Dwarkesh Patel, Dwarkesh Podcast, 25 August 2026. Transcript. https://www.dwarkesh.com/p/dylan-patel-3

SemiAnalysis (2025). How AI labs are solving the power crisis: the onsite gas deep dive. 30 December 2025. Link. https://newsletter.semianalysis.com/p/how-ai-labs-are-solving-the-power

SemiAnalysis (2026a). Datacenters 2027: the year of acceleration, +50 GW in a single year. 19 October 2026. Link. https://newsletter.semianalysis.com/p/datacenters-2027-the-year-of-acceleration

SemiAnalysis (2026b). The datacenter backlash goes bipartisan: next week's midterms and the 2028 buildout. 1 November 2026. Link. https://newsletter.semianalysis.com/p/the-datacenter-backlash-goes-bipartisan

SemiAnalysis (2026c). Elon is doing it again: SpaceX will be the world's biggest AI cloud. 3 November 2026. Link. https://newsletter.semianalysis.com/p/elon-is-doing-it-again-spacex-will

TechSpot (2026). Epoch AI: cost of fixed AI performance falling ~13× a year. https://www.techspot.com/news/114055-cost-ai-performance-falling-fast-could-reshape-entire.html

A Running the model

The simulation is one Python file, `agentic_growth_v3.py`. It integrates four stocks in logs and C in levels with SciPy's `solve_ivp` (RK45); the lead time makes the system a delay-differential equation, solved by the method of steps, in chunks of length τ with the completed-capital term read from the dense output of earlier chunks. A run stops when compute has grown e^{40}-fold or its growth rate passes 50 a year, and reports a finite-time blow-up only when $\tau = 0$ and the growth rate at the stop exceeds 20 a year (a double-exponential path crosses e^{40} at about $\psi g_E \times 40 \approx 11$ a year). `BASE` holds the 2026 anchors and unit costs; `SCEN` holds ρ, g_E, g_F, s_{max}, T_a, w, LR, ψ, m_P, τ, the absorption cap and the displacement share. It writes the regime dates to `v3_summary.csv`; `v3_sens.py` produces the grids in 6.3.

Effective agent instances, used only for intuition, are $\varphi KE/c_0$ with $\varphi = 0.6$ of compute on inference and $c_0 = 0.1$ H100e per frontier agent at 2026 efficiency: about 120 million at the start of 2026 and 160 billion at the end of 2030 in the baseline.