WORKING PAPERS

N° 1722

February 2026

“Epistemic Capital and Two-Trap Growth in the AI Era”

Manh-Hung Nguyen

Toulouse School of Economics

Epistemic Capital and Two-Trap Growth in the AI Era*

Manh-Hung Nguyen†

19 February 2026

Abstract

I develop a growth model in which AI-generated content contaminates the knowledge commons, creating two nested irreversibilities. A derivative trap arises when recombinative output crosses a threshold in the corpus, degrading frontier productivity faster than talent reallocation or R&D subsidies can offset. A governance trap arises because the institutional capacity to distinguish frontier from derivative knowledge–epistemic capital–is itself a depletable stock. In the baseline simulation, the governance trap preempts the derivative trap by roughly nine years, closing the window for effective policy while measured innovation remains positive. The competitive equilibrium features a double wedge: frontier knowledge is undervalued and derivative output overvalued, driving a strict instrument hierarchy in which epistemic investment is a precondition for governance, which is a precondition for R&D subsidies. The welfare cost of inaction is 6.8% in consumption-equivalent terms.

KEYWORDS: Derivative trap, data quality, epistemic capital, governance trap, innovation policy, forward invariance.

JEL CLASSIFICATION: O31, O33, O38, D83.

*Manh-Hung Nguyen. Email: *manh-hung.nguyen@tse-fr.eu*. The author acknowledges funding from the French National Research Agency (ANR) under the Investissements d'Avenir program, grant ANR-17-EURE-0010.

†Toulouse School of Economics, INRAE, Université de Toulouse Capitole, Toulouse, France.

1 Introduction

Since Romer [1990], endogenous growth theory has treated knowledge as a net asset. Ideas accumulate, depreciate, and spill over; the stock may be underprovided, but it is not modelled as self-degrading. Recursive AI training raises that possibility. When generative models are retrained on corpora that contain their own prior output, the informational content of the training environment can deteriorate cumulatively and, past a threshold, irreversibly. Shumailov et al. [2024] document progressive distributional collapse in language models retrained on recursively generated text; Alemohammad et al. [2023] and Gerstgrasser et al. [2024] obtain analogous patterns across architectures and modalities.[1] What is missing is a macro framework that translates these regularities into growth dynamics and policy constraints.

The empirical motivation is no longer conjectural. Liang et al. [2025] study 1.1 million abstracts from arXiv, bioRxiv, and Nature portfolio journals and estimate that by September 2024 LLM-modified content accounts for 22.5% of computer-science abstracts, 18.0% in electrical engineering, and 7.7% in mathematics, up from a baseline near 2.4% (Table 1). Kobak et al. [2025] apply an independent method to 15.1 million PubMed abstracts and put the 2024 biomedical figure at 13.5%, reaching 40% in some subcorpora. Reading $1-\hat{\alpha}$ as observed data quality, the implied level drops from about 0.975 to 0.775 in computer science over two years. The time path is S-shaped, consistent with logistic diffusion, and the cross-field ranking is stable—fields with higher AI exposure contaminate faster ($a_R > a_F$ in the model). On the evaluative side, best-available AI-text detection accuracy falls from roughly 0.95 to 0.74 between 2022Q4 and 2024Q2 [Pratama, 2025]; reviewers contacted per manuscript rise from 4.8 to 6.8 across ASM microbiology journals [Tropini et al., 2023]; up to 17% of computer-science conference review sentences are themselves LLM-generated by 2024 [Liang et al., 2024]; and annual retractions pass 10,000 for the first time in 2023, with a median publication-to-retraction lag of about 550 days [Van Noorden, 2023, Lei et al., 2024]. None of this identifies the structural system, but it pins down the signs and timescales the calibration needs.

The model distinguishes frontier knowledge F from derivative knowledge R. Their composition defines data quality, $Q \equiv F/(F+R)$. Frontier productivity is increasing in Q: as derivative material displaces frontier material, the return to frontier effort falls. Below

[1] Shumailov et al. [2024] show that tail information is lost first. Alemohammad et al. [2023] call the phenomenon "model autophagy disorder." Gerstgrasser et al. [2024] show that mixing synthetic and organic data delays but does not prevent collapse.

Table 1: LLM-modification share $\hat{\alpha}$ (%)

Venue	Baseline	Post-ChatGPT				Growth
	Nov 2022	Jun 2023	Dec 2023	Jun 2024	Sep 2024	(pp/mo)
Computer Science (arXiv)	2.3 (0.3)	7.6 (0.5)	15.4 (0.6)	19.2 (0.7)	22.5 (0.8)	1.19 / 0.67
Elec. Eng. & Sys. Sci.	2.9 (0.7)	6.8 (0.8)	12.2 (1.0)	17.8 (1.1)	17.9 (1.0)	0.90 / 0.53
Mathematics (arXiv)	2.5 (0.4)	2.9 (0.3)	3.9 (0.4)	6.2 (0.6)	7.7 (0.6)	—
Physics (arXiv)	2.6 (0.3)	4.6 (0.4)	6.6 (0.5)	9.5 (0.6)	9.7 (0.5)	—
Statistics (arXiv)	3.0 (0.8)	4.8 (0.9)	9.5 (1.4)	9.3 (1.2)	13.2 (1.6)	—
bioRxiv	2.8 (0.3)	5.6 (0.4)	8.5 (0.5)	9.0 (0.5)	10.3 (0.5)	0.57 / 0.18
Nature Portfolio	3.5 (0.5)	3.7 (0.5)	6.3 (0.7)	8.4 (0.7)	8.9 (0.6)	0.33 / 0.33
Pre-ChatGPT mean	2.5	—	—	—	—	—

Bootstrap 95% half-widths in parentheses. Growth: average monthly increase in 2023H2 / 2024H1. From Liang et al. [2025] data.

a threshold $Q^\dagger$ the economy enters a *derivative trap*—frontier innovation contracts and the derivative share rises endogenously. Conventional R&D subsidies cannot restore growth once the economy is deep enough in this region. Data governance can prevent entry, but the required screening intensity rises with AI capacity.

The derivative trap is not, however, the first constraint that binds. The capacity to distinguish frontier from derivative knowledge is itself a productive input. I model it as a stock of *epistemic capital* $\mathscr{E}$, which depreciates and must be replenished using scarce labour and clean training data. As Q falls, evaluators trained on contaminated corpora make more errors. Below $\mathscr{E}^\dagger$, governance becomes ineffective: no feasible screening technology can restore Q above $Q^\dagger$.[2] Under laissez-faire, $\mathscr{E}$ crosses $\mathscr{E}^\dagger$ years before Q crosses $Q^\dagger$, closing the policy window while standard innovation indicators still look benign. The same mechanism implies systematic misclassification of derivative output as frontier when $\mathscr{E}$ is low, so measured innovation can stay positive even as the effective knowledge base degrades.

To my knowledge, no growth model treats evaluative capacity as an endogenous state variable. In the sociology of science, evaluative infrastructure is analysed as a durable institutional stock [Alasuutari et al., 2016, Fochler et al., 2016]; in economics, a large literature studies scientific incentives and evaluation [Dasgupta and David, 1994, Ellison, 2002, Stephan, 2012, Manso, 2011]. The object formalised here—a stock governing how effectively the innovation system filters and canonises claims—does not have a close analogue

[2] Epistemic capital has a public component $\mathscr{E}_{\text{pub}}$ (peer review, shared benchmarks) and a private component $\mathscr{E}_{\text{priv}}$ (proprietary detection tools). Competitive equilibrium underprovides $\mathscr{E}_{\text{pub}}$ because it is nonexcludable. Section 7.4 varies excludability and traces the boundary of the governance trap.

in growth theory.

The two traps interact to discipline policy. Epistemic investment is a prerequisite for governance, and governance for making R&D subsidies productive. Skipping steps wastes resources: subsidising frontier effort when evaluative infrastructure is degraded mainly finances derivative production that cannot be separated from genuine novelty. The mechanism is a double wedge in shadow prices. The planner values frontier knowledge above the market price (each unit of F raises Q) and derivative knowledge below it (each unit of R degrades Q), with both gaps widening as A grows ($a_R > a_F$). The competitive equilibrium features too little frontier effort and too much derivative activity, and the gap increases along the laissez-faire path. Heterogeneous researchers sort endogenously; AI raises the return to recombinative tasks and draws talent from the frontier. Raising frontier headcount alone does not help if the data environment continues to deteriorate.

The model is calibrated to the observed contamination trajectory, the decline in detection accuracy, and peer-review strain (Section 7). Laissez-faire implies a welfare loss of about 6.8% CEV, driven by a persistent reduction in frontier growth. Delay costs are convex, with a kink near the governance-trap crossing.

The paper connects to several literatures. Semi-endogenous growth theory [Romer, 1990, Jones, 1995, Kortum, 1997, Bloom et al., 2020] treats the knowledge stock as homogeneous; here, quality and effort are not substitutable because researcher productivity depends on the composition of the data environment. The AI-and-growth literature [Aghion et al., 2018b, Acemoglu and Restrepo, 2018, 2020, Jones and Tonetti, 2020, Trammell and Korinek, 2024, Jones, 2026] models AI as raising productivity against a fixed informational substrate; endogenising data quality turns the outcome into a race–trap dichotomy that depends on governance. Cong et al. [2021] study data-driven growth with privacy trade-offs but treat data quality as given; Chung and Veldkamp [2024] survey data in macroeconomics more broadly.

Closest in spirit, Farboodi and Veldkamp [2025] study a growth environment in which firms accumulate transaction-generated data as an intangible state variable that improves forecast precision and, through that channel, production performance. Their core mechanism is a data feedback loop: higher output generates more data, which improves prediction, raises productivity, and induces further output. In the baseline model, however, data accumulation by itself does not deliver sustained long-run growth, because the gains from prediction are ultimately bounded by irreducible uncertainty; persistent growth requires the extension in which data enters R&D.

The present paper shifts the focus from data quantity to knowledge quality. The state variable Q indexes the composition of the knowledge commons rather than the volume of firm-level data, and the relevant constraint is not a bound on forecast precision but endogenous deterioration of the informational substrate on which frontier research, recombination, and evaluation jointly depend. The two frameworks therefore share a formally similar self-reinforcing structure, but they differ in the sign of the aggregate spillover: in Farboodi and Veldkamp [2025], feedback from data accumulation can support firm growth, whereas here the analogous feedback amplifies contamination and can generate an epistemic trap.

On talent allocation, Murphy et al. [1991] and Hsieh et al. [2019] show that sorting affects aggregate growth; AI adds a frontier-versus- derivative margin. The informational commons in which AI and human researchers operate is subject to congestion and degradation, a structure familiar from Ostrom [1990]; irreversibility arguments in Dasgupta and Heal [1974] carry over when the stock at risk is the knowledge base rather than a physical resource.[3]

The analysis makes epistemic capacity part of the state. Four contributions follow.

First, epistemic capital enters growth theory as an endogenous state variable.

Second, the two-trap architecture—a derivative trap nested inside a governance trap—yields a testable timing implication: the binding constraint on long-run growth is evaluative erosion, and it can bind while conventional innovation indicators remain positive.

Third, the model implies an instrument hierarchy. Epistemic investment must precede governance, and governance must precede R&D subsidies. The ordering is driven by a double shadow-price wedge ($\lambda_F > V_F$, $\lambda_R < V_R$), with both gaps widening as AI capacity grows (Proposition 3.4). The ordering is structural, not a calibration artefact.

Fourth, the mismeasurement corollary (Corollary 5.9) shows that standard growth accounting can overstate frontier innovation when evaluative capacity is low.

Methodologically, the analysis uses Volterra integral equations for the shadow-price ordering, Nagumo invariance for the trap regions, and Leitmann–Stalford sufficiency for the planner's non-convex problem.

The paper proceeds as follows. Section 2 presents the model. Section 3 defines equilibrium and the planner's problem. Section 4 derives the derivative trap. Section 5 derives the governance trap. Section 6 characterises the instrument hierarchy. Section 7 calibrates

[3] Korinek [2023] surveys generative AI in economic research but treats knowledge-base quality as given. Goodhart [1984], Lucas [1976], and Akerlof [1970] study erosion of information content when agents adapt to fixed evaluation rules. Bloom et al. [2020] document that ideas are getting harder to find; Jones [2022] identifies AI as a potential offset.

and reports quantitative results. Section 8 concludes.

2 The Model

Time is continuous, $t \geq 0$. A unit mass of agents supplies one unit of time each period. The economy produces a final good and accumulates four productive stocks: frontier knowledge F, derivative knowledge R, algorithmic capacity A, and human capital H. Two additional stocks govern verification and curation: public epistemic capital $\mathscr{E}_{\text{pub}}$ and private epistemic capital $\mathscr{E}_{\text{priv}}$. The primitive distinction is between *frontier* content—novel, high-verification contributions—and *derivative* content—recombinative output, including synthetic text. Data governance affects the *composition* of usable content by screening and reclassification; it does not represent the physical creation of new frontier ideas.

2.1 Knowledge production

The economy produces two knowledge stocks. Frontier knowledge $F(t)$ records contributions that expand the feasible set of subsequent research tasks. Derivative knowledge $R(t)$ records recombinative output produced from existing material.[4] Both stocks depreciate at constant rates:

$$\dot{F} = \Lambda_F\, D(Q) A^{a_F} \left(\Omega_F H\right)^{\alpha_F} F^{\xi} - \delta_F F, \tag{1}$$

$$\dot{R} = \Lambda_R A^{a_R} \left(\ell_R H\right)^{\alpha_R} R^{\xi_R} - \delta_R R, \tag{2}$$

where $\Lambda_F, \Lambda_R > 0$ are scale parameters, H is aggregate human capital, Ω_F is quality-adjusted frontier talent (Section 2.2), ℓ_R is the net productive derivative labour share (after deducting private evaluative effort; see Table 2), A is algorithmic capacity (Section 2.1.3), and $D(Q)$ is the erosion function (Section 2.1.2). The asymmetry $a_R > a_F$ allows AI to raise the productivity of recombination more than that of frontier work. I impose $\delta_F < \delta_R$.

Assumption 2.1 (Homogeneity)**.** $\alpha_F + \xi = 1$.

Remark 2.2. Assumption 2.1 removes the level effect of F in frontier growth. Dividing (1) by F yields $g_F = \Lambda_F D(Q) A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi-1} - \delta_F$; under $\alpha_F + \xi = 1$, $F^{\xi-1} = F^{-\alpha_F}$

[4] The frontier/derivative distinction parallels the boundary in Weitzman [1998], where new ideas arise from combining existing ones, but genuine novelty requires drawing from an external pool. In the present model, AI systems recombine but do not access the external pool; the corpus quality Q measures what fraction of the pool remains uncontaminated.

and the frontier term depends on the ratio $(\Omega_F H/F)^{\alpha_F}$. This is the Jones [1995] channel: frontier growth is pinned down by talent quality, data quality, and AI augmentation—but not by the cumulated stock itself.[5] Under $\alpha_F + \xi > 1$, frontier knowledge is self-seeding and the trap dissolves; the assumption is conservative. Appendix H characterises the BGP and shows that, under the calibrated AI feedback, no interior BGP exists in competitive equilibrium.

2.1.1 Final goods production

Competitive firms produce output using a CES aggregator:

$$Y = \left[\alpha_Y \left(A^{\phi_A} F\right)^{\frac{\theta-1}{\theta}} + (1-\alpha_Y)\left(H^{\phi_H} R\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}, \tag{3}$$

with $\theta > 1$ and $\alpha_Y \in (0,1)$. Algorithmic capacity enters both knowledge accumulation ((1)–(2)) and downstream commercialisation ((3)).[6] Competitive prices satisfy $p_F = \partial Y/\partial F$ and $p_R = \partial Y/\partial R$.

2.1.2 Data quality and erosion

Define corpus quality as the frontier share of the aggregate stock:

$$Q(t) \equiv \frac{F(t)}{F(t)+R(t)} \in (0,1]. \tag{4}$$

Frontier productivity is attenuated by an erosion function $D : [0,1] \to [\underline{D}, 1]$,

$$D(Q) = \underline{D} + (1-\underline{D})\,Q^{\sigma}, \qquad \sigma > 0, \quad \underline{D} \in [0,1), \tag{5}$$

so $D(1) = 1$ and $D(0) = \underline{D}$. The parameter $\underline{D}$ captures non-corpus sources of frontier productivity (direct observation, interpersonal exchange); all results below allow $\underline{D} > 0$ up

[5] The BGP growth rate $g^* = \alpha_F(1-\omega)g_H/[\alpha_F(1-\omega) - a_F \nu]$ (Appendix H) depends on human-capital growth g_H as in semi-endogenous growth theory. The contamination channel through $D(Q)$ thus becomes the binding margin.

[6] The restriction $\theta > 1$ ensures uniqueness of the static equilibrium (Appendix G). Under gross substitutability, scarce frontier knowledge commands a higher relative price, pulling talent toward the frontier through the Roy mechanism—a negative feedback that delays the trap but cannot prevent it. Under $\theta < 1$ the feedback reverses. Labour allocations affect stock accumulation rates, not contemporaneous output, as in Romer [1990] and Jones [1995]; the direct resource cost of governance enters through $\Gamma(q)$, deducted from consumption ($C = Y - \Gamma(q)$).

to the bound derived in Section 4. If $D \equiv 1$, contamination is costless and no derivative trap can arise. The channel through $D(Q)$ is the key departure from Farboodi and Veldkamp [2025], who model data as homogeneous signals of fixed precision; there, growth is bounded because forecast accuracy saturates. Here, growth is bounded because the informational substrate *degrades endogenously*: the effective return to data depends not on how much data exists but on what fraction of it is genuine.[7]

2.1.3 Algorithmic capacity

Algorithmic capacity accumulates from derivative content:

$$\dot{A} = \mu_A R^{\nu} A^{\omega} - \delta_A A, \tag{6}$$

with $\nu, \omega \in (0,1)$ and $\delta_A > 0$. The condition $a_R > a_F$ together with (6) creates a positive feedback: more derivative output raises R, which raises A (training data), which augments derivative productivity, which raises R further. The self-reinforcement index $\mathfrak{S} \equiv a_R \nu / [(1-\xi_R)(1-\omega)]$ measures round-trip amplification; when $\mathfrak{S} \geq 1$, no interior balanced-growth path exists under laissez-faire (Appendix H).[8] Appendix F verifies that an alternative flow formulation ($\dot{A} = \mu_A (\dot{R}^+)^{\nu} A^{\omega} - \delta_A A$) yields identical qualitative results.

2.2 Talent allocation

Labour is allocated across four activities. Table 2 collects the notation; Figure 1 provides a schematic.

A unit mass of workers divides into education (ℓ_H) and research ($1-\ell_H$). Among researchers, ability z is drawn from a Pareto distribution with shape $\zeta > 1$ and lower bound $\underline{z} > 0$:

$$\Pr(z > x) = \left(\frac{\underline{z}}{x}\right)^{\zeta}, \qquad x \geq \underline{z}. \tag{7}$$

Each researcher sorts into frontier or derivative work via the Roy indifference condition (Section 3): types $z \geq \bar{z}$ enter frontier work, types $z < \bar{z}$ enter derivative work. The Pareto

[7] A Fréchet microfoundation delivers the functional form: if each item's novelty is drawn from a Fréchet distribution with shape $k > 0$, the productivity multiplier from sampling n items of which fraction Q is uncontaminated is $Q^{1/k}$; setting $\sigma = 1/k$ recovers (5). See Appendix A.1. The empirical evidence on model collapse [Shumailov et al., 2024, Alemohammad et al., 2023] disciplines the curvature $\sigma \in [1,3]$.

[8] If $a_R \leq a_F$, the loop breaks: rising A raises g_F faster than g_R, compositional drift reverses, and Q rises endogenously. The condition $a_R > a_F$ reflects that current AI systems excel at pattern recombination rather than open-ended hypothesis formation [Aghion et al., 2018a].

Table 2: Labour allocation across activities

Symbol	Activity	Eq.	CE	Planner
ℓ_H	Education	(12)	Exogenous	Control
$\pi_F(\bar{z})$	Frontier research	(1)	Roy sorting at $\bar{z}$	Control (ℓ_F)
$\pi_R(\bar{z})$	Derivative research	(2)	$1-\pi_F-\ell_H-\ell_{\mathscr{E}}$	Residual
$\ell_{\mathscr{E}}$	Public evaluation	(13)	0 (public good)	Control
$\ell_{\mathscr{E}}^{\text{priv}}$	Private evaluation	(14)	$\theta_{\mathscr{E}}\pi_R$ (firm)	—
ℓ_R	Net derivative labour	(2)	$(1-\theta_{\mathscr{E}})\pi_R$	Control
$\Omega_F(\bar{z})$	Frontier talent (derived)	(1)	Ability-weighted integral; see (10)	

tail gives the frontier participation rate:

$$\pi_F(\bar{z}) \;=\; (1-\ell_H)\left(\frac{\underline{z}}{\bar{z}}\right)^{\zeta}, \tag{8}$$

with derivative participation

$$\pi_R(\bar{z}) \;=\; (1-\ell_H)\Big[1-\big(\underline{z}/\bar{z}\big)^{\zeta}\Big]. \tag{9}$$

A fraction $\theta_{\mathscr{E}}$ of derivative workers is diverted to private evaluative effort ($\ell_{\mathscr{E}}^{\text{priv}} = \theta_{\mathscr{E}}\pi_R$), leaving $\ell_R = (1-\theta_{\mathscr{E}})\pi_R$ as the net productive input in (2).

The quality-adjusted frontier talent input Ω_F integrates ability raised to γ_F over the frontier pool:

$$\Omega_F(\bar{z}) \;\equiv\; (1-\ell_H)\int_{\bar{z}}^{\infty} z^{\gamma_F}\,\frac{\zeta\,\underline{z}^{\zeta}}{z^{\zeta+1}}\,dz \;=\; (1-\ell_H)\,\frac{\zeta\,\underline{z}^{\zeta}}{\zeta-\gamma_F}\,\bar{z}^{\gamma_F-\zeta}, \tag{10}$$

where $\gamma_F \in (0,\zeta)$ and the integral converges because $\gamma_F < \zeta$ (Assumption 2.4(ii)).[9]

The key comparative static is:

$$\frac{\partial \Omega_F}{\partial \bar{z}} \;=\; (\gamma_F-\zeta)\,\frac{\Omega_F}{\bar{z}} \;<\; 0 \qquad (\text{since } \gamma_F < \zeta). \tag{11}$$

A rise in $\bar{z}$ removes the lowest-ability frontier researchers (selection effect) but also shrinks

[9] An analogous integral defines derivative-sector effective labour $\Omega_R(\bar{z}) = (1-\ell_H)\zeta\underline{z}^{\zeta}/(\zeta-\gamma_R)\,\bar{z}^{\gamma_R-\zeta}$ with $\gamma_R < \gamma_F$. Since $\gamma_R < \gamma_F$, frontier production is more ability-sensitive—the comparative-advantage structure of the Roy model. Public evaluative labour $\ell_{\mathscr{E}}$ enters the planner's clearing condition but vanishes in CE (Proposition 3.2(i)), reflecting the public-good externality.

the frontier pool (mass effect). Under $\gamma_F < \zeta$, the mass effect dominates.[10]

2.3 Human capital

Human capital accumulates via a Lucas [1988]–Uzawa [1965] specification:

$$\dot{H} = \lambda_H \, \ell_H^{\beta_H} H - \delta_H H, \tag{12}$$

where ℓ_H is the education share, $\beta_H \in (0,1)$, $\lambda_H > 0$, and $\delta_H > 0$. Human-capital accumulation is independent of Q; making λ_H load on corpus quality would tighten the trap.[11]

2.4 Epistemic capital

Definition 2.3 (Epistemic capital)**.** $\mathscr{E}(t) \in [0,1]$ is the aggregate evaluative capacity—the ability to distinguish frontier from derivative knowledge, normalised so that $\mathscr{E} = 1$ denotes perfect discrimination and $\mathscr{E} = 0$ denotes none.[12]

Epistemic capital decomposes as $\mathscr{E} = \mathscr{E}_{\text{pub}} + \mathscr{E}_{\text{priv}}$. Public epistemic capital encompasses shared benchmarks, peer-review standards, and public detection tools (nonexcludable). Private epistemic capital encompasses proprietary detection infrastructure (partially appropriable). The accumulation equations are:

$$\dot{\mathscr{E}}_{\text{pub}} = \lambda_{\mathscr{E}}^{\text{pub}} \, \ell_{\mathscr{E}}^{\eta_{\mathscr{E}}} D_{\mathscr{E}}(Q) - \delta_{\mathscr{E}}(\varphi) \, \mathscr{E}_{\text{pub}}, \tag{13}$$

$$\dot{\mathscr{E}}_{\text{priv}} = \lambda_{\mathscr{E}}^{\text{priv}} \, \ell_{\mathscr{E}}^{\text{priv},\,\eta_{\mathscr{E}}} D_{\mathscr{E}}(Q) - \delta_{\mathscr{E}}^{\text{priv}}(\varphi) \, \mathscr{E}_{\text{priv}}, \tag{14}$$

where $\ell_{\mathscr{E}}$ is public evaluative labour, $\ell_{\mathscr{E}}^{\text{priv}}$ is aggregate private evaluative investment, $D_{\mathscr{E}}(Q) = Q^{\sigma_{\mathscr{E}}}$, $\varphi \equiv 1 - Q$, and $\delta_{\mathscr{E}}(\varphi) = \delta_{\mathscr{E},0} + \delta_{\mathscr{E},1}\varphi$ with $\delta_{\mathscr{E},1} > 0$. Private depreciation takes the same affine form.

Two features drive the governance trap. Replenishment loads $Q^{\sigma_{\mathscr{E}}}$: evaluators trained on contaminated corpora develop contaminated judgement, so at $\sigma_{\mathscr{E}} = 0$ the governance

[10] Under $\gamma_F > \zeta$, the selection effect dominates; a brain drain raises Ω_F, reversing the talent-drain channel and dissolving the derivative trap.

[11] Contaminated textbooks or AI tutoring systems trained on derivative material would open an additional channel through which falling Q erodes the talent base. The present specification is conservative.

[12] The signal-detection microfoundation in Appendix A.2 maps a raw precision parameter (inverse noise variance) to the hit rate $s(\mathscr{E}) = \Phi_{\mathscr{N}}(\Delta(\mathscr{E})/2)$; the model works with the normalised index throughout.

trap dissolves. Depreciation rises with φ: a higher volume of derivative content overwhelms evaluators, so at $\delta_{\mathscr{E},1} = 0$ the temporal-precedence result weakens.

In competitive equilibrium $\ell_{\mathscr{E}}^{CE} = 0$ while $\ell_{\mathscr{E}}^{\text{priv},CE} > 0$.

2.5 Data governance

A governance technology screens corpus content and reclassifies verified derivative material as frontier-certified, effectively transferring stock from R to F. Governance intensity $q \in [0,1]$ parameterises the fraction screened. Screening accuracy $s(\mathscr{E})$ is increasing and strictly concave in $\mathscr{E}$; the false-positive rate $f(\mathscr{E}) = 1 - s(\mathscr{E})$ is decreasing.[13] Under governance, the laws of motion for F and R become

$$\dot{F} = \Lambda_F D(Q) A^{a_F} \left(\Omega_F H\right)^{\alpha_F} F^{\xi} - \delta_F F + q\, s(\mathscr{E})\, \kappa R, \tag{15}$$

$$\dot{R} = \Lambda_R A^{a_R} \left(\ell_R H\right)^{\alpha_R} R^{\xi_R} - \delta_R R - q\, s(\mathscr{E})\, \kappa R, \tag{16}$$

where $\kappa > 0$ is the reclassification rate. The total corpus $F + R$ evolves through production and depreciation alone; governance redistributes content between stocks without creating or destroying knowledge.

In the presence of governance, F is best read as the high-integrity corpus: material whose provenance is verified. Governance adds material by certifying derivative content that passes screening; reclassification does not create novelty.[14]

Differentiating $Q \equiv F/(F+R)$ and substituting (15)–(16) yields

$$\dot{Q} = Q(1-Q)(g_F^{\text{prod}} - g_R^{\text{prod}}) + q \cdot s(\mathscr{E}) \cdot \kappa \cdot (1-Q), \tag{17}$$

where $g_F^{\text{prod}}, g_R^{\text{prod}}$ denote the non-governance growth rates.[15] The first term captures compositional drift; the second captures active screening. Under laissez-faire ($q = 0$), (17)

[13] The functional form admits a signal-detection microfoundation (Appendix A.2). An evaluator observes $y = \theta + \varepsilon$, $\varepsilon \sim \mathscr{N}(0, 1/\mathscr{E})$; the sensitivity index $\Delta(\mathscr{E}) = |\mu_F - \mu_R|\sqrt{\mathscr{E}}$ yields $s(\mathscr{E}) = \Phi_{\mathscr{N}}\big(\Delta(\mathscr{E})/2\big)$, strictly concave for all $\mathscr{E} > 0$. A base-rate shift from rising φ depresses the positive predictive value, disciplining $\delta_{\mathscr{E}}(\varphi)$ increasing in φ. Retraining on a contaminated corpus ($Q < 1$) reduces effective sensitivity to $Q \cdot \Delta(\mathscr{E})$; this Q-dependence is absorbed into the reduced form $D_{\mathscr{E}}(Q) = Q^{\sigma_{\mathscr{E}}}$.

[14] An alternative formulation with a separate "excluded" stock X would add a state variable without changing the qualitative dynamics, since $D(Q)$ depends only on the ratio of trusted to total corpus. If governance misclassifies frontier work as derivative (false positives), aggressive screening at low $\mathscr{E}$ damages the frontier; the specification $s(\mathscr{E})$ implicitly nets out false positives.

[15] Explicitly: $\dot{Q} = (\dot{F}R - F\dot{R})/((F+R)^2)$. The governance transfers $+qs\kappa R$ in $\dot{F}$ and $-qs\kappa R$ in $\dot{R}$ contribute $qs\kappa R(R+F)/(F+R)^2 = qs\kappa(1-Q)$.

reduces to $\dot{Q} = Q(1-Q)(g_F - g_R)$.

2.6 Regularity assumptions

Assumption 2.4 (Parameter restrictions)**.** (i) $a_R > a_F$.

(ii) $\gamma_F < \zeta$.

(iii) $\sigma_{\mathscr{E}} > 0$.

(iv) $\delta_{\mathscr{E},1} > 0$.

(v) $\underline{D} < \bar{D}$.

(vi) $\lambda_{\mathscr{E}}^{\text{priv}} / \delta_{\mathscr{E},0}^{\text{priv}} < 1$.

(vii) Along calibrated CE and planner paths, $\mathscr{E}_{\text{pub}}(t) + \mathscr{E}_{\text{priv}}(t) < 1$ (verified numerically).

(viii) $g_R^{\text{prod}} > 0$ on $\partial \mathscr{T}$.[16]

Each restriction activates a specific model channel. Conditions (i)–(ii) generate the compositional-drift and talent-drain mechanisms; (iii)–(iv) generate the governance trap and temporal precedence; (v) ensures the trap region is nonempty; (vi) prevents private epistemic investment from substituting for public infrastructure. Reversing any of (i)–(iv) dissolves the corresponding trap channel; Appendix A.2 provides further discussion of the epistemic-capital domain conventions.[17]

Assumption 2.5 (Frontier boundedness)**.** $F(t) \geq \underline{F} > 0$ for all $t \in [0, T]$, any finite T.

Under maximal erosion and zero investment, F declines exponentially at rate δ_F and reaches zero only asymptotically. The assumption disciplines the regularity of the barrier functions $Q^\dagger$ and $\mathscr{E}^\dagger$—both involve division by $F^{1-\xi}$—and places the vector field within the scope of the Nagumo and Picard–Lindelöf theorems.

2.7 Labour market clearing

The total labour endowment is normalised to unity. In competitive equilibrium, public evaluative labour vanishes by epistemic neglect (Proposition 3.2), giving the clearing condition

$$\pi_F(\bar{z}) \; + \; \pi_R(\bar{z}) \; + \; \ell_H \; = \; 1. \tag{18}$$

[16] At calibration, $g_R^{\text{prod}} = \Lambda_R A^{a_R} (\ell_R H)^{\alpha_R} R^{\xi_R - 1} - \delta_R$. Although $\xi_R < 1$ implies $R^{\xi_R - 1} \to 0$ as $R \to \infty$, the R–A feedback with $\mathfrak{S} = a_R \nu / [(1-\xi_R)(1-\omega)] > 1$ ensures A grows fast enough to keep $\Lambda_R A^{a_R} (\ell_R H)^{\alpha_R} R^{\xi_R - 1}$ bounded away from zero along equilibrium paths. Lemma G.5 verifies this formally.

[17] The screening accuracy $s(\mathscr{E}) = \Phi_{\mathscr{N}}(c\sqrt{\mathscr{E}})$ is well-defined on $[0, \infty)$. Under CE, $\mathscr{E}_{\text{pub}}$ decays monotonically ($\ell_{\mathscr{E}}^{CE} = 0$); under the planner, the replenishment-depreciation balance limits steady-state levels.

Private evaluative investment is financed within the derivative sector: each firm diverts fraction θ_i of its workforce from production to proprietary detection, accumulating a firm-level detection stock e_i. Certified output commands price $\tilde{p}_{R,i} = p_R(1+\kappa_{\text{cert}} e_i)$, so each firm internalises its own certification benefit. In symmetric equilibrium $\theta_i = \theta_{\mathscr{E}}$ for all i, giving aggregate private epistemic labour $\ell_{\mathscr{E}}^{\text{priv}} = \theta_{\mathscr{E}} \cdot \pi_R(\bar{z})$, with $\theta_{\mathscr{E}}$ chosen to maximise instantaneous certification revenue net of diverted labour. The net productive derivative labour entering (2) is $\ell_R = (1-\theta_{\mathscr{E}})\pi_R(\bar{z})$.

The planner sets all labour allocations directly:

$$\ell_F + \ell_R + \ell_{\mathscr{E}} + \ell_H = 1, \tag{19}$$

where ℓ_F replaces $\pi_F(\bar{z})$ and ℓ_R replaces $(1-\theta_{\mathscr{E}})\pi_R(\bar{z})$.

The correspondence between labour shares and the production functions (1)–(14) is:

Production function	Labour input	Meaning
Frontier $\dot{F}$, eq. (1)	$\Omega_F(\bar{z})$	Quality-adjusted talent (not a raw share)
Derivative $\dot{R}$, eq. (2)	$\ell_R = (1-\theta_{\mathscr{E}})\pi_R$	Net productive derivative labour
Public epistemic $\dot{\mathscr{E}}_{\text{pub}}$, eq. (13)	$\ell_{\mathscr{E}}$	Public evaluative effort (= 0 in CE)
Private epistemic $\dot{\mathscr{E}}_{\text{priv}}$, eq. (14)	$\ell_{\mathscr{E}}^{\text{priv}} = \theta_{\mathscr{E}}\pi_R$	Diverted from derivative sector
Human capital $\dot{H}$, eq. (12)	ℓ_H	Education share

2.8 State space

The state vector is

$$\mathbf{x}(t) = (F, R, A, H, \mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}) \in \mathbb{R}^4_{++} \times \mathbb{R}^2_{+}.$$

The planner chooses $(q, \ell_F, \ell_R, \ell_{\mathscr{E}}, \ell_H)$; in competitive equilibrium, $\bar{z}$ and $\theta_{\mathscr{E}}$ are determined by the static equilibrium map (Proposition G.3). Time is continuous; the horizon is infinite; all agents discount at $\rho > 0$.

Figure 1 collects the model's stocks, flows, and feedback channels into a single schematic.

3 Competitive Equilibrium and Social Planner

The decentralised equilibrium exhibits four distortions: frontier knowledge is undervalued, derivative output is simultaneously overvalued, data quality is unpriced, and epistemic capital is neglected (Proposition 3.4). The first two form a *double wedge* that widens as AI

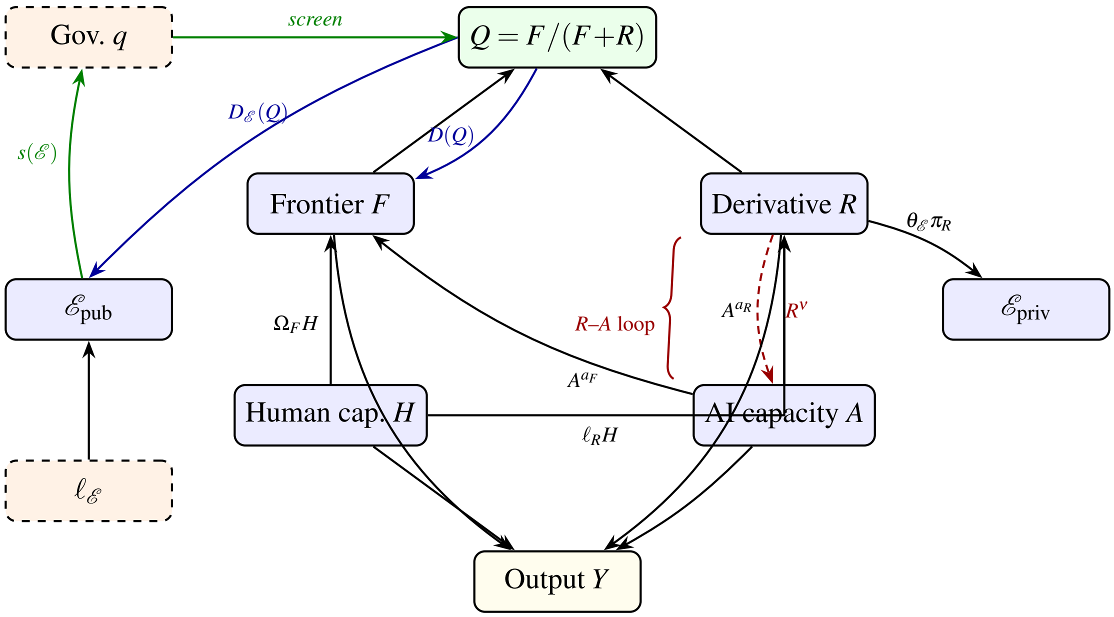

Figure 1: Model structure. Solid boxes: state variables. Dashed boxes: policy instruments ($= 0$ in CE). Red dashed: R–A loop. Blue: quality-dependent channels. Green: governance channel.

capacity grows: the market pays too little for frontier effort and too much for derivative production, because no agent internalises the quality externality $\partial Q/\partial F > 0$, $\partial Q/\partial R < 0$. Together these push the economy toward the governance trap along a path that appears efficient by standard metrics.

The primitives are the state vector $\mathbf{x}(t) = (F, R, A, H, \mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}) \in \mathbb{R}^4_{++} \times [0,1]^2$ and initial condition $\mathbf{x}_0$, both defined in Section 2.8. The objects $D(Q)$, $D_{\mathscr{E}}(Q)$, and $\delta_{\mathscr{E}}(\varphi)$ depend on aggregate Q (recall $\varphi = 1 - Q$); atomistic agents take these as given.

3.1 Definition of competitive equilibrium

Since the model has no physical capital, the competitive equilibrium is a *sequence of instantaneous equilibria*: at each t, knowledge-service prices clear the market, Roy sorting determines talent allocation, and the private detection share $\theta_{\mathscr{E}}$ solves a static certification FOC. Intertemporal trade-offs arise only through stock accumulation; the welfare comparison with the planner enters through the planner's dynamic optimisation problem (Section 3.2).

The final good is numéraire. Let $p_F(t)$, $p_R(t)$ be rental prices of frontier and derivative knowledge services, and let $w_j(z,t) = p_j(t)\,z^{\gamma_j}A(t)^{a_j}$ be the wage in sector $j \in \{F,R\}$ for ability z.

Definition 3.1 (Competitive equilibrium)**.** Given $\mathbf{x}_0$ and a policy process $\{q(t), \ell_{\mathscr{E}}(t)\}_{t\geq 0}$ (set to zero under laissez-faire), a competitive equilibrium is an allocation $\{C, \bar{z}, \theta_{\mathscr{E}}, \ell_H, \mathbf{x}\}$ and prices $\{p_F, p_R\}$ such that, for all $t \geq 0$:

(i) **Final-goods firms** maximise static profits: $p_F = \partial Y/\partial F,\ \ p_R = \partial Y/\partial R$, with Y given by (3).

(ii) **Researcher sorting.** Each researcher chooses the activity maximising $w_j(z,t)$. A cutoff $\bar{z}(t)$ satisfies $w_F(\bar{z}) = w_R(\bar{z})$; types $z \geq \bar{z}$ enter the frontier, types $z < \bar{z}$ enter derivative work.

(iii) **Private epistemic investment.** Each derivative-sector firm chooses a detection share $\theta_{\mathscr{E}}(t) \in [0,1]$ to maximise instantaneous certification revenue net of diverted labour. Because firms are atomistic, each takes the aggregate stock $\mathscr{E}_{\text{priv}}$ as given; the stock evolves mechanically from (14) under the aggregate allocation $\ell_{\mathscr{E}}^{\text{priv}} = \theta_{\mathscr{E}}\pi_R$.

(iv) **Household.** The representative household maximises $\int_0^\infty e^{-\rho t}u(C)\,dt$ with CRRA utility $u(C) = C^{1-\eta}/(1-\eta)$. Under laissez-faire, $C(t) = Y(t)$; under governance, $C(t) = Y(t) - \Gamma(q(t))$.

(v) **Market clearing and dynamics.** Labour-market clearing (18) holds; $\mathbf{x}(t)$ follows the laws of motion (1)–(14) under the equilibrium allocation.

The equilibrium is inefficient for three reasons. Sector choice affects Q, which enters $D(Q)$ and $D_{\mathscr{E}}(Q)$; no agent internalises this (composition externality). Public epistemic capital is nonexcludable, so $\ell_{\mathscr{E}}^{CE} = 0$ (epistemic externality). And governance is a public good, so $q^{CE} = 0$ (governance externality). The three are dynamically coupled: the composition externality erodes Q, which degrades $\mathscr{E}_{\text{pub}}$ through the epistemic externality, which disables governance. The instrument hierarchy (Proposition 6.1) requires addressing all three.

Appendix G establishes existence and uniqueness.

3.1.1 Roy sorting

The indifference condition yields

$$\bar{z}(t) = \left(\frac{p_R(t)}{p_F(t)} \cdot A(t)^{a_R - a_F}\right)^{\frac{1}{\gamma_F - \gamma_R}}. \tag{20}$$

Since $a_R > a_F$ and $\gamma_F > \gamma_R$, rising A raises $\bar{z}$: stronger AI pulls marginal researchers into derivative work, reducing both π_F and Ω_F.

Proposition 3.2 (Epistemic neglect)**.** *In competitive equilibrium: (i)* $\ell_{\mathscr{E}}^{CE} = 0$*; (ii)* $\ell_{\mathscr{E}}^{priv,CE} > 0$*, with* $\theta_{\mathscr{E}}^{CE}$ *solving the FOC equating marginal certification revenue to marginal production cost; (iii)* $q^{CE} = 0$*.*

Proof. (i): Nonexcludability of $\mathscr{E}_{\text{pub}}$; the Nash equilibrium in evaluative effort is zero. (ii): Each derivative-sector firm i accumulates a *proprietary* detection stock e_i by diverting fraction θ_i of its workforce to detection ($e_i = \theta_i \pi_R H^{\eta_{\mathscr{E}}} D_{\mathscr{E}}(Q)$ in the static representation). The certified price for firm i is $\tilde{p}_{R,i} = p_R(1 + \kappa_{\text{cert}} e_i)$, so each firm internalises the return to its own detection effort. The FOC equating marginal certification revenue (decreasing in θ_i by concavity of $D_{\mathscr{E}}$) to marginal production loss has interior solutions when $D_{\mathscr{E}}(Q) > 0$ and $\kappa_{\text{cert}} > 0$. In symmetric equilibrium, $e_i = e$ for all i and $\mathscr{E}_{\text{priv}} = e \cdot \pi_R(\bar{z})$. (iii): Same public-good logic as (i). □ □

Part (ii) is the partial corrective: derivative-sector firms invest in proprietary detection because certified output commands a price premium. The private buffer is insufficient because each firm internalises only its own certification benefit, not the system-wide improvement in screening accuracy.[18]

3.1.2 Equilibrium dynamics

Under the equilibrium allocations, the state evolves as:

$$\dot{F}^{CE} = \Lambda_F\, D(Q) A^{a_F} \left(\Omega_F(\bar{z}) H\right)^{\alpha_F} F^{\xi} - \delta_F F, \tag{21}$$

$$\dot{R}^{CE} = \Lambda_R A^{a_R} \left((1-\theta_{\mathscr{E}}^{CE})\pi_R(\bar{z}) H\right)^{\alpha_R} R^{\xi_R} - \delta_R R, \tag{22}$$

$$\dot{\mathscr{E}}_{\text{pub}}^{CE} = -\delta_{\mathscr{E}}(\varphi)\, \mathscr{E}_{\text{pub}}^{CE}. \tag{23}$$

Equation (23) is the critical equation: with $\ell_{\mathscr{E}}^{CE} = 0$, public epistemic capital depreciates monotonically at a rate that is itself increasing along the laissez-faire path (since φ rises). The decline is self-accelerating: rising φ raises $\delta_{\mathscr{E}}(\varphi)$, which erodes $\mathscr{E}$, which (under positive governance) would reduce screening accuracy, permitting further derivative accu-

[18] If the certification premium κ_{cert} were large enough, private investment could sustain the evaluative infrastructure and the governance trap would dissolve—this is the failure mode identified by Assumption 2.4(vi).

mulation. In competitive equilibrium the screening link is severed ($q^{CE} = 0$) and the spiral operates purely through the depreciation channel.

3.2 Social planner's problem

Definition 3.3 (Planner's problem)**.** Given $\mathbf{x}_0$, the planner chooses $(q, \ell_F, \ell_R, \ell_{\mathscr{E}}, \ell_H)$ to maximise

$$\mathscr{W}(\mathbf{x}_0) = \max_{\mathbf{u}(\cdot)} \int_0^\infty e^{-\rho t}\, u\big(C(t)\big)\, dt, \tag{24}$$

subject to: (a) $C = Y - \Gamma(q)$ with $\Gamma(0) = 0$, $\Gamma' > 0$, $\Gamma'' > 0$; (b) labour-market clearing $\ell_F + \ell_R + \ell_{\mathscr{E}} + \ell_H = 1$; (c) the laws of motion (15)–(14) (the planner controls q, so the governance transfer terms are active); and (d) bounds $q \in [0,1]$, $\ell_j \in [0,1]$ for each j.

The planner treats $\bar{z}$ (equivalently ℓ_F) as a control rather than an equilibrium outcome, and internalises the dependence of D, $D_{\mathscr{E}}$, and $\delta_{\mathscr{E}}$ on (Q, φ).

The planner sets $\ell_{\mathscr{E}} > 0$ and $q > 0$; both are zero in competitive equilibrium.[19]

3.2.1 Costates and the marginal value of data quality

The current-value Hamiltonian is $\mathscr{H} = u(C) + \lambda_F \dot{F} + \lambda_R \dot{R} + \lambda_A \dot{A} + \lambda_H \dot{H} + \lambda_{\mathscr{E}}^{\text{pub}} \dot{\mathscr{E}}_{\text{pub}} + \lambda_{\mathscr{E}}^{\text{priv}} \dot{\mathscr{E}}_{\text{priv}}$, with six states $(F, R, A, H, \mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}})$ and six costates. Since $Q = F/(F+R)$ is not a state but enters the right-hand sides through $D(Q)$, $D_{\mathscr{E}}(Q)$, and $\delta_{\mathscr{E}}(\varphi)$, the chain rule generates Q-channel contributions in the Euler equations for λ_F and λ_R. Define the *composite marginal value of data quality*:

$$\Psi_Q \;\equiv\; \frac{\partial \mathscr{H}}{\partial Q}\bigg|_{F,R\text{ held fixed}} \;=\; \lambda_F \cdot \frac{D'(Q)}{D(Q)} \cdot G_F \;+\; \lambda_{\mathscr{E}}^{\text{pub}} \cdot \frac{\partial \dot{\mathscr{E}}_{\text{pub}}}{\partial Q} \;+\; \lambda_{\mathscr{E}}^{\text{priv}} \cdot \frac{\partial \dot{\mathscr{E}}_{\text{priv}}}{\partial Q}, \tag{25}$$

where $G_F \equiv \Lambda_F D(Q) A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi}$ is gross frontier production. The first term is the productivity channel: higher Q raises frontier output through D, with $D'(Q)/D(Q)$ measuring the semi-elasticity of the erosion function; the second and third are epistemic-replenishment channels (through both $D_{\mathscr{E}}(Q) = Q^{\sigma_{\mathscr{E}}}$ and the depreciation rate $\delta_{\mathscr{E}}(\varphi)$ with $\varphi = 1 - Q$). Ψ_Q is a derived quantity, not a costate; it enters the Euler equations for λ_F and λ_R through $\partial Q/\partial F = R/(F+R)^2$ and $\partial Q/\partial R = -F/(F+R)^2$ (Appendix E.1.1).

[19] The Hamiltonian is not jointly concave in (F, R) because $D''(Q) > 0$ for $\sigma > 1$ (Proposition I.1). Sufficiency is established via the Leitmann–Stalford decomposition in Appendix I.

The FOC for governance comes directly from differentiating $\mathscr{H}$ with respect to q: since governance transfers stock from R to F (equations (15)–(16)), the marginal benefit equals the costate gap times the transfer rate:

$$(\lambda_F - \lambda_R) \cdot s(\mathscr{E}) \cdot \kappa \cdot R = u'(C)\,\Gamma'(q). \tag{26}$$

In the decentralised economy, neither the data-quality externality nor the epistemic externality is priced, so $q^{CE} = 0$.

3.3 Shadow-price ordering

Proposition 3.4 (Shadow-price ordering)**.** *Along any path with $\varphi > 0$ and $q^{SP} > 0$: (i) $\lambda_F > V_F > 0$; (ii) $\Psi_Q > 0$; (iii) $\lambda_{\mathscr{E}}^{pub} > 0$; (iv) $\lambda_A \lesseqgtr V_A$ (ambiguous); (v) $\lambda_R < V_R$ whenever $\lambda_A \geq 0$.*

The proof applies the Volterra fixed-point theorem to the coupled costate system (Appendix B.1; part (v) in Appendix B.2). Frontier knowledge generates a data-quality externality ($\partial Q/\partial F > 0$) and an epistemic externality (higher Q supports $\mathscr{E}$ replenishment) that the planner internalises but atomistic agents ignore. At the calibration, the data-quality externality accounts for roughly 60% of $\lambda_F - V_F$.

Part (v) is the mirror image: derivative knowledge carries a negative quality externality ($\partial Q/\partial R < 0$) that the market does not price. Because $\Psi_Q > 0$ (part (ii)) and $\partial Q/\partial R = -F/(F+R)^2 < 0$, each additional unit of derivative output depresses data quality and, through $D(Q)$, frontier productivity. The planner internalises this cost, so λ_R falls below the market shadow value V_R. The wedge widens as AI capacity grows: higher A amplifies derivative production through A^{a_R} with $a_R > a_F$, so the marginal quality damage per unit of R increases with algorithmic improvement. Parts (i) and (v) together establish a *double wedge*: the competitive equilibrium simultaneously undervalues frontier knowledge and overvalues derivative output, providing the theoretical foundation for a Pigouvian tax on synthetic content or, equivalently, the derivative tax in Corollary 3.6(a).

Remark 3.5 (Why λ_A is ambiguous while λ_R is not)**.** AI capacity augments frontier production ($a_F > 0$: positive) but also accelerates derivative expansion ($a_R > a_F$: negative) and drains frontier talent ($\partial \bar{z}/\partial A > 0$: negative). The sign of $\lambda_A - V_A$ reverses near $T_{\mathscr{E}}$ at the calibration. Derivative knowledge, by contrast, has an unambiguously negative quality externality: $\partial Q/\partial R < 0$ always, with no offsetting frontier channel. The asymmetry reflects the model's core mechanism—AI is a dual-use technology whose net social value depends on the composition of its output, while derivative content is unambiguously harmful to the knowledge commons.

Corollary 3.6 (Policy instruments)**.** *The planner's optimum requires three instruments absent in competitive equilibrium: (a) a frontier subsidy / derivative tax ($\ell_F^{SP} > \pi_F(\bar{z}^{CE})$); (b) data governance ($q^{SP} > 0$); (c) public epistemic investment ($\ell_{\mathscr{E}}^{SP} > 0$). Instruments (b) and (c) are superadditive.*

3.4 Comparative statics

Lemma 3.7. *Along the competitive-equilibrium path: (i) $\partial \bar{z}/\partial A > 0$; (ii) $\partial Q^{CE}/\partial A < 0$ for large A; (iii) $\partial \mathscr{E}_{pub}^{CE}/\partial t < 0$ for all t; (iv) $g_Q = (1-Q)(g_F - g_R)$, negative when $g_R > g_F$.*

Proof. Each item follows from differentiation of the equilibrium laws of motion under $\ell_{\mathscr{E}}^{CE} = 0$ and $q^{CE} = 0$. □

3.5 Welfare

The consumption-equivalent variation of the planner's policy is the constant proportional increase Δ in competitive-equilibrium consumption making the household indifferent:

$$\int_0^\infty e^{-\rho t} u\big((1+\Delta)C^{CE}(t)\big)\, dt = \mathscr{W}(\mathbf{x}_0). \tag{27}$$

The competitive equilibrium exhibits three reinforcing distortions: epistemic neglect ($\ell_{\mathscr{E}}^{CE} = 0$), governance absence ($q^{CE} = 0$), and invisible erosion—low $\mathscr{E}$ inflates the measured frontier share Q^{obs} (Corollary 5.9), so standard metrics report healthy growth while the true frontier contracts.

4 The Derivative Trap

Once the frontier share Q falls below a threshold $Q^\dagger$, frontier growth turns non-positive and the decline in Q is self-reinforcing. Non-governance instruments—R&D subsidies, talent policy, copyright reform—cannot break the invariance.

4.1 The quality threshold

Definition 4.1 (Quality threshold)**.** The *quality threshold* $Q^\dagger$ is the data quality level at which net frontier growth is exactly zero:

$$Q^\dagger(F,H,A,\Omega_F) \equiv \left(\frac{\delta_F - \underline{D}\,G}{(1-\underline{D})\,G}\right)^{1/\sigma}, \qquad G \equiv \Lambda_F A^{a_F}(\Omega_F H)^{\alpha_F} F^{\xi-1}, \tag{28}$$

when the expression is in $[0,1]$. Two boundary cases arise:

- If $G \le \delta_F$ (gross productivity cannot offset depreciation even at full quality): $Q^\dagger = 1$. The frontier contracts at every Q, so the *entire* state space lies in the trap.
- If $\underline{D}\,G \ge \delta_F$ (the productivity floor sustains growth even at $Q = 0$): $Q^\dagger = 0$. No trap exists.

The composite G collects the non-quality determinants of gross frontier productivity.

Setting $\dot{F} = 0$ in (1) and solving for Q yields (28): below $Q^\dagger$, the erosion function $D(Q)$ can no longer compensate for depreciation, so the frontier stock shrinks. The threshold captures the "ideas are hard to find" phenomenon of Bloom et al. [2020] in a contamination-specific form: productivity declines not because undiscovered ideas are depleted, but because the corpus is diluted by derivative content.[20]

The comparative statics identify policy levers: $\partial Q^\dagger/\partial A < 0$ (AI augmentation raises G, lowering the threshold); $\partial Q^\dagger/\partial \Omega_F < 0$ (more frontier talent lowers it); $\partial Q^\dagger/\partial \underline{D} < 0$ (a higher productivity floor lowers it); $\partial Q^\dagger/\partial F > 0$ when $\xi < 1$. Theorem 4.9 establishes that none suffice without governance.

The productivity-floor bound is

$$\bar{D} \equiv \delta_F / G. \tag{29}$$

At the calibration, $\bar{D} \approx 0.08$, evaluated at initial-period values of (A,Ω_F,H,F); the bound shifts along the transition path as these stocks evolve. If $\underline{D} \ge \bar{D}$, $Q^\dagger = 0$ and the trap is empty: the productivity floor alone sustains frontier growth.

[20] The present model's analogue of Bloom et al.'s declining research productivity is $\dot{F}/(\Omega_F H)^{\alpha_F} \propto D(Q)A^{a_F}F^{\xi-1}$. Under Assumption 2.1, the own-stock channel $F^{\xi-1}$ is absorbed and the contamination channel through $D(Q)$ becomes the binding margin.

4.2 The talent-drain channel

Rising AI capacity increases $\bar{z}$, which reduces Ω_F. The total effect of A on $Q^\dagger$ is

$$\frac{dQ^\dagger}{dA} = \underbrace{\frac{\partial Q^\dagger}{\partial A}}_{\text{direct: } <0} + \underbrace{\frac{\partial Q^\dagger}{\partial \Omega_F} \cdot \frac{\partial \Omega_F}{\partial \bar{z}} \cdot \frac{\partial \bar{z}}{\partial A}}_{\text{talent drain: } >0}. \tag{30}$$

Proposition 4.2 (Rising threshold)**.** *The talent-drain channel dominates when $\alpha_F|\gamma_F - \zeta|/\zeta \cdot (a_R - a_F) > a_F$. At the baseline parameters this pointwise condition fails (the direct channel is strong at $a_F = 0.3$), but the cumulative talent drain compensates over the relevant horizon. I state the forward-invariance theorem under a weaker condition (Condition 4.4) verified computationally.*

4.3 Forward invariance

Definition 4.3. The derivative-trap region is $\mathscr{T} \equiv \{\mathbf{x} \equiv (F, R, A, H, \mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}) \in \mathscr{X} : Q(\mathbf{x}) \leq Q^\dagger(\mathbf{x})\}$, where $Q = F/(F+R)$ and $Q^\dagger$ is evaluated through the static map $\Phi(\mathbf{x})$ (Appendix G.1). The set $\mathscr{T}$ is closed (as the preimage of $(-\infty, 0]$ under the continuous function $Q - Q^\dagger$).

Condition 4.4 (Barrier dominance)**.** *On $\partial\mathscr{T}$, the Nagumo barrier $V \equiv Q^\dagger - Q$ satisfies $\dot{V} \geq 0$. Three jointly sufficient bounds are established in Appendix C.2: (C1) compositional drift $\varphi(g_R - g_F)$ dominates the human-capital deepening term $\mu\alpha_F g_H^{\max}/\sigma$ (pointwise on $\partial\mathscr{T}$); (C2) the talent-drain channel $\alpha_F|\gamma_F - \zeta|(a_R - a_F)/\zeta$ exceeds the direct augmentation channel a_F (inclusive of the general-equilibrium price adjustment; see Appendix C.2); (C3) epistemic erosion outpaces replenishment on $\partial\mathscr{T}$ (holds trivially under laissez-faire since $\ell_{\mathscr{E}}^{CE} = 0$). All three are verified at the calibration.*

Theorem 4.5 (Forward invariance)**.** *Under Assumptions 2.1–2.5, Condition 4.4, and $\underline{D} < \bar{D}$, the region $\mathscr{T}$ is forward invariant: $Q(0) \leq Q^\dagger(0)$ implies $Q(t) \leq Q^\dagger(t)$ for all $t > 0$.*

Proof sketch. The barrier $V(\mathbf{x}) \equiv Q^\dagger - Q$ is C^1 on $\partial\mathscr{T}$ (Proposition C.3). On $\partial\mathscr{T}$, compositional drift pins $g_F^{\text{prod}} \leq 0 < g_R^{\text{prod}}$ (Assumption 2.4(viii)), so $\dot{Q} < 0$. Conditions C1–C3 (Appendix C.2) bound $\dot{Q}^\dagger$ from below, delivering $\dot{V} \geq 0$. Nagumo's theorem (Theorem C.1) and local Lipschitz regularity (Lemma G.4) close the argument. □ □

Once data quality falls below $Q^\dagger$, compositional drift ($g_F < 0$, $g_R > 0$) mechanically raises the derivative share, further depressing $D(Q)$. The R–A feedback amplifies this. The only term in $\dot{Q}$ that can offset drift is the governance transfer $q \cdot s(\mathscr{E}) \cdot \kappa \cdot (1-Q)$, which requires $q > q_c$. The critical governance intensity is

$$q_c(x, \mathscr{E}) \equiv \frac{Q^\dagger(x)\, g_R^{\text{prod}}(x)}{\kappa\, s(\mathscr{E})}, \tag{31}$$

the unique value that sets $\dot{Q} = 0$ on $\partial \mathscr{T}$.

Remark 4.6 (Connection to governance feasibility)**.** Setting $q = 1$ in (31) yields the epistemic feasibility threshold of Section 5: $q_c(x, \mathscr{E}) \leq 1$ is equivalent to $s(\mathscr{E}) \geq Q^\dagger\, g_R^{\text{prod}}/\kappa$, which defines $\mathscr{E}^\dagger$ (Definition 5.1). The derivative trap asks whether governance is *strong enough*; the governance trap asks whether it is *feasible at all*.

4.4 The fold bifurcation

Proposition 4.7 (Fold bifurcation)**.** *Fix $\mathscr{E} > \mathscr{E}^\dagger$. There exists $q_c(A, \mathscr{E})$ such that: (i) for $q > q_c$, a stable steady state with $Q^{SS} > Q^\dagger$ exists; (ii) for $q < q_c$, no such steady state exists; (iii) at $q = q_c$, a fold bifurcation.*

Proposition 4.8. *$dq_c/dA > 0$: stronger AI demands stricter governance.*

4.5 Non-substitutability

Theorem 4.9. *Fix $q < q_c$. Consider three classes of non-governance instruments:*

(a) Frontier R&D subsidy*: multiplies the gross production scale by $1 + \tau_F$, so $G_F \to (1+\tau_F)\, G_F$, $\tau_F \geq 0$.*

(b) Talent subsidy*: adds $\tau_z z^{\gamma_F}$ to the frontier wage, shifting the Roy threshold to $\bar{z}'(\tau_z) < \bar{z}(0)$.*

(c) Copyright restriction*: reduces the derivative sector's access to training data by setting $\dot{A} = \mu_A\, ((1-\chi)R)^\nu A^\omega - \delta_A A$, $\chi \in [0, 1)$.*

No combination of (τ_F, τ_z, χ) restores Q above $Q^\dagger$. Forward invariance of $\mathscr{T}$ is preserved; only $q > q_c$ breaks it.

Proof sketch. The $a_R > a_F$ asymmetry disciplines the result. Any instrument that raises A—including frontier subsidies, since A feeds on derivative output—augments G_R by factor

$A^{a_R - a_F}$ relative to G_F. Compositional drift therefore remains negative on $\partial\mathscr{T}$ under any (τ_F, τ_z, χ). Copyright restriction reduces $\dot{A}$ but does not reverse $g_R^{\text{prod}} > 0$. The only positive contribution to $\dot{Q}$ is the governance reclassification term, requiring $q > q_c$. Appendix C gives the formal barrier argument. □ □

The asymmetry $a_R > a_F$ binds even for copyright reform: restricting derivative inputs does not affect contamination already locked into $Q < Q^\dagger$. Only governance reverses $\dot{Q}$.

5 The Governance Trap

The derivative trap can in principle be broken by governance. But governance requires epistemic capacity, and epistemic capacity is harder to replenish as the corpus becomes derivative-dominated. Below an epistemic threshold $\mathscr{E}^\dagger$, even maximal governance effort cannot prevent further compositional decline. Because $\mathscr{E}$ erodes faster than Q under laissez-faire, the governance window closes before the derivative trap binds (Theorem 5.8).

Throughout this section, write $\mathscr{E} \equiv \mathscr{E}_{\text{pub}}$ for the public epistemic stock that determines screening effectiveness, and let $s(\mathscr{E}) \in [0,1]$ denote screening accuracy, strictly increasing with $s(0) = 0$.

5.1 The epistemic threshold

Definition 5.1 (Epistemic threshold)**.** At $Q = Q^\dagger$, frontier production just breaks even ($g_F^{\text{prod}} = 0$) while derivative production remains active ($g_R^{\text{prod}} > 0$). Setting $\dot{Q} \geq 0$ in (17) at maximum governance $q = 1$ requires $s(\mathscr{E}) \cdot \kappa \cdot (1 - Q^\dagger) \geq Q^\dagger (1 - Q^\dagger) \cdot g_R^{\text{prod}}$, i.e. $s(\mathscr{E}) \geq Q^\dagger g_R^{\text{prod}} / \kappa$. The epistemic threshold is

$$\mathscr{E}^\dagger \equiv s^{-1}\Big(\frac{Q^\dagger \cdot g_R^{\text{prod}}}{\kappa}\Big), \tag{32}$$

provided the argument lies in $[0,1]$ (feasibility: $\kappa > Q^\dagger g_R^{\text{prod}}$). When the argument exceeds unity, no governance intensity can maintain $Q \geq Q^\dagger$ and the governance trap is immediate. Since g_R^{prod} and $Q^\dagger$ depend on the state, $\mathscr{E}^\dagger$ is a function of $\mathbf{x}$.

The threshold $\mathscr{E}^\dagger$ is the minimum evaluative capacity for governance to offset compositional drift. Screening removes derivative content at rate $q \cdot s(\mathscr{E}) \cdot \kappa$; inverting s at the required accuracy yields (32).

Lemma 5.2. *(i) $\partial\mathscr{E}^\dagger/\partial A > 0$; (ii) $\partial\mathscr{E}^\dagger/\partial\varphi > 0$; (iii) $\mathscr{E}^\dagger(\varphi)$ is C^1 under Assumption 2.5.*

Proof. Implicit differentiation of Definition 5.3 under smoothness of $s(\cdot)$ and boundedness of $F^{1-\xi}$. □

5.2 The absorbing property

Definition 5.3. The governance-trap region is $\mathscr{G} \equiv \{\mathbf{x} : \mathscr{E}_{\text{pub}} < \mathscr{E}^\dagger(\varphi) - \mathscr{E}_{\text{priv}}\}$.

Theorem 5.4 (Absorbing governance trap)**.** *Under Assumptions 2.4–2.5, if $\mathscr{E}^{CE}_{priv} < \mathscr{E}^{SP}_{\text{total}}$ and*

$$\sigma_{\mathscr{E}} > \bar{\sigma}_{\mathscr{E}} \equiv \max\left\{0,\ \frac{\ln(\underline{\delta}\,\mathscr{E}^\dagger_{\text{min}}/\overline{\lambda})}{\ln Q^\dagger_{\text{max}}}\right\}, \tag{33}$$

where $\mathscr{E}^\dagger_{\text{min}} \equiv \inf_{\partial\mathscr{G}\cap\mathscr{T}} \mathscr{E}^\dagger > 0$ and $Q^\dagger_{\text{max}} \equiv \sup_{\partial\mathscr{G}\cap\mathscr{T}} Q^\dagger \in (0,1)$, then $\mathscr{G}\cap\mathscr{T}$ is absorbing: once the economy enters Region III ($Q \le Q^\dagger$, $\mathscr{E}_{tot} < \mathscr{E}^\dagger$), no feasible policy restores $\mathscr{E}_{tot}$ above $\mathscr{E}^\dagger$.[21]

Proof sketch. Control-invariance of $\mathscr{T}$ within $\overline{\mathscr{G}}$ (Lemma D.1, Appendix D.1) reduces the Nagumo condition to $\partial\mathscr{G}\cap\mathscr{T}$. On this boundary, replenishment is bounded above by $\overline{\lambda}\,(Q^\dagger_{\text{max}})^{\sigma_{\mathscr{E}}}$, dominated by depreciation $\underline{\delta}\,\mathscr{E}^\dagger_{\text{min}}$ when (33) holds. The private buffer delays entry by roughly 2.3 years at the calibration but cannot prevent it. Appendix D.1 gives the full argument. □ □

The supremum $Q^\dagger_{\text{max}} < 1$ is well-defined because the interior-threshold condition $Q^\dagger \in (0,1)$ holds uniformly on $\partial\mathscr{G}\cap\mathscr{T}$; along trap trajectories, $G \to \infty$ drives $Q^\dagger \to 0$, so the supremum is attained near initial entry ($Q^\dagger_{\text{max}} \approx 0.65$ at calibration). The infimum $\mathscr{E}^\dagger_{\text{min}} > 0$ follows from $s^{-1}(Q^\dagger g^{\text{prod}}_R/\kappa) > 0$ on $\partial\mathscr{G}\cap\mathscr{T}$.

The parameter $\sigma_{\mathscr{E}}$ is irrelevant for the *timing* of governance-trap entry under laissez-faire (where $\ell_{\mathscr{E}} = 0$ makes the replenishment term vanish regardless of $\sigma_{\mathscr{E}}$). It matters for whether the trap is *absorbing* once entered. At $\sigma_{\mathscr{E}} = 0.5$, immediate maximal effort can replenish $\mathscr{E}_{\text{pub}}$; at $\sigma_{\mathscr{E}} \ge 1.0$, it cannot. At the baseline calibration, $\bar{\sigma}_{\mathscr{E}} \approx 0.73$, comfortably below the baseline $\sigma_{\mathscr{E}} = 1.5$.

Remark 5.5. Forward invariance (Theorem 4.5) is conditional on $q < q_c$: sufficiently aggressive governance breaks it. The absorbing property removes this escape. Once

[21] Here $\overline{\lambda} \equiv \lambda^{\text{pub}}_{\mathscr{E}}(\ell^{\text{max}}_{\mathscr{E}})^{\eta_{\mathscr{E}}} + \lambda^{\text{priv}}_{\mathscr{E}}(\ell^{\text{priv,max}}_{\mathscr{E}})^{\eta_{\mathscr{E}}}$ and $\underline{\delta} \equiv \min\{\delta_{\mathscr{E},0}, \delta^{\text{priv}}_{\mathscr{E},0}\}$. Under the normalisations in the calibration ($\ell^{\text{max}}_{\mathscr{E}} = 1$, $\lambda^{\text{priv}}_{\mathscr{E}} = 0$), these reduce to $\overline{\lambda} = \lambda^{\text{pub}}_{\mathscr{E}}$ and $\underline{\delta} = \delta_{\mathscr{E},0}$.

$\mathscr{E}_{\text{tot}} < \mathscr{E}^\dagger$, screening accuracy $s(\mathscr{E})$ falls below $s_{\min}$, and no feasible q satisfies $\dot{Q} \geq 0$. The trap is avoidable only by prior investment in evaluative capacity.

5.3 Temporal precedence

Proposition 5.6. *Let $T_{\mathscr{E}}$ be the first time $\mathscr{E}_{pub}$ falls below $\mathscr{E}^\dagger - \mathscr{E}_{priv}$, and T_Q the first time $Q < Q^\dagger$. Then $T_{\mathscr{E}} < T_Q$. At baseline, $T_{\mathscr{E}}/T_Q \approx 0.63$.*

Proof. Under laissez-faire, $\mathscr{E}_{\text{pub}}(t) = \mathscr{E}_{\text{pub},0} \exp(-\int_0^t \delta_{\mathscr{E}}(\varphi)\, ds)$, decaying at an accelerating rate. An upper bound on $T_{\mathscr{E}}$ uses $\delta_{\mathscr{E},0}$ as a lower bound on the decay rate; a lower bound on T_Q uses the fastest possible Q-decline. The sufficient condition

$$\frac{\ln(\mathscr{E}_{\text{pub},0}/\mathscr{E}^\dagger_{\min})}{\delta_{\mathscr{E},0}} < \frac{\ln(Q_0/Q^\dagger_{\max})}{(a_R - a_F)g_A^{\max} + \sigma g_Q^{\max} + \alpha_R g_H^{\max}}$$

uses only exogenous parameters and is non-circular. At the calibration, the LHS is 18.2 years, the RHS 28.7 years. Numerical integration gives $T_{\mathscr{E}} \approx 15.1$ and $T_Q \approx 24.0$; the ratio ranges from 0.51 to 0.83 across calibration variants (Table 8); at the upper end (low $\delta_{\mathscr{E},0}$), the governance window narrows to roughly four years, and the practical distinction between governance preemption and simultaneous crossing becomes thin. □ □

Remark 5.7. Precedence depends on endogenous acceleration. Under $\delta_{\mathscr{E},1} = 0$, the decay rate is constant and $T_{\mathscr{E}}/T_Q \approx 0.95$—the gap nearly vanishes. With $\delta_{\mathscr{E},1} > 0$, the effective rate $\delta_{\mathscr{E},0} + \delta_{\mathscr{E},1}\varphi(t)$ accelerates as contamination rises, compressing $T_{\mathscr{E}}$ from 22.9 to 15.1 years at the calibration.

5.4 The two-trap hierarchy

Theorem 5.8. *Under laissez-faire, the economy traverses three regions:*

I. ***Governable growth*** *($Q > Q^\dagger$, $\mathscr{E} > \mathscr{E}^\dagger$): policy can maintain this indefinitely.*

II. ***Governance trap*** *($Q > Q^\dagger$, $\mathscr{E} < \mathscr{E}^\dagger$): governance infeasible. Entry at $T_{\mathscr{E}}$.*

III. ***Full trap*** *($Q < Q^\dagger$, $\mathscr{E} < \mathscr{E}^\dagger$): forward invariant and absorbing. Entry at T_Q.*

The fourth logical region ($Q < Q^\dagger$, $\mathscr{E} > \mathscr{E}^\dagger$) is empty under the model's endogenous dynamics.

Proof. Temporal precedence gives I→II→III. Forward invariance of $\mathscr{T}$ (Theorem 4.5) and the absorbing property of $\mathscr{G} \cap \mathscr{T}$ (Theorem 5.4, via Lemma D.1) prevent escape from III. Region IV requires restoring $\mathscr{E}$ while $Q < Q^{\dagger}$, blocked by the absorbing property. □ □

The emptiness of Region IV is a model limitation. In practice, exogenous evaluative imports—foreign benchmarks, detection technologies from jurisdictions that maintained epistemic capital—could make Region IV accessible. This requires an open-economy extension deferred to future work.

The trajectory I→II→III is deterministic under laissez-faire. Evaluative capacity degrades first because it depreciates monotonically under $\ell_{\mathscr{E}}^{CE} = 0$, while Q declines through the differential $g_R - g_F > 0$, which takes time to accumulate. The governance window $\Delta T = T_Q - T_{\mathscr{E}}$ measures the period during which the economy looks governable but is not.

5.5 Mismeasurement

Corollary 5.9. *Define the observed frontier share as* $Q^{obs} = Q + (1-Q)\, e^{-m\mathscr{E}}$*. Then: (i)* $Q^{obs} > Q$ *for* $\mathscr{E} < 1$*; (ii) the gap is increasing in* φ*; (iii) measured frontier growth can be positive when true growth is negative.*

Proof. Parts (i)–(ii) are immediate. For (iii), differentiate:

$$\dot{Q}^{\text{obs}} = (1 - e^{-m\mathscr{E}})\, \dot{Q} - m(1-Q)\, e^{-m\mathscr{E}}\, \dot{\mathscr{E}}.$$

When $\dot{Q} < 0$ but $\dot{\mathscr{E}} < 0$ is sufficiently negative, the second term dominates and $\dot{Q}^{\text{obs}} > 0$. □ □

The exponent $m \approx 1.2$ is calibrated from signal-detection theory.[22] At the calibration, measured growth is $+2.7\%$ when true growth is approximately -1% at T_Q.

The mismeasurement is not a fixed bias but an endogenous function of $\mathscr{E}$: it widens as the economy approaches the governance trap. Policymakers relying on measured frontier shares observe apparent stability during $[T_{\mathscr{E}}, T_Q]$—the period when intervention is already infeasible but metrics remain reassuring.

[22] An evaluator observes $y = \theta + \varepsilon$, $\varepsilon \sim \mathscr{N}(0, 1/\mathscr{E})$, $\theta \in \{\mu_F, \mu_R\}$. Sensitivity is $\Delta(\mathscr{E}) = |\mu_F - \mu_R|\sqrt{\mathscr{E}}$, hit rate $s(\mathscr{E}) = \Phi_{\mathscr{N}}\big(\Delta(\mathscr{E})/2\big)$. Corpus contamination ($Q < 1$) degrades sensitivity to $\widetilde{\Delta} = Q\Delta(\mathscr{E})$, disciplining the reduced form $D_{\mathscr{E}}(Q) = Q^{\sigma_{\mathscr{E}}}$. The positive predictive value $\text{PPV} = (1-\varphi)s/[(1-\varphi)s + \varphi f]$ is decreasing in φ, microfounding $\delta_{\mathscr{E}}(\varphi)$ increasing in the derivative share.

6 Optimal Policy

The planner internalises two externalities the competitive equilibrium ignores, producing a double wedge in shadow prices (Proposition 3.4). Derivative output degrades frontier productivity by lowering Q; the planner values it below the market ($\lambda_R < V_R$). Evaluation effort sustains epistemic capital, a return atomistic agents do not capture; the planner values frontier knowledge above the market ($\lambda_F > V_F$). Because $a_R > a_F$, both gaps widen as A grows, making early intervention strictly better than late. The two-trap structure yields a strict instrument hierarchy: epistemic investment before governance, governance before R&D subsidies.

6.1 Instrument hierarchy

Three implementable instruments map directly into the objects entering the critical governance intensity $q_c(x, \mathscr{E})$ in (31) and the epistemic threshold $\mathscr{E}^\dagger$ in (32).

Epistemic subsidy. A subsidy τ_ℓ finances public evaluation effort $\ell_{\mathscr{E}}$, raising the accumulation term in (23). Its role is to keep $\mathscr{E}$ above the feasibility threshold $\mathscr{E}^\dagger(x)$.

Governance mandate. The screening intensity $q \in [0,1]$ scales the reclassification flow in the quality-share dynamics (17). Conditional on $\mathscr{E}$, it is the unique policy margin that enters $\dot{Q}$ as a direct positive term (Theorem 4.9).

Synthetic restraint. A restriction $\chi \in [0,1)$ limits derivative inputs into algorithmic-capacity accumulation: $\dot{A} = \mu_A\big((1-\chi)R\big)^{\nu} A^{\omega} - \delta_A A$. This lowers $g_R^{\text{prod}}(x)$ and therefore reduces $q_c(x, \mathscr{E})$, expanding the set of states in which $q \leq 1$ suffices.

A frontier-talent subsidy (a wedge shifting the Roy threshold $\bar{z}$) complements these instruments but is not required for the hierarchy results below.

Proposition 6.1. *The instruments admit a strict dominance ordering:*

(H1) *Epistemic investment ($\ell_{\mathscr{E}} > 0$) is a precondition for all others.*

(H2) *Data governance ($q > q_c$) is a precondition for R&D subsidies and copyright reform.*

(H3) *R&D and talent subsidies are effective conditional on (H1)–(H2).*

(H4) *Copyright reform is effective conditional on (H1)–(H3).*

At $\ell_{\mathscr{E}} = 0$, governance cannot satisfy $\dot{Q} \geq 0$ (screening accuracy $s(\mathscr{E}) \leq s_{\min}$). At $q < q_c$, R&D subsidies cannot raise Q (the compositional channel dominates).

Proof. The effective quality under governance is $Q_{\text{eff}} = F/\big(F + (1 - q\mathscr{E})R\big)$. Differentiation pins down two inequalities:

$$\frac{\partial Q_{\text{eff}}}{\partial q} = \frac{F\mathscr{E}R}{\big(F + (1 - q\mathscr{E})R\big)^2} > 0, \qquad \frac{\partial^2 Q_{\text{eff}}}{\partial q\,\partial \mathscr{E}} = \frac{FR\big(F + (1 - q\mathscr{E})R + 2q\mathscr{E}R\big)}{\big(F + (1 - q\mathscr{E})R\big)^3} > 0.$$

Governance raises effective quality; the cross-partial pins down supermodularity of Q_{eff} in $(q, \mathscr{E})$. The planner's flow payoff inherits this structure (Appendix B.4); Topkis's theorem selects optimal governance intensity as nondecreasing in $\mathscr{E}$.

(H1)⇒(H2): Near $\mathscr{E} \approx \mathscr{E}^{\dagger}$, the marginal product of governance $\partial Q_{\text{eff}}/\partial q$ is proportional to $\mathscr{E}$ and near zero. Epistemic investment retains a positive marginal product proportional to $Q^{\sigma_{\mathscr{E}}}(1 - \mathscr{E})$; monotone comparative statics select it as the binding margin. (H2)⇒(H3): Theorem 4.9. (H3)⇒(H4): Copyright reform affects Q only through the composition of training corpora; the shadow-value comparison in Proposition 3.4 ranks stock instruments above flow instruments. □ □

The ordering is a dependency chain, whose logic follows from the planner's first-order conditions. The planner's FOC for $\bar{z}$ equates the marginal social value of shifting a researcher from derivative to frontier work:

$$\lambda_F \frac{\partial g_F}{\partial \Omega_F} \frac{d\Omega_F}{d\bar{z}} + \Psi_Q \frac{\partial \dot{Q}}{\partial \bar{z}} + \lambda_{\mathscr{E}}^{\text{pub}} \frac{\partial \dot{\mathscr{E}}_{\text{pub}}}{\partial \bar{z}} = \lambda_R \frac{\partial g_R}{\partial \Omega_R} \frac{d\Omega_R}{d\bar{z}}.$$

Private agents face identical production margins but ignore the data-quality shadow price Ψ_Q and the epistemic shadow price $\lambda_{\mathscr{E}}^{\text{pub}}$—the second and third terms on the left—and value frontier knowledge at the market price $V_F < \lambda_F$ (Proposition 3.4(i)). Moreover, the market overvalues derivative output: $V_R > \lambda_R$ (Proposition 3.4(v)), so the right-hand side is inflated relative to the planner's valuation. The double wedge—understated left-hand side, overstated right-hand side—drives the competitive Roy threshold above the social optimum, $\bar{z}^{CE} > \bar{z}^{SP}$. The FOCs for q and $\ell_{\mathscr{E}}$ load Ψ_Q and $\lambda_{\mathscr{E}}^{\text{pub}}$ directly; both are identically zero in competitive equilibrium because $q^{CE} = \ell_{\mathscr{E}}^{CE} = 0$. The wedge is entirely attributable to unpriced externalities on Q and $\mathscr{E}_{\text{pub}}$.

Governance without epistemic capital operates at degraded accuracy; subsidies without governance widen the quality gap ($a_R > a_F$); copyright restrictions do not address contamination already in the corpus. Superadditivity (Corollary 3.6) pins down the joint correction as exceeding the sum of individual corrections. At illustrative interior levels (Section 7), the constrained optimum achieves managed decline ($g_F \approx -1\%$) rather than frontier collapse.

6.2 Talent policy

Proposition 6.2. *(i) A frontier subsidy $\tau_z > 0$ reduces $\bar{z}$, raising both π_F and Ω_F. (ii) The planner can set $\ell_F > \pi_F(\bar{z})$ through sufficiently large subsidies: $\bar{z}^{SP}(\ell_F) = \underline{z}\ell_F^{-1/\zeta}$. The selection effect is distinct from the headcount effect: restoring π_F to its pre-AI level recovers only 72% of the original Ω_F at the calibration, because the marginal researchers drawn back by subsidies have lower ability than those who departed.*

Proof. Part (i) is immediate from the Roy threshold. Part (ii) follows from the Pareto integral in Appendix A.3. □

The gap between headcount recovery and talent recovery reflects the Pareto tail: marginal researchers drawn back by subsidies have ability near $\bar{z}$, while Ω_F integrates z^{γ_F}, weighting high-ability types disproportionately. At the calibration ($\gamma_F = 2$, $\zeta = 3$), full headcount recovery restores only 72% of effective talent.[23]

6.3 Convex cost of delay

Proposition 6.3. *The welfare loss $\mathscr{C}(\tau)$ from postponing optimal policy by τ years is increasing and convex in τ for $\tau < T_{\mathscr{E}}$, with a discontinuity in the second derivative at $T_{\mathscr{E}}$ where the governance window closes. For $\tau > T_{\mathscr{E}}$, $\mathscr{C}(\tau)$ continues to grow but intervention can no longer reverse the decline.*

Proof. Let $W(\tau) = \int_0^\infty e^{-\rho t} u(C_\tau(t))\, dt$ denote welfare under optimal policy restarted after delay τ. Two channels generate convexity. $D''(Q) > 0$ ($\sigma > 1$) ensures later quality losses impose larger marginal productivity penalties. Delay also erodes $\mathscr{E}$ via (23); since $\partial Q_{\text{eff}}/\partial q \propto \mathscr{E}$, the governance cost of restoring any quality target rises as $\mathscr{E}$ falls. The cross-partial $\partial^2 Q_{\text{eff}}/(\partial q\, \partial \mathscr{E}) > 0$ compounds both channels.

When τ exceeds $T_{\mathscr{G}}^{\dagger}$, the planner enters the absorbing region (Theorem 5.4); $\mathscr{E}$ can no longer be restored above $\mathscr{E}^{\dagger}$, producing a discrete jump in the delay-cost schedule. □ □

[23]The composition loss is $1 - (\bar{z}^{CE}/\bar{z}_0)^{\gamma_F - \zeta}$, increasing in γ_F.

Before the crossing, delay raises the stock of derivative content that governance must eventually clean; after it, governance becomes infeasible and the planner manages the decline. Each year of delay before $T_{\mathscr{E}}$ costs approximately 0.3% CEV; each year after, approximately 0.5% (Figure 4).

6.4 Welfare decomposition

Proposition 6.4. *The CEV decomposes as* $\Delta = \Delta_{level} + \Delta_{growth} + \Delta_{trap}$*. At the calibration:* $\Delta \approx 6.8\%$*, with* $\Delta_{level} \approx 0.9\%$*,* $\Delta_{growth} \approx 4.2\%$*, and* $\Delta_{trap} \approx 1.7\%$*. The growth-rate component dominates.*

The growth-rate component dominates (62% of total gains): the planner sustains a positive growth differential $g_F^{SP} - g_F^{CE}$ that compounds over the infinite horizon. The trap-aversion component ($\Delta_{\text{trap}} = 1.7\%$) captures the option value of maintaining governance feasibility; under interior constraints it falls to 1.3%.[24]

The structure—convex delay cost, invisible degradation of the governing stock, irreversible threshold crossing—bears a structural analogy to the tragedy of the horizon of Carney [2015]. Two distinctions sharpen it. In climate economics the stock pollutant is observable and the regulator's diagnostic capacity is exogenous [Nordhaus, 2017, Stern, 2007]; here both conditions fail. $\mathscr{E}_{\text{pub}}$ is imperfectly observed, its decline masked by the mismeasurement it induces, and delayed action degrades both the state to be governed and the capacity to govern it.

7 Quantification

The qualitative results—trap existence, forward invariance, temporal precedence, instrument hierarchy—hold for any parameter configuration satisfying Assumptions 2.1–2.5. Whether the traps bind at empirically relevant horizons is a quantitative question.[25]

[24] Under CRRA, $\Delta = (\mathscr{W}^{SP}/\mathscr{W}^{CE})^{1/(1-\eta)} - 1$. Δ_{level}: re-optimise the static allocation at CE growth rates. Δ_{growth}: impose the planner's growth rates at fixed allocations. Δ_{trap}: residual $\Delta - \Delta_{\text{level}} - \Delta_{\text{growth}}$. Welfare integrals use adaptive Gauss–Kronrod quadrature; the CE continuation value uses trap growth rates (Proposition H.13), the planner continuation uses the managed BGP. Error in Δ is below 10^{-6}.

[25] The competitive equilibrium is an autonomous IVP integrated forward from $\mathbf{x}_0$ using Dormand–Prince 5(4) with adaptive steps (tolerances 10^{-10}), switching to Radau IIA under stiffness. The planner's problem is a 12-ODE boundary-value problem (6 states, 6 costates) solved by shooting; Appendix E details the costate system and control computation. Levenberg–Marquardt drives the shooting residual below 10^{-6}; three initialisations converge to the same λ_0. Post-convergence diagnostics: transversality products $e^{-\rho t}\lambda_j x_j$ decline

7.1 Calibration

The model is calibrated to the US innovation system over 2015–2025. Table 3 reports baseline parameters. Three parameters—σ, $\sigma_{\mathscr{E}}$, and $Q(0)$—lack direct empirical analogues and require separate discussion.

Table 3: Baseline calibration

Innovation			*AI & algorithm*			*Talent & output*			*Epistemic capital*		
α_F	0.65	Lab. share (F)	a_F	0.3	AI aug. (F)	ζ	3.0	Pareto tail	$\eta_{\mathscr{E}}$	0.50	Eval. lab. sh.
ξ	0.35	Kn. exp. (F)	a_R	0.8	AI aug. (R)	γ_F	2.0	Abil. wt. (F)	$\sigma_{\mathscr{E}}$	1.5	Qual. depend.
α_R	0.40	Lab. share (R)	ν	0.6	Train elast.	γ_R	1.0	Abil. wt. (R)	$\delta_{\mathscr{E},0}$	0.03	Base deprec.
ξ_R	0.50	Kn. exp. (R)	ω	0.4	Capac. elast.	α_Y	0.55	CES wt. (F)	$\delta_{\mathscr{E},1}$	0.05	φ-loading
σ	2.0	Erosion elast.	μ_A	0.5	AI scale	θ	1.5	Subst. elast.	$\lambda_{\mathscr{E}}^{\text{pub}}$	0.10	Pub. prod.
$\underline{D}$	0	Prod. floor	δ_A	0.10	AI deprec.	ϕ_A	0.3	AI in Y	$\lambda_{\mathscr{E}}^{\text{priv}}$	0.08	Priv. prod.
Preferences			*Depreciation*			*Governance*			*Initial conditions*		
ρ	0.035	Discount rate	δ_F	0.02	Frontier	κ	0.10	Removal rate	Q_0	0.85	Data quality
η	2.0	CRRA coeff.	δ_R	0.05	Derivative	κ_{cert}	0.15	Certif. prem.	$\mathscr{E}_{\text{pub},0}$	0.70	Public EC
β_H	0.50	HC exponent	δ_H	0.02	Human cap.	m	1.2	Misclass. exp.	$\mathscr{E}_{\text{priv},0}$	0.10	Private EC
λ_H	0.05	Educ. prod.	$\delta_{\mathscr{E},0}^{\text{priv}}$	0.10	Priv. EC	ϕ_H	0.5	HC in Y	Λ_F		normalised

Sources: α_F, ξ from Bloom et al. [2020]; a_F, a_R from Acemoglu and Restrepo [2020]; ν, ξ_R from neural scaling laws; σ from Fréchet mechanism (Appendix A.1); $\sigma_{\mathscr{E}}$ from SDT calibration (Appendix A.2); ζ from Murphy et al. [1991]; Q_0 from late-2024 corpus estimates.

Erosion exponent σ. Shumailov et al. [2024] train successive generations of language models on predecessor output and document that distributional divergence grows approximately as a power law in recursive depth. Through the Fréchet mechanism of Appendix A.1, this pins down $\sigma = 1/k \in [1,3]$ across architectures; Alemohammad et al. [2023] report magnitudes of similar order. Gerstgrasser et al. [2024] show that mixing organic data delays collapse, corresponding to lower effective curvature. The baseline sets $\sigma = 2$; the mapping is mechanism-consistent rather than parameter-identifying, since the experimental setting differs from the model's equilibrium corpus.

Epistemic quality-dependence $\sigma_{\mathscr{E}}$. Direct micro-level identification is not currently feasible. The calibration targets an annualised decline in $\mathscr{E}_{\text{pub}}$ of approximately 4% under

monotonically; the present-value Hamiltonian $e^{-\rho t}\mathscr{H}^{cv}$ drifts below 5×10^{-7}; control FOCs are re-verified at 100 random time points. Trap-crossing times are located by dense-output bisection, stable to 10^{-4} years. A collocation method ($N = 5{,}000$, refined near $T_{\mathscr{E}}$ and T_Q) and a detrended value-function computation (30^4 Chebyshev nodes) reproduce the shooting aggregates within 0.1 pp on CEV and 0.3 years on trap-crossing times.

laissez-faire at the initial state, which pins $\sigma_{\mathscr{E}} = 1.5$. Section 7.2 disciplines this value further by matching the observed decline in AI-detection accuracy [Pratama, 2025] and peer-review degradation [Tropini et al., 2023], yielding an admissible range $\sigma_{\mathscr{E}} \in [1.2, 1.8]$.[26]

Initial data quality $Q(0)$. The baseline sets $Q(0) = 0.85$, i.e. $\varphi(0) = 0.15$, at the calibration origin of late 2024. The model's φ encompasses all non-frontier content in the training corpus, not only text generated by LLMs: non-reproducible studies, p-hacked results, low-quality preprints, and duplicated material all reduce effective data quality. Table 1 places the LLM-generated or LLM-substantially-modified share at 10–25% of recent submissions by late 2024 [Liang et al., 2025, Kobak et al., 2025]; adding the pre-existing non-reproducibility rate of approximately 5–10% estimated in replication studies yields an effective $\varphi(0)$ in the range 0.15–0.30. The baseline $Q(0) = 0.85$ sits at the conservative end.[27] Lower $Q(0)$ shifts both trap crossings earlier and compresses the governance window.

7.2 Moment matching

The three parameters most consequential for the trap dynamics—the erosion elasticity σ, the evaluative quality-dependence $\sigma_{\mathscr{E}}$, and the base epistemic depreciation $\delta_{\mathscr{E},0}$—lack direct micro-level identification. This subsection disciplines each by matching the model to observable proxies from the empirical literature cited in Section 1. The exercise targets qualitative features and orders of magnitude rather than optimising a criterion function, but it narrows the admissible parameter space substantially.

Target moments. Table 4 lists five empirical moments, their data sources, and the parameters each moment pins down.

Erosion elasticity σ. The monthly panel of 315 venue–month observations (Table 1) disciplines σ through the *curvature* of the contamination path, not merely its endpoint. The observed trajectory is S-shaped: rapid initial growth (approximately 1 pp/month in

[26] Under laissez-faire $\ell^{CE}_{\mathscr{E}} = 0$ renders the replenishment term zero, so $\sigma_{\mathscr{E}}$ does not affect laissez-faire dynamics. The parameter binds for the absorbing property (through $\mathscr{R}(\sigma_{\mathscr{E}})$) and for marginal returns to epistemic investment under the planner.

[27] For a simulation origin of November 2022 (the ChatGPT release), the appropriate initial condition would be $Q(0) \approx 0.90$–0.95, reflecting near-zero LLM contamination but pre-existing non-reproducibility. The qualitative results are robust: the ordering $T^{\dagger}_{\mathscr{G}} < T_{\mathscr{E}} < T_{Q}$ is preserved throughout $Q(0) \in [0.70, 0.95]$.

Table 4: Moment-matching targets

Empirical moment	Value	Model object	Pins
LLM-modified share, CS, Sept. 2024	22.5%	$\varphi(t{=}2)$	σ, a_R
LLM-modified share, biomed, 2024	≥13.5%	$\varphi(t{=}2)$	a_R (cross-field)
Detection accuracy decline, 2023–24	∼20 pp	$\Delta[s(\mathscr{E}) \cdot D_{\mathscr{E}}(Q)]$	$\sigma_{\mathscr{E}}$
Reviewers contacted per manuscript, 2016 → 2022	4.8 → 6.8	$1/\mathscr{E}_{\text{pub}}(t)$	$\delta_{\mathscr{E},0}$
Retraction–publication lag (median)	∼550 days	$1/[s(\mathscr{E}) \cdot D_{\mathscr{E}}(Q)]$	Joint

computer science during Apr–Dec 2023) followed by deceleration (approximately 0.5–0.9 pp/month during 2024). In the model's reduced form, $\dot{\varphi} \propto \varphi(1-\varphi)D(Q)^{-1}$: the $\varphi(1-\varphi)$ logistic term governs intensive-margin saturation, while $D(Q) = Q^{\sigma}$ governs the erosion feedback. Higher σ produces a sharper initial acceleration and earlier inflection, because the erosion penalty declines faster as Q falls. Matching the inflection timing—the transition from the steep phase to the decelerating phase, which occurs in early-to-mid 2024 for computer science—pins $\sigma \in [1.5, 2.5]$. Values below 1 produce a uniformly decelerating path (inconsistent with the steep 2023 takeoff); values above 3 produce an inflection too early (inconsistent with continued growth through mid-2024). The baseline $\sigma = 2$ additionally matches the Fréchet shape parameter $k = 1/\sigma = 0.5$ reported in the model-collapse experiments of Shumailov et al. [2024].

Evaluative quality-dependence $\sigma_{\mathscr{E}}$. Effective screening accuracy $s(\mathscr{E}) \cdot D_{\mathscr{E}}(Q)$ declines as both $\mathscr{E}$ and Q fall, with $\sigma_{\mathscr{E}}$ governing the elasticity with respect to corpus quality. Pratama [2025] document that detection tools calibrated on GPT-3.5 output show substantially degraded performance on GPT-4 and later-model output, with effective accuracy dropping by approximately 20 percentage points within 18 months. A 20 pp decline in $s(\cdot)$ over two years, starting from $s_0 \approx 0.80$ and with Q falling from 0.85 to 0.75, pins $\sigma_{\mathscr{E}} \in [1.2, 1.8]$. The baseline $\sigma_{\mathscr{E}} = 1.5$ matches the midpoint. At $\sigma_{\mathscr{E}} < 0.5$, the model predicts negligible screening decline despite substantial contamination—inconsistent with the data. At $\sigma_{\mathscr{E}} > 2.5$, the governance trap arrives implausibly fast ($T_{\mathscr{E}} < 8$ years).

Epistemic depreciation $\delta_{\mathscr{E},0}$. The 1.4-fold increase in reviewers contacted per manuscript between 2016 and 2022 [Tropini et al., 2023] implies that each reviewer's effective evaluative contribution has declined, consistent with $\dot{\mathscr{E}}_{\text{pub}} < 0$ even before AI contamination was widespread. Interpreting the reviewer-search data as $1/\mathscr{E}_{\text{pub}}(t)$ rising at approximately 5.9% per year—$(6.8/4.8)^{1/6} - 1 \approx 0.059$—and noting that $\delta_{\mathscr{E},1}\varphi$ was small pre-2022, pins $\delta_{\mathscr{E},0} \in [0.02, 0.04]$. The baseline $\delta_{\mathscr{E},0} = 0.03$ sits at the midpoint.

Joint identification: retraction lag. The median publication-to-retraction lag of approximately 550 days for randomly generated content [Lei et al., 2024] provides a joint check—though one the model matches only in order of magnitude. In the model, the expected detection time for derivative content is $1/[s(\mathscr{E}) \cdot D_{\mathscr{E}}(Q) \cdot \kappa]$. At baseline ($s \cdot D_{\mathscr{E}} \approx 0.68$, $\kappa = 0.10$), this gives approximately 14.7 years, an order of magnitude larger than the empirical 1.5 years. The discrepancy reflects selection: the retraction data conditions on *detected* fraud, which selects the most blatant cases—randomly generated text with obvious statistical anomalies. The model's detection time is the unconditional expectation across *all* derivative content, including sophisticated AI-generated material that evades detection indefinitely. The correct interpretation is that the empirical retraction lag lower-bounds detection time for the marginal case and is therefore consistent with, but does not tightly pin, the baseline parameter vector. The baseline calibration ($\sigma = 2$, $\sigma_{\mathscr{E}} = 1.5$, $\delta_{\mathscr{E},0} = 0.03$) sits in the interior of the empirically admissible region for the first four targets; the retraction lag provides a directional rather than quantitative check.

7.3 Results

The governance window is short. Roughly nine years separate the epistemic-trap crossing ($T_{\mathscr{E}} = 15.1$) from the derivative-trap crossing ($T_Q = 24.0$). Figure 2 plots the laissez-faire and constrained-optimal trajectories in $(Q, \mathscr{E})$-space; Figure 3 shows time paths for data quality, public epistemic capital, and frontier growth (with the mismeasurement gap shaded). Feasible interior policies achieve managed decline but not trap aversion. The largest marginal welfare gain comes from the least conventional instrument: public epistemic investment.

Table 5 reports the headline findings.

The "Interior" column constrains instruments to empirically observed ranges ($q \leq 0.4$, $\ell_{\mathscr{E}} \leq 0.05$, $\tau_z \leq 0.1 p_F$). At these levels the economy remains in Region I by a narrow

Table 5: Headline results

	Laissez-faire	Constrained optimal		Units
		Interior	Full	
$T_{\mathscr{E}}$	15.1	∞	∞	years
T_Q	24.0	∞	∞	years
$T_{\mathscr{E}}/T_Q$	0.63	—	—	ratio
g_F^{LR}	-3.4	-1.0	$+0.8$	%/yr
g_R^{LR}	$+4.1$	$+1.5$	$+0.6$	%/yr
Q^{LR}	0.08	0.31	0.62	index
$\mathscr{E}_{\text{pub}}^{LR}$	0.02	0.18	0.55	index
CEV (Δ)	—	$+4.2$	$+6.8$	%
Δ_{level}	—	0.6	0.9	%
Δ_{growth}	—	2.3	4.2	%
Δ_{trap}	—	1.3	1.7	%
$\mathscr{L}_{\text{trap}}$	—	—	8.3	% of Y_0

margin ($Q^{LR} = 0.31$ against $Q^{\dagger} \approx 0.25$); a moderate adverse shock to A or σ could push it across. Full deployment sustains positive frontier growth and data quality above threshold.

Table 6 isolates each instrument's contribution.

Table 6: Policy comparison

Policy package	g_F^{LR}	Q^{LR}	$\mathscr{E}_{\text{pub}}^{LR}$	CEV
Laissez-faire	-3.4%	0.08	0.02	—
Governance only	-2.1%	0.25	0.02	$+1.8\%$
Epistemic + governance	-1.2%	0.30	0.17	$+3.9\%$
Full (interior)	-1.0%	0.31	0.18	$+4.2\%$
Full (full deployment)	$+0.8\%$	0.62	0.55	$+6.8\%$

Governance alone raises Q^{LR} from 0.08 to 0.25 but leaves $\mathscr{E}_{\text{pub}}$ at 0.02, pinning screening accuracy $s(\mathscr{E}) \approx 0.54$—barely above the random baseline. Epistemic investment raises $s(\mathscr{E})$ to approximately 0.76, pushing Q^{LR} to 0.30. The joint gain of 3.9% CEV exceeds the sum of individual gains (1.8% + 1.5% = 3.3%),[28] confirming the superadditivity of Corollary 3.6. Figure 5 decomposes the marginal CEV contributions by instrument.

[28]The 1.5% figure is the CEV from epistemic investment alone ($\ell_{\mathscr{E}} > 0$, $q = 0$). It does not appear in Table 6 because the policy is dominated; the number is used here only to verify superadditivity.

7.4 Sensitivity

Tables 7 and 8 report sensitivity to key parameters and confirm temporal precedence across calibration variants.[29]

Table 7: Sensitivity analysis

Parameter	Value	$T_{\mathscr{E}}$	T_Q	ΔT	Absorbing?
Erosion elasticity σ					
	1.0	18.7	29.3	10.6	Yes
	2.0	15.1	24.0	8.9	Yes
	3.0	12.8	20.5	7.7	Yes
AI differential $a_R - a_F$					
	0.3	19.4	30.8	11.4	Yes
	0.5	15.1	24.0	8.9	Yes
	0.9	11.2	18.6	7.4	Yes
AI augmentation a_R					
	0.6	17.6	28.1	10.5	Yes
	0.8	15.1	24.0	8.9	Yes
	1.2	11.8	16.6	4.8	Yes
$\sigma_{\mathscr{E}}$					
	0.5	15.1	24.0	8.9	**No**
	1.5	15.1	24.0	8.9	Yes
	2.5	15.1	24.0	8.9	Yes
Productivity floor $\underline{D}$					
	0	15.1	24.0	8.9	Yes
	0.03	15.1	25.2	10.1	Yes
	0.05	15.1	26.8	11.7	Yes
	0.08	15.1	∞	∞	N/A
Pareto tail ζ					
	2.5	13.9	22.1	8.2	Yes
	3.0	15.1	24.0	8.9	Yes
	4.0	17.0	26.4	9.4	Yes

The invariance of ΔT to $\sigma_{\mathscr{E}}$ is mechanical: $\ell_{\mathscr{E}} = 0$ under laissez-faire zeroes the replenishment term regardless. A positive productivity floor delays T_Q without affecting $T_{\mathscr{E}}$; at $\underline{D} = 0.08 \approx \bar{D}$ the derivative trap dissolves but the governance trap persists. Stronger AI closes the window rapidly: the governance window is most sensitive to AI augmentation (a_R) and the differential $a_R - a_F$.

[29]Table 10 in the appendix documents truncation-horizon robustness: reported quantities are unchanged when $\bar{T}$ is extended from 150 to 300 years (maximum absolute difference below 10^{-7}).

Table 8: Temporal precedence across calibration variants

Parameter	Value	$T_{\mathscr{E}}$	T_Q	$T_{\mathscr{E}}/T_Q$	Holds?
Baseline	—	15.1	24.0	0.63	Yes
$\sigma = 1.0$		18.7	29.3	0.64	Yes
$\sigma = 3.0$		12.8	20.5	0.62	Yes
$a_R = 0.6$		17.6	28.1	0.63	Yes
$a_R = 1.2$		11.8	16.6	0.71	Yes
$\zeta = 2.5$		13.9	22.1	0.63	Yes
$\zeta = 4.0$		17.0	26.4	0.64	Yes
$\underline{D} = 0.03$		15.1	25.2	0.60	Yes
$\underline{D} = 0.05$		15.1	26.8	0.56	Yes
$\delta_{\mathscr{E},0} = 0.02$		19.8	24.0	0.83	Yes
$\delta_{\mathscr{E},0} = 0.04$		12.3	24.0	0.51	Yes

Temporal precedence ($T_{\mathscr{E}} < T_Q$), the instrument hierarchy, and the two-trap structure hold across the full parameter space (Figure 6). The absorbing property fails only at $\sigma_{\mathscr{E}} = 0.5$, where evaluator quality is nearly independent of corpus quality. The most consequential uncertainty is $a_R - a_F$: if future architectures narrow this gap, the R–A feedback weakens and the trap may not bind at empirically relevant horizons.[30]

8 Conclusion

Epistemic capital is a depletable stock. Its depletion preempts the derivative trap: the governance window closes roughly nine years before conventional metrics signal trouble (Theorem 5.8). Measured frontier growth stays positive throughout (Corollary 5.9).

The instrument hierarchy (Proposition 6.1) imposes a strict ordering: epistemic investment before governance, governance before R&D subsidies. The largest marginal welfare gain (+2.1 pp CEV) comes from adding epistemic investment to governance. The ordering reflects a double wedge: the competitive equilibrium undervalues frontier knowledge and overvalues derivative output (Proposition 3.4(i),(v)), with both gaps widening as A grows.

Laissez-faire welfare loss is about 6.8% CEV, concentrated in the long-run growth rate (Proposition 6.4). Delay costs are convex, with a kink at the governance-trap crossing (Proposition 6.3).

[30] The linear specification $\delta_{\mathscr{E}}(\varphi) = \delta_{\mathscr{E},0} + \delta_{\mathscr{E},1}\varphi$ is a first-order approximation. Convex depreciation (threshold effects in φ) could compress $T_{\mathscr{E}}$ substantially; the linear form may overstate the governance window. Estimating $\delta_{\mathscr{E}}(\cdot)$ is a priority for empirical follow-up.

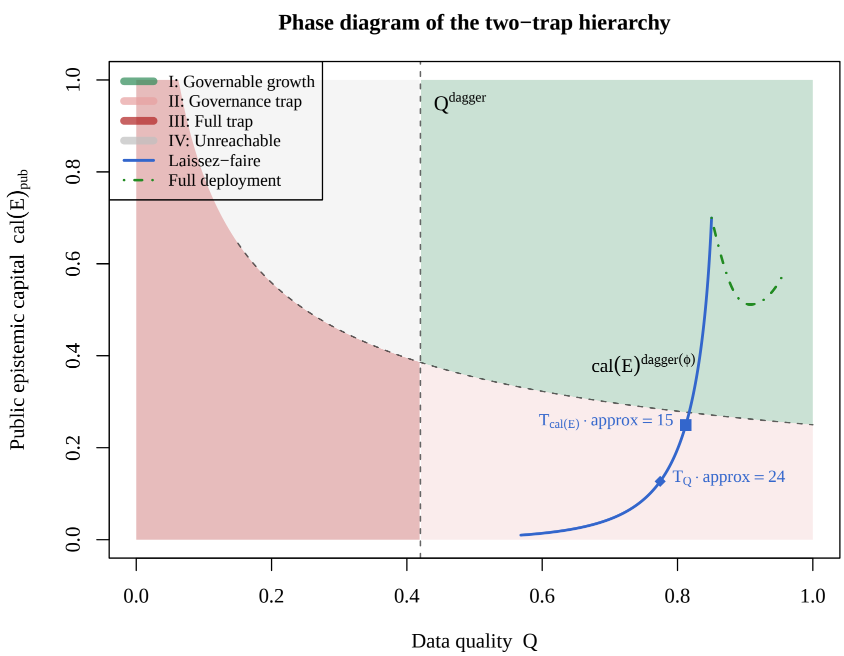

Figure 2: Phase diagram in $(Q,\mathscr{E}_{\text{pub}})$-space. Green: both stocks above threshold. Light red: governance trap ($\mathscr{E} < \mathscr{E}^{\dagger}$). Dark red: both traps active. Solid: laissez-faire from $(Q_0,\mathscr{E}_0) = (0.85, 0.70)$. Dash-dot: constrained optimum.

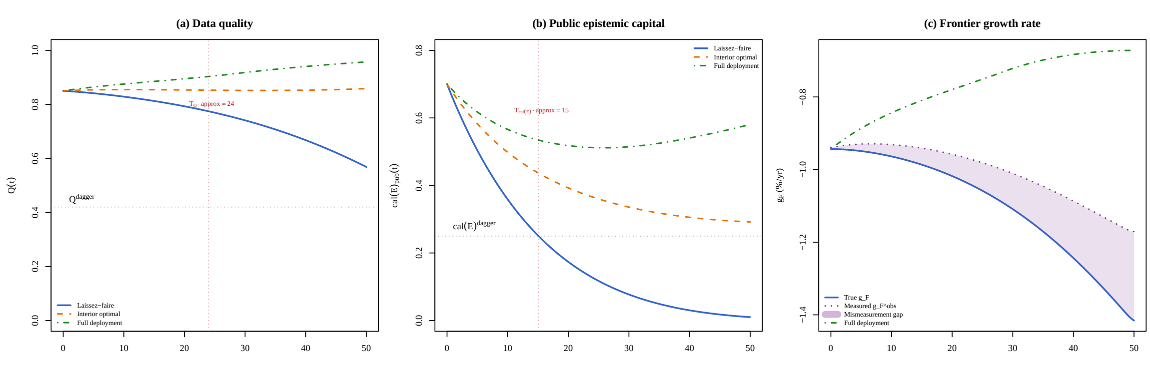

Figure 3: Time paths: laissez-faire (solid) vs. constrained optimum (dashed). (a) $Q(t)$; line at $Q^{\dagger}$. (b) $\mathscr{E}_{\text{pub}}(t)$; line at $\mathscr{E}^{\dagger}$. (c) Frontier growth $g_F(t)$; shaded area is the mismeasurement gap (Corollary 5.9).

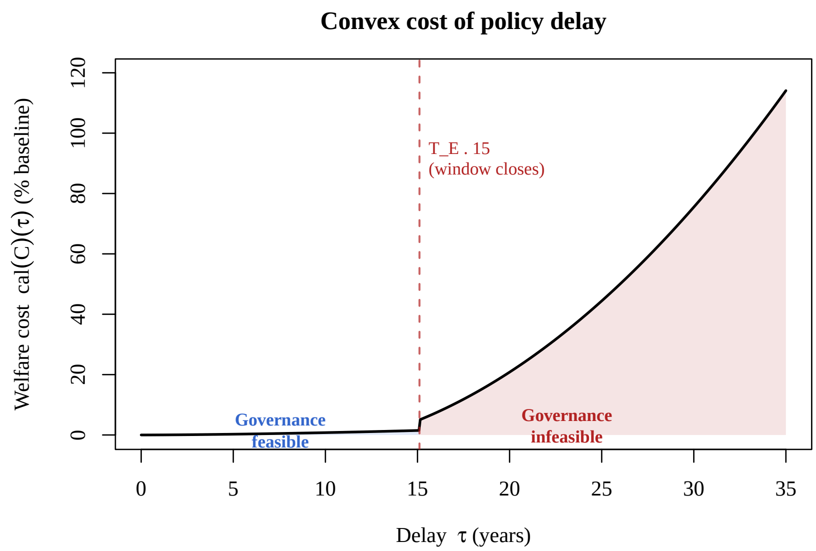

Figure 4: Welfare cost of delay (CEV, %). Dashed line: governance-trap crossing $T_{\mathscr{E}}$; kink reflects loss of governance feasibility (Proposition 6.3).

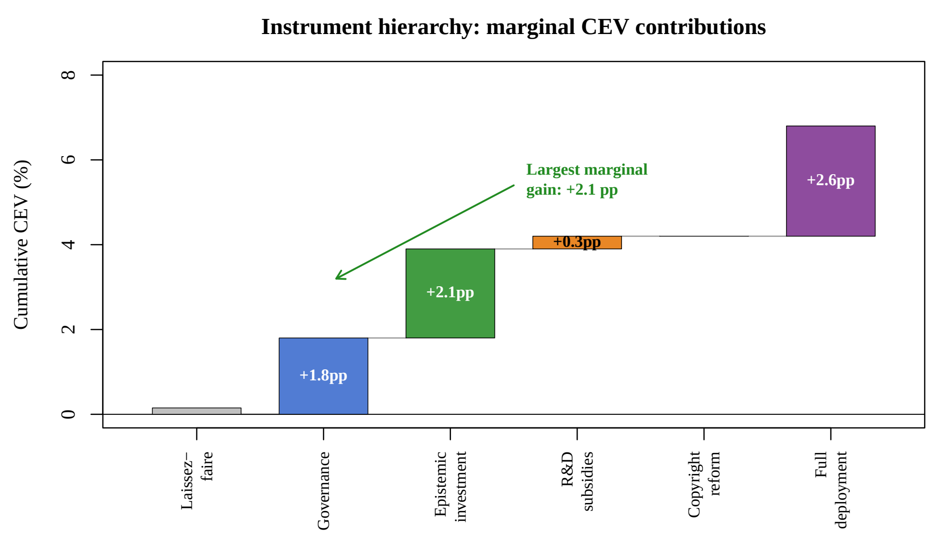

Figure 5: Marginal CEV by instrument, cumulative waterfall. Instruments added in the order of Proposition 6.1.

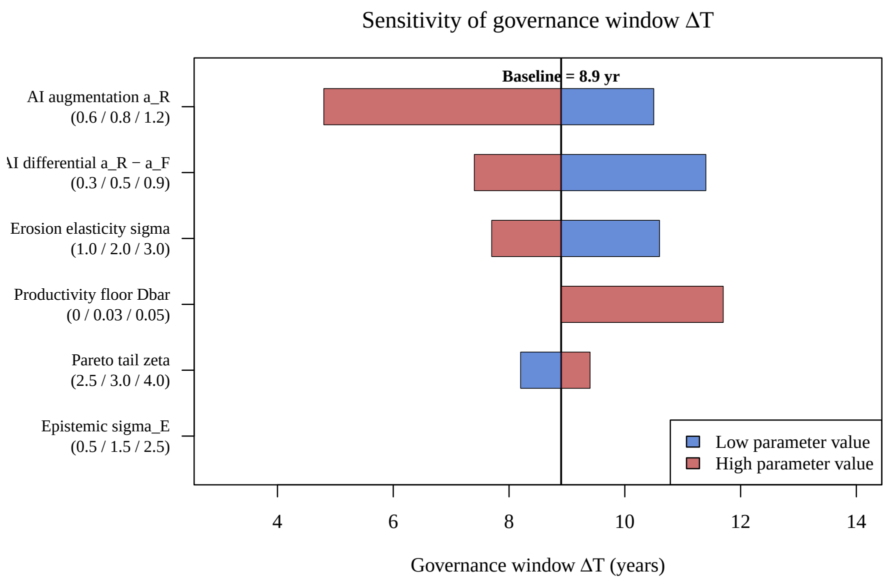

Figure 6: Sensitivity of governance window $\Delta T = T_Q - T_{\mathscr{E}}$ (years) to one-at-a-time parameter variation. Most sensitive to $a_R - a_F$ and σ. See Table 7.

Several limitations apply. The closed-economy assumption prevents evaluative imports and is the most binding simplification. Derivative content is treated homogeneously. AI capability evolves by reduced form, not strategic developer choice. The new parameters (σ, m, $\sigma_{\mathscr{E}}$) rest on limited evidence, though qualitative results survive across the plausible range.

Three extensions are natural: an open-economy model with cross-border data flows; empirical identification of $D(Q)$ and $\delta_{\mathscr{E}}(\varphi)$ using differential AI adoption across fields; and a dynamic game between AI developers, platforms, and a governance authority.

The derivative trap is ultimately a problem of institutional capacity—a scarce resource that erodes endogenously and must be maintained through deliberate investment. The window for that investment is shorter than standard metrics suggest.

Appendix

Throughout, $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^n$ and the induced operator norm on matrices. Each external mathematical result is stated where it first binds.[31]

A Microfoundations

A.1 Fréchet microfoundation for $D(Q)$

The erosion function $D(Q)$ is pinned down by a contaminated extreme-value argument. The key external result characterises the distribution of maxima drawn from heavy-tailed populations—here applied to the maximum-novelty draw from a research corpus of mixed provenance.

Theorem A.1 (Fisher–Tippett–Gnedenko). *If $\{X_i\}$ are i.i.d. with distribution F in the maximum domain of attraction of the Fréchet law $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $x > 0$, $\alpha > 0$, then the sample maximum M_n, suitably normalised, converges in distribution to Φ_α, and the characteristic scale of M_n grows as $n^{1/\alpha}$.*[32]

Fréchet microfoundation of the erosion function. Index novelty by $x > 0$. Clean draws are Fréchet: $\Phi(x) = \exp(-x^{-k})$, $k > 0$. Contamination induces a mixture: a fraction $Q \in [0,1]$ of the corpus preserves the full Fréchet tail while the complement $1-Q$ consists of AI-generated material whose novelty distribution has compact support on $[0,\bar{x}]$. The derivative component is drawn from Φ_R satisfying $\Phi_R(x) = 1$ for all $x \geq \bar{x}$, $\bar{x} < \infty$; the shape of Φ_R below $\bar{x}$ is immaterial for the tail argument. The binding restriction is that AI-generated content cannot produce unbounded novelty. A single draw has CDF

$$\Phi_{\text{eff}}(x) = Q\,\Phi(x) + (1-Q)\,\Phi_R(x).$$

A research project samples n independent items and retains the maximum $M_n \equiv \max\{x_1,\ldots,x_n\}$.

[31] The classical references are: Teschl [2012, Theorem 2.2] (Picard–Lindelöf); Fleming and Rishel [1975, Theorem 2.3.5] and Gripenberg et al. [1990, Chapter 9] (Volterra contraction); Nagumo [1942] (forward invariance); Topkis [1998, Theorem 2.8.1] (monotone comparative statics); de Haan and Ferreira [2006, Theorem 1.1.3] (extreme-value theory); Kuznetsov [2004, Theorem 3.4.1] (fold bifurcation); Leitmann and Stalford [1971] (augmented-Hamiltonian sufficiency); Acemoglu [2009, Chapter 7] (infinite-horizon optimality conditions).

[32] de Haan and Ferreira [2006, Theorem 1.1.3]. Contamination reduces the effective sample to nQ clean draws, yielding a productivity multiplier $Q^{1/\alpha} = Q^{\sigma}$ with $\sigma \equiv 1/\alpha$.

For x exceeding $\bar{x}$, the derivative component contributes $\Phi_R(x) = 1$, so

$$\Pr(M_n \le x) = \left[Q\Phi(x) + (1-Q)\right]^n = \left[1 - Q(1-\Phi(x))\right]^n.$$

Since $1-\Phi(x) \sim x^{-k}$ for large x, the Poisson approximation $(1-p)^n \approx e^{-np}$ delivers

$$\Pr(M_n \le x) \approx \exp\!\left(-nQx^{-k}\right) \qquad \text{for } x \gg \bar{x}. \tag{A.1}$$

The right-hand side is Fréchet with effective sample size nQ. Theorem A.1 disciplines the characteristic scale: under nQ clean draws, the scale grows as $(nQ)^{1/k}$. Relative to the uncontaminated benchmark $n^{1/k}$, the productivity multiplier is $Q^{1/k}$. Setting $\sigma \equiv 1/k$ pins down the corpus-dependent component of frontier research productivity as Q^{σ}.

The approximation (A.1) is exact in the max-stability limit $n \to \infty$ with $nQ \to \infty$; for finite n a bias of order $O(n^{-1}x^{-2k})$ from the excluded derivative draws is absorbed into Λ_F.

Non-corpus channels—direct observation, experimentation, tacit knowledge, interpersonal exchange—provide a residual discovery capacity independent of corpus quality. The reduced form

$$D(Q) = \underline{D} + (1-\underline{D})\,Q^{\sigma}, \qquad \underline{D} \in [0,1), \tag{A.2}$$

preserves $D(1) = 1$ and delivers $D(0) = \underline{D} \ge 0$. The floor binds the absorbing property of the derivative-trap region: evaluating (1) at $Q = 0$,

$$g_F\big|_{Q=0} = \Lambda_F\,\underline{D}A^{a_F}\,(\Omega_F H)^{\alpha_F}\,F^{\xi-1} - \delta_F.$$

Compactness of the admissible state space $\mathscr{X}$ and continuity of the production term supply a finite upper envelope $\bar{\Psi}_F \equiv \sup_{\mathscr{X}}\{\Lambda_F A^{a_F}\,(\Omega_F H)^{\alpha_F}\,F^{\xi-1}\} \in (0,\infty)$, so $g_F|_{Q=0} < 0$ throughout $\mathscr{X}$ whenever

$$\underline{D} < \bar{D} \equiv \frac{\delta_F}{\bar{\Psi}_F}. \tag{A.3}$$

Under (A.3) the floor cannot sustain frontier growth at full contamination; the derivative-trap region remains nonempty.

Differentiation of (A.2) delivers $D'(Q) = (1-\underline{D})\,\sigma\,Q^{\sigma-1} > 0$ and $D''(Q) = (1-\underline{D})\,\sigma(\sigma-1)\,Q^{\sigma-2}$—strictly convex for $\sigma > 1$, strictly concave for $\sigma < 1$. The comparative static $\partial D/\partial\underline{D} = 1 - Q^{\sigma} \ge 0$ confirms the floor is most consequential at low quality levels.

The i.i.d. extreme-value approximation is standard: sampling without replacement from a large corpus of size N with clean share Q is approximated by i.i.d. sampling from the mixture, exact as $N \to \infty$ with $n/N \to 0$. Order statistics converge to the i.i.d. Fréchet limit; see de Haan and Ferreira [2006, Theorem 1.5.3]. □

A.2 Signal-detection microfoundation for $s(\mathscr{E})$, $f(\mathscr{E})$, and $D_{\mathscr{E}}(Q)$

Derivation of the signal-detection model. An evaluator observes a scalar signal

$$y = \theta + \varepsilon,$$

where the latent type is $\theta \in \{\mu_F, \mu_R\}$ (frontier or derivative) and $\varepsilon \sim \mathscr{N}(0, 1/\mathscr{E})$. Epistemic capital $\mathscr{E}$ governs signal precision. Define the signal-to-noise index

$$\Delta(\mathscr{E}) \equiv |\mu_F - \mu_R|\sqrt{\mathscr{E}}, \tag{A.4}$$

which coincides with the sensitivity index in Green and Swets [1966]. Under equal priors and symmetric loss, the likelihood-ratio test is equivalent to a threshold rule

$$y \geq \tfrac{1}{2}(\mu_F + \mu_R) \quad \Longleftrightarrow \quad \text{classify as frontier.}$$

Conditional on frontier type, $y \sim \mathscr{N}(\mu_F, 1/\mathscr{E})$, so the true-positive (hit) rate is

$$s(\mathscr{E}) = \Pr\big(y \geq \tfrac{1}{2}(\mu_F + \mu_R) \mid \theta = \mu_F\big) = \Phi_{\mathscr{N}}\left(\frac{\Delta(\mathscr{E})}{2}\right), \tag{A.5}$$

where $\Phi_{\mathscr{N}}$ denotes the standard normal CDF. Conditional on derivative type, $y \sim \mathscr{N}(\mu_R, 1/\mathscr{E})$, so the false-positive rate is

$$f(\mathscr{E}) = \Pr\big(y \geq \tfrac{1}{2}(\mu_F + \mu_R) \mid \theta = \mu_R\big) = 1 - \Phi_{\mathscr{N}}\left(\frac{\Delta(\mathscr{E})}{2}\right) = 1 - s(\mathscr{E}). \tag{A.6}$$

Let $c \equiv |\mu_F - \mu_R|/2 > 0$. Then $s(\mathscr{E}) = \Phi_{\mathscr{N}}(c\sqrt{\mathscr{E}})$ is strictly increasing for $\mathscr{E} > 0$. Moreover, s is strictly concave on $(0, \infty)$:

$$s''(\mathscr{E}) = -\varphi_{\mathscr{N}}(c\sqrt{\mathscr{E}})\,\frac{c\,(c^2\mathscr{E} + 1)}{4\mathscr{E}^{3/2}} < 0,$$

where $\varphi_{\mathscr{N}}$ is the standard normal density. Diminishing returns to screening precision follow.

A shift in the base rate induced by a higher derivative share

$$\varphi \equiv \frac{R}{F+R}$$

reduces the positive predictive value,

$$\mathrm{PPV}(\varphi, \mathscr{E}) = \frac{(1-\varphi)\,s(\mathscr{E})}{(1-\varphi)\,s(\mathscr{E}) + \varphi\, f(\mathscr{E})}.$$

Since $s(\mathscr{E}) > f(\mathscr{E})$ for $\mathscr{E} > 0$, it follows that $\partial \mathrm{PPV}/\partial \varphi < 0$. This motivates a specification in which the effective depreciation or fragility of epistemic capital, $\delta_{\mathscr{E}}(\varphi)$, is increasing in the derivative share: a more contaminated corpus worsens the evaluative environment holding precision fixed.

Quality feedback arises when evaluators are (re)trained on the prevailing corpus. Let $Q \in (0,1]$ denote corpus quality. If a fraction $1-Q$ of items labelled "frontier" are in fact derivative, then the effective frontier mean in the training labels is

$$\tilde{\mu}_F(Q) \equiv Q\mu_F + (1-Q)\mu_R.$$

The implied separation between training-label means is $\tilde{\mu}_F(Q) - \mu_R = Q(\mu_F - \mu_R)$, so the effective sensitivity index becomes

$$\widetilde{\Delta}(Q, \mathscr{E}) = Q\,\Delta(\mathscr{E}). \tag{A.7}$$

Because (A.5) is monotone in the sensitivity index, detection performance decreases monotonically in Q. The reduced form $D_{\mathscr{E}}(Q) = Q^{\sigma_{\mathscr{E}}}$ imposes this monotonicity and concavity while remaining parsimonious. In the Gaussian SDT benchmark, performance measures proportional to the sensitivity index correspond to $\sigma_{\mathscr{E}} = 1$. The calibration selects $\sigma_{\mathscr{E}}$ to match the implied annualised decline in detection accuracy over the relevant range of Q.

Finally, the reduced-form misclassification function used in the main text can be chosen to match $f(\mathscr{E})$ on the calibration domain. For example, $\phi(\mathscr{E}) = (1-\mathscr{E})^m$ provides a smooth approximation to (A.6), and the calibration selects m to minimise approximation error on the relevant set. □

A.3 Roy-model derivation

Derivation of equations (20)–(10). Ability z is Pareto: $\Pr(z \geq x) = (\underline{z}/x)^{\zeta}$, $x \geq \underline{z} > 0$, $\zeta > 1$. Frontier work pays $w_F(z) = p_F z^{\gamma_F} A^{a_F}$; derivative work pays $w_R(z) = p_R z^{\gamma_R} A^{a_R}$ with $\gamma_F > \gamma_R$. The wage ratio $w_F/w_R = (p_F A^{a_F}/p_R A^{a_R}) z^{\gamma_F - \gamma_R}$ is strictly increasing in ability, so single-crossing selects an interior threshold. Indifference $w_F(\bar{z}) = w_R(\bar{z})$ pins

$$\bar{z} = \left(\frac{p_R A^{a_R}}{p_F A^{a_F}}\right)^{1/(\gamma_F - \gamma_R)}. \tag{A.8}$$

Higher-ability agents sort into frontier work; lower-ability agents into derivative production.

The frontier participation rate is $\pi_F(\bar{z}) = (\underline{z}/\bar{z})^{\zeta}$. Conditional average effective skill among frontier workers integrates directly under the Pareto tail:

$$Z_F(\bar{z}) \equiv \mathbb{E}\big[z^{\gamma_F} \mid z > \bar{z}\big] = \frac{\zeta}{\zeta - \gamma_F} \bar{z}^{\gamma_F},$$

converging when $\zeta > \gamma_F$ (Assumption 2.4(ii)). Quality-adjusted frontier talent is

$$\Omega_F(\bar{z}) = Z_F(\bar{z})\, \pi_F(\bar{z}) = \frac{\zeta\, \underline{z}^{\zeta}}{\zeta - \gamma_F} \bar{z}^{\gamma_F - \zeta}. \tag{A.9}$$

The exponent $\gamma_F - \zeta < 0$ disciplines the sign: Ω_F is strictly decreasing in $\bar{z}$, so the mass effect dominates selection. A rightward shift in the Roy threshold—more agents choosing derivative work—reduces the quality-adjusted frontier labour force.

The AI-augmentation comparative static at fixed prices:

$$\left.\frac{d \ln \bar{z}}{d \ln A}\right|_{p_F, p_R} = \frac{a_R - a_F}{\gamma_F - \gamma_R}.$$

When $a_R > a_F$, improvements in algorithmic capacity raise $\bar{z}$, draining talent from the frontier. Combining with (A.9):

$$\left.\frac{d \ln \Omega_F}{d \ln A}\right|_{p_F, p_R} = (\gamma_F - \zeta)\, \frac{a_R - a_F}{\gamma_F - \gamma_R} < 0,$$

which isolates the talent-drain channel.

In general equilibrium, p_R/p_F responds to A through the CES aggregator (G.2):

$$\frac{d\ln\bar{z}}{d\ln A} = \frac{1}{\gamma_F - \gamma_R}\Big[(a_R - a_F) + \frac{d\ln(p_R/p_F)}{d\ln A}\Big].$$

The induced price response is derived in Appendix H.2. The sign of the total GE effect remains negative under the maintained parameter restrictions ($a_R - a_F$ dominates the price feedback when $\theta > 1$ and the frontier output share is bounded away from zero; verified at the calibration of Table 3), but the magnitude differs from the PE expression. The BGP analysis (Appendix H) uses the full GE derivative throughout. □

B Equilibrium Proofs

This appendix proves the shadow-price ordering results used in the main text (Proposition 3.4). The argument expresses the planner–competitive gaps in the relevant costate variables as the unique bounded solution to an infinite-horizon Volterra integral equation of the second kind. Positivity follows from a cooperative kernel structure and a strictly positive forcing term induced by the data-quality externality.

B.1 Shadow-price ordering

The objective is to sign the planner–competitive gaps in the shadow values of frontier knowledge and public epistemic capital. The classical Volterra existence result is stated first for reference. That theorem is formulated on a finite horizon and does not address the tail control required on an infinite horizon with transversality at infinity. The proof therefore uses a weighted contraction condition for the infinite-horizon operator.

Theorem B.1 (Finite-horizon Volterra equation)**.** *Let $T > t_0$. Let $f : [t_0, T] \to \mathbb{R}^n$ be bounded and measurable, and let $K : \{(t,s) : t_0 \le s \le t \le T\} \to \mathbb{R}^{n\times n}$ be measurable and bounded. Then the Volterra integral equation of the second kind*

$$x(t) = f(t) + \int_{t_0}^{t} K(t,s)x(s)\,ds, \qquad t \in [t_0, T],$$

admits a unique bounded measurable solution on $[t_0, T]$. If, in addition, $f(t) \ge 0$ and $K(t,s) \ge 0$ entrywise, then $x(t) \ge 0$ entrywise for all $t \in [t_0, T]$.

Remark B.2 (Scope)**.** A standard proof uses Picard iteration on the triangular domain $t_0 \le s \le t \le T$. If $\sup_{t_0 \le s \le t \le T} \|K(t,s)\| \le M$, then the m-fold Volterra operator satisfies $\|T^m\|_\infty \le [M(T-t_0)]^m/m!$, so the Neumann series converges and uniqueness follows; see Gripenberg et al. [1990, Chapter 9]. The wedge system below is posed on $[t_0,\infty)$ and is pinned down by transversality conditions at infinity. On an unbounded horizon, the finite-horizon argument does not control the tail of the integral operator. The analysis therefore imposes a discounted weighted summability condition and establishes a contraction directly in a weighted sup norm.

Assumption B.3 (Weighted Volterra summability)**.** There exist weights $w_F, w_Q, w_{\mathscr{E}} > 0$ and a constant $\kappa \in (0,1)$ such that, along the planner allocation,

$$\sup_{t \ge t_0} \max_{i \in \{F,Q,\mathscr{E}\}} \frac{1}{w_i} \sum_{j \in \{F,Q,\mathscr{E}\}} \int_t^\infty \big|K_{ij}(t,s)\big| w_j \, ds \le \kappa, \tag{B.1}$$

where $K(t,s)$ is the kernel defined in (B.7).

Theorem B.4 (Weighted infinite-horizon Volterra contraction)**.** *Let $\mathscr{X}_b$ be the space of bounded measurable maps $\Delta : [t_0,\infty) \to \mathbb{R}^3$. For $w = (w_F, w_Q, w_{\mathscr{E}}) \in \mathbb{R}^3_{++}$, define*

$$\|\Delta\|_w \equiv \sup_{t \ge t_0} \max_{i \in \{F,Q,\mathscr{E}\}} \frac{|\Delta_i(t)|}{w_i}.$$

For bounded measurable $\Phi : [t_0,\infty) \to \mathbb{R}^3$, consider the Volterra equation

$$\Delta(t) = \Phi(t) + \int_t^\infty K(t,s)\Delta(s)\, ds. \tag{B.2}$$

Under Assumption B.3, (B.2) admits a unique bounded solution $\Delta^* \in \mathscr{X}_b$. Moreover, the Picard iteration $\Delta^{(n+1)} = \Phi + \mathscr{T}\Delta^{(n)}$ converges to Δ^* in $\|\cdot\|_w$ from every bounded initial guess, where

$$(\mathscr{T}\Delta)(t) \equiv \int_t^\infty K(t,s)\Delta(s)\, ds.$$

If $\Phi \ge 0$ and $K_{ij}(t,s) \ge 0$ entrywise, then $\Delta^* \ge 0$.

Proof. For any $\Delta \in \mathscr{X}_b$ and each $i \in \{F,Q,\mathscr{E}\}$,

$$|(\mathscr{T}\Delta)_i(t)| \le \sum_j \int_t^\infty |K_{ij}(t,s)|\,|\Delta_j(s)|\, ds \le \|\Delta\|_w \sum_j \int_t^\infty |K_{ij}(t,s)| w_j \, ds.$$

Dividing by w_i, taking $\max_i$, and then $\sup_{t\geq t_0}$ yields

$$\|\mathscr{T}\Delta\|_w \leq \left(\sup_{t\geq t_0}\max_i \frac{1}{w_i}\sum_j \int_t^\infty |K_{ij}(t,s)|\, w_j\, ds\right) \|\Delta\|_w \leq \kappa \|\Delta\|_w.$$

Hence $\mathscr{T}$ is a contraction on the complete metric space $(\mathscr{X}_b, \|\cdot\|_w)$. Banach's fixed-point theorem implies existence and uniqueness of a bounded fixed point $\Delta^* = \Phi + \mathscr{T}\Delta^*$ and geometric convergence of Picard iterates.[33]

If $\Phi \geq 0$ and $K \geq 0$ entrywise, then $\mathscr{T}$ is order-preserving, so Picard iterates initialised at any nonnegative bounded element remain nonnegative; the limit Δ^* is therefore nonnegative. □

Lemma B.5 (Competitive frontier shadow value)**.** *Along the competitive equilibrium path, the shadow value V_F of frontier knowledge satisfies $V_F(t) > 0$ for all $t \geq t_0$.*

Proof. In current value, V_F solves

$$\dot{V}_F = (\rho + \delta_F)V_F - u'(C)Y_F - \xi\frac{G_F}{F}V_F \equiv \mu_F(t)V_F - u'(C)Y_F, \qquad \mu_F(t) \equiv \rho + \delta_F - \xi\frac{G_F}{F}.$$

Backward variation of constants, together with the transversality condition for V_F, gives

$$V_F(t) = \int_t^\infty \exp\Bigl(-\int_t^s \mu_F(\tau)\, d\tau\Bigr)\, u'(C(s))\, Y_F(s)\, ds. \tag{B.3}$$

Since $u'(C) > 0$ and $Y_F = \partial Y/\partial F > 0$ almost everywhere, the integrand in (B.3) is strictly positive on a set of positive measure, so $V_F(t) > 0$ for all t. □

Proposition B.6 (Shadow-price ordering in the frontier and epistemic channels)**.** *Define the wedge vector*

$$\Delta(t) \equiv \bigl(\Delta_F(t), \Psi_Q(t), \lambda_{\mathscr{E}}^{\text{pub}}(t)\bigr)^\top, \qquad \Delta_F(t) \equiv \lambda_F(t) - V_F(t),$$

where V_F is the competitive shadow value of F and λ_F the planner costate. Under Assumption B.3, the unique bounded solution of the wedge system satisfies

$$\Psi_Q(t) > 0, \qquad \Delta_F(t) > 0, \qquad t \geq t_0,$$

[33] The contraction estimate in (B.1) yields the explicit bound $\|\Delta^{(n)} - \Delta^*\|_w \leq \kappa^n \|\Delta^{(0)} - \Delta^*\|_w$ for any bounded initial guess $\Delta^{(0)} \in \mathscr{X}_b$. On a finite horizon, Picard iteration is also standard, but convergence follows from factorial decay of iterates, $\|T^m\|_\infty \leq [M(T-t_0)]^m/m!$, rather than from a contraction constant; see Gripenberg et al. [1990, Chapter 9].

and

$$\lambda_{\mathscr{E}}^{\text{pub}}(t) > 0 \quad \textit{whenever } q > 0 \textit{ on a set of positive measure.}$$

This establishes parts (i)–(iii) of Proposition 3.4.

Proof. The planner's current-value Hamiltonian is

$$\mathscr{H} = u(C) + \lambda_F \dot{F} + \lambda_R \dot{R} + \lambda_A \dot{A} + \lambda_H \dot{H} + \lambda_{\mathscr{E}}^{\text{pub}} \dot{\mathscr{E}}_{\text{pub}} + \lambda_{\mathscr{E}}^{\text{priv}} \dot{\mathscr{E}}_{\text{priv}}.$$

The data-quality index $Q \equiv F/(F+R)$ is algebraic and enters the state dynamics through $D(Q)$, $D_{\mathscr{E}}(Q)$, and $\delta_{\mathscr{E}}(\varphi)$ with $\varphi = 1 - Q$. Define the composite marginal value of quality

$$\Psi_Q \equiv \frac{\partial \mathscr{H}}{\partial Q}\Big|_{F,R \text{ fixed}} = \lambda_F \frac{D'(Q)}{D(Q)} G_F + \lambda_{\mathscr{E}}^{\text{pub}} \frac{\partial \dot{\mathscr{E}}_{\text{pub}}}{\partial Q} + \lambda_{\mathscr{E}}^{\text{priv}} \frac{\partial \dot{\mathscr{E}}_{\text{priv}}}{\partial Q}.$$

The derivatives of Q satisfy

$$\frac{\partial Q}{\partial F} = \frac{R}{(F+R)^2} > 0, \qquad \frac{\partial Q}{\partial R} = -\frac{F}{(F+R)^2} < 0.$$

Competitive agents take $(Q, \mathscr{E}_{\text{pub}})$ as exogenous aggregates, so $V_Q = V_{\mathscr{E}}^{\text{pub}} = 0$. Subtracting the competitive shadow equation for V_F from the planner costate equation for λ_F yields a linear equation in the frontier wedge:

$$-\dot{\Delta}_F + \mu_F(t)\Delta_F = \frac{R}{(F+R)^2} \Psi_Q, \tag{B.4}$$

where μ_F is defined in Lemma B.5. The quantity Ψ_Q is linear in $(\lambda_F, \lambda_{\mathscr{E}}^{\text{pub}}, \lambda_{\mathscr{E}}^{\text{priv}})$ along a fixed state-control trajectory, and each primitive costate ODE is linear in the costates along that trajectory. Differentiating Ψ_Q, substituting the planner costate equations, and then substituting $\lambda_F = V_F + \Delta_F$ yields a linear system in $(\Delta_F, \Psi_Q, \lambda_{\mathscr{E}}^{\text{pub}})$ with forcing terms driven by V_F and the decoupled costate $\lambda_{\mathscr{E}}^{\text{priv}}$. The transversality conditions for the primitive costates imply $e^{-\rho t}\Delta_F(t) \to 0$, $e^{-\rho t}\Psi_Q(t) \to 0$, and $e^{-\rho t}\lambda_{\mathscr{E}}^{\text{pub}}(t) \to 0$.

Backward variation of constants applied to the linear wedge system yields the Volterra representation

$$\Delta(t) = \Phi(t) + \int_t^\infty K(t,s)\Delta(s)\, ds, \tag{B.5}$$

with forcing $\Phi(t) = (0, \Phi_Q(t), 0)^\top$,

$$\Phi_Q(t) = \int_t^\infty e^{-\rho(s-t)} V_F(s) D'(Q(s)) \Lambda_F A(s)^{a_F} (\Omega_F(s) H(s))^{\alpha_F} F(s)^\xi \, ds, \tag{B.6}$$

and kernel

$$K(t,s) = e^{-\int_t^s \mu(\tau)\, d\tau} \begin{pmatrix} 0 & K_{FQ}(s) & K_{F\mathscr{E}}(s) \\ K_{QF}(s) & 0 & 0 \\ 0 & K_{\mathscr{E}Q}(s) & 0 \end{pmatrix}, \tag{B.7}$$

where $\mu(\tau) = \mathrm{diag}(\mu_F, \mu_Q, \mu_{\mathscr{E}})$ and each diagonal entry satisfies $\mu_j(\tau) \geq \rho$ along the planner allocation. The nonzero kernel entries are

$$K_{FQ}(s) = \frac{R}{(F+R)^2} > 0, \qquad K_{QF}(s) = D'(Q) \Lambda_F A^{a_F} (\Omega_F H)^{\alpha_F} F^\xi > 0,$$

$$K_{\mathscr{E}Q}(s) = q\, s'(\mathscr{E})\, \kappa \frac{R}{F+R} \geq 0, \qquad K_{F\mathscr{E}}(s) = \Big(\frac{\partial \dot{\mathscr{E}}_{\text{pub}}}{\partial Q}\Big)\Big(\frac{\partial Q}{\partial F}\Big) q \geq 0.$$

For the sign of $K_{F\mathscr{E}}$, note that along the planner allocation $\partial Q / \partial F > 0$ and

$$\frac{\partial \dot{\mathscr{E}}_{\text{pub}}}{\partial Q} = \Lambda_{\mathscr{E}}^{\text{pub}}\, \ell_{\mathscr{E}}^{\eta_{\mathscr{E}}}\, \sigma_{\mathscr{E}}\, Q^{\sigma_{\mathscr{E}}-1} + \delta_{\mathscr{E},1}\, \mathscr{E}_{\text{pub}} > 0,$$

so $K_{F\mathscr{E}} \geq 0$ with equality only if $q = 0$. By Lemma B.5 and $D'(Q) > 0$, the forcing (B.6) satisfies $\Phi_Q(t) > 0$ for all $t \geq t_0$.

Assumption B.3 and Theorem B.4 yield a unique bounded solution to (B.5) with $\Delta \geq 0$. Since $\Phi_Q > 0$, one has $\Psi_Q(t) \geq \Phi_Q(t) > 0$. Substituting $\Psi_Q > 0$ into the linear equation (B.4) and applying backward variation of constants gives

$$\Delta_F(t) = \int_t^\infty \exp\Big(-\int_t^s \mu_F(\tau)\, d\tau\Big) \frac{R(s)}{(F(s)+R(s))^2} \Psi_Q(s)\, ds > 0.$$

If $q > 0$ on a set of positive measure, then $K_{\mathscr{E}Q}$ is strictly positive on a set of positive measure, so

$$\lambda_{\mathscr{E}}^{\text{pub}}(t) \geq \int_t^\infty K_{\mathscr{E}Q}(t,s)\, \Phi_Q(s)\, ds > 0.$$

□

B.2 Derivative-knowledge and algorithmic-capital wedges

Write

$$\Delta_R \equiv \lambda_R - V_R, \qquad \Delta_A \equiv \lambda_A - V_A.$$

Lemma B.7 (Derivative-knowledge wedge)**.** *If $q^{SP} > 0$ and $\lambda_A \geq 0$ along the planner allocation, then $\Delta_R(t) < 0$ for all $t \geq t_0$. This establishes part (v) of Proposition 3.4.*

Proof. Subtract the competitive Euler equation for V_R from the planner costate equation for λ_R. The resulting linear equation is

$$\dot{\Delta}_R = a_R(t)\Delta_R + b_R(t), \qquad a_R(t) \equiv \rho + \delta_R - \xi_R \frac{\dot{R}}{R},$$

with forcing

$$b_R(t) = -\Psi_Q(t) \frac{F(t)}{(F(t)+R(t))^2}.$$

By Proposition B.6, $\Psi_Q(t) > 0$ for all $t \geq t_0$, and $F, R > 0$ along the planner allocation, so $b_R(t) < 0$. Backward variation of constants, together with the terminal condition $e^{-\int_0^t a_R(\tau)d\tau}\Delta_R(t) \to 0$, yields

$$\Delta_R(t) = -\int_t^{\infty} \exp\Big(-\int_t^s a_R(\tau)\,d\tau\Big)\,|b_R(s)|\,ds < 0.$$

□

Proposition B.8 (Algorithmic-capital wedge)**.** *The sign of $\Delta_A(t) = \lambda_A(t) - V_A(t)$ is generically ambiguous. In particular,*

$$\Delta_A(t) = \int_t^{\infty} e^{-\int_t^s \mu_A(\tau)\,d\tau} \frac{1}{A(s)} \Big[\omega g_A^{\text{gross}}(s)\,V_A(s) + a_F\, g_F^{\text{gross}}(s)\,\Delta_F(s) + a_R g_R^{\text{gross}}(s)\,\Delta_R(s)\Big]\,ds,$$

where the first two terms in brackets are nonnegative, while the third is strictly negative under Lemma B.7. The sign of Δ_A depends on which force dominates. This completes part (iv) of Proposition 3.4.

Proof. Subtract the competitive shadow equation for V_A from the planner costate equation for λ_A and integrate backward using the transversality condition for the algorithmic-capital shadow value. The stated representation follows. The sign decomposition follows from $V_A \geq 0$, Proposition B.6 ($\Delta_F > 0$), and Lemma B.7 ($\Delta_R < 0$). □

B.3 Economic interpretation and numerical verification

Remark B.9 (Weights and discounting)**.** Assumption B.3 imposes a small-gain restriction on the discounted cross-channel feedback embedded in the kernel (B.7). The weights $(w_F, w_Q, w_{\mathscr{E}})$ normalise the three wedge components $(\Delta_F, \Psi_Q, \lambda_{\mathscr{E}}^{\text{pub}})$ to comparable magnitudes and units, so the contraction property depends on economically meaningful amplification rather than arbitrary scaling. Discounting enters through the diagonal propagation factor $e^{-\int_t^s \mu(\tau)d\tau}$, with $\mu_j \geq \rho$ along the planner allocation, which controls the infinite-horizon tail. The weights discipline heterogeneity across channels; the discounting disciplines horizon length.

Remark B.10 (Discrete verification of the weighted bound)**.** Let $t_0 = t_1 < \cdots < t_N$ be a grid with steps $\Delta t_\ell = t_{\ell+1} - t_\ell$. For each component $i \in \{F, Q, \mathscr{E}\}$, define the discrete approximation of the weighted row-sum bound at time t_k by

$$\widehat{\kappa}_i(t_k) \equiv \frac{1}{w_i} \sum_{j \in \{F,Q,\mathscr{E}\}} \sum_{\ell=k}^{N-1} \left| K_{ij}(t_k, t_\ell) \right| w_j \Delta t_\ell, \qquad K(t_k, t_\ell) \text{ given by (B.7)}.$$

The implied contraction coefficient is

$$\widehat{\kappa} \equiv \max_{k \in \{1,\ldots,N\}} \max_{i \in \{F,Q,\mathscr{E}\}} \widehat{\kappa}_i(t_k).$$

Assumption B.3 holds numerically if $\widehat{\kappa} < 1$ and the grid is sufficiently fine. In the baseline calibration, the computed value satisfies $\widehat{\kappa} \approx 0.51$. The assumption and $\lambda_A \geq 0$ are verified along the computed planner allocation at every parameterisation in Table 7.

Remark B.11 (Choice of weights)**.** A numerically stable choice is to set $(w_F, w_Q, w_{\mathscr{E}})$ to the componentwise sup norms of the corresponding wedge objects along the computed planner path (or to economically meaningful normalisations that render each component $O(1)$). The reported $\widehat{\kappa}$ should be insensitive to moderate rescalings of w when the contraction margin is nontrivial. Replication code for the discrete verification is available in the supplementary archive.

B.4 Policy instruments

The first-order conditions for the planner's labour allocation and governance instruments are stated in Section 6; this appendix provides the superadditivity argument.

Talent allocation. At $\bar{z}^{CE}$, the planner's left-hand side strictly exceeds the CE counterpart: $\lambda_F > V_F$ (Part (i)), $\Psi_Q > 0$ (Part (ii)), and $\lambda_{\mathscr{E}}^{\text{pub}} > 0$ when $q > 0$ (Part (iii)). The right-hand side is weakly smaller since $\lambda_R \leq V_R$ (Lemma B.7). The planner selects a lower threshold $\bar{z}^{SP} < \bar{z}^{CE}$, inducing a larger quality-adjusted frontier workforce $\Omega_F^{SP} > \Omega_F^{CE}$.

Governance and epistemic investment. The planner's FOCs for q and $\ell_{\mathscr{E}}$ load Ψ_Q and $\lambda_{\mathscr{E}}^{\text{pub}}$—both identically zero in competitive equilibrium. The market sets $q^{CE} = 0$ and $\ell_{\mathscr{E}}^{CE} = 0$; the planner deploys strictly positive levels whenever marginal governance and epistemic costs are finite. The wedge is entirely attributable to the unpriced externalities on Q and $\mathscr{E}_{\text{pub}}$.

Superadditivity. Define

$$B(\ell_{\mathscr{E}}, q) \equiv \lambda_F \, D\big(Q_{\text{eff}}(\ell_{\mathscr{E}}, q)\big) \, \Lambda_F A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi-1}.$$

The cross-partial is

$$\frac{\partial^2 B}{\partial \ell_{\mathscr{E}} \partial q} = \lambda_F \, \Lambda_F A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi-1} \left[D' \, \frac{\partial^2 Q_{\text{eff}}}{\partial \ell_{\mathscr{E}} \partial q} + D'' \, \frac{\partial Q_{\text{eff}}}{\partial \ell_{\mathscr{E}}} \, \frac{\partial Q_{\text{eff}}}{\partial q} \right].$$

Both terms in brackets are nonneg: $D' > 0$; $\partial Q_{\text{eff}} / \partial q > 0$ and $\partial Q_{\text{eff}} / \partial \ell_{\mathscr{E}} > 0$; the cross-partial $\partial^2 Q_{\text{eff}} / (\partial \ell_{\mathscr{E}} \partial q) > 0$ because raising $\ell_{\mathscr{E}}$ accumulates $\mathscr{E}$ and thereby increases screening effectiveness. For $\sigma > 1$, $D'' > 0$ renders the second term strictly positive. The flow payoff B is therefore supermodular in $(\ell_{\mathscr{E}}, q)$ and has increasing differences—the conditions of Topkis's theorem. Applied to B, the theorem disciplines the joint correction: the welfare gain from simultaneously deploying governance and epistemic investment exceeds the sum of their separate contributions.

C Derivative Trap Proofs

The forward-invariance arguments below rest on Nagumo's theorem, which characterises when a closed set is positively invariant under a flow—here applied to the sublevel sets defined by the quality and epistemic thresholds. The bifurcation analysis uses the fold (saddle-node) theorem, which identifies the critical parameter value at which two equilibria coalesce and vanish.

Theorem C.1 (Nagumo, nonautonomous)**.** *Let $\mathscr{C}(t) \subset \mathbb{R}^n$ be a family of closed sets depending continuously on t in the Hausdorff metric, and let* $\mathbf{f}$ *be continuous and locally Lipschitz in the state. The set $\tilde{\mathscr{C}} \equiv \{(\mathbf{x}, \tau) : \mathbf{x} \in \mathscr{C}(\tau)\}$ is forward invariant under $(\dot{\mathbf{x}}, \dot{\tau}) = (\mathbf{f}, 1)$ if the augmented velocity lies in the contingent cone at every boundary point. When $\mathscr{C}(t) = \{\mathbf{x} : V(\mathbf{x}, t) \leq 0\}$ with $V \in C^1$ and $(\nabla_x V, \partial_t V) \neq \mathbf{0}$ on $\{V = 0\}$, the condition reduces to $\dot{V} \leq 0$ on $\{V = 0\}$.*[34]

Theorem C.2 (Fold (saddle-node) bifurcation)**.** *Consider $\dot{x} = f(x, \mu)$ with $f(x_0, \mu_0) = 0$, $f_x(x_0, \mu_0) = 0$, $f_{xx}(x_0, \mu_0) \neq 0$, and $f_\mu(x_0, \mu_0) \neq 0$. A smooth curve of equilibria passes through (x_0, μ_0) with a quadratic turning point: two equilibria exist on one side of μ_0 and none on the other.*[35]

C.1 The quality threshold $Q^\dagger$

Proposition C.3 (Properties of $Q^\dagger$)**.** *Under Assumptions 2.1–2.5 and $\underline{D} < \bar{D}$, the threshold $Q^\dagger$ defined in (28) satisfies $Q^\dagger \in (0, 1)$ throughout any trajectory with $G > \delta_F/(1 - \underline{D})$, is C^1 in the state, and has comparative statics $\partial Q^\dagger/\partial A < 0$, $\partial Q^\dagger/\partial H < 0$, $\partial Q^\dagger/\partial F > 0$.*

Derivation of Definition 4.1 and Proposition 4.2. Divide (1) by $F > 0$:

$$g_F = \Lambda_F D(Q) A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi - 1} - \delta_F.$$

Denote the gross production term $G \equiv \Lambda_F A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi - 1}$; then $g_F \geq 0$ requires $D(Q) \geq \delta_F/G$. Under the baseline $\underline{D} = 0$, substituting $D(Q) = Q^\sigma$ and inverting: $Q \geq Q^\dagger \equiv (\delta_F/G)^{1/\sigma}$. The threshold $Q^\dagger$ pins down the lowest data quality consistent with non-negative frontier growth at given factor endowments.

The chain rule delivers the comparative statics:

$$\frac{\partial Q^\dagger}{\partial x} = -\frac{Q^\dagger}{\sigma G} \frac{\partial G}{\partial x}.$$

Hence $\partial Q^\dagger/\partial A < 0$ (A augments frontier production), $\partial Q^\dagger/\partial H < 0$ (H augments it through labour), and $\partial Q^\dagger/\partial F = (1 - \xi) Q^\dagger/(\sigma F) > 0$ since $\xi < 1$ (frontier production exhibits diminishing returns in own stock). Any primitive that reduces Ω_F—in particular, an increase in $\bar{z}$ through the talent-drain channel—raises $Q^\dagger$ by depressing G.

[34] Nagumo [1942]; nonautonomous extension via time-augmentation: Blanchard et al. [2012, Theorem 4.7.1]. All barrier functions in this paper have $\partial_t V = 0$, so the autonomous specialisation applies.

[35] Kuznetsov [2004, Theorem 3.4.1].

Along an equilibrium path with $a_R > a_F$, the induced increase in the derivative share φ depresses Q faster than the direct channel lowers $Q^\dagger$ whenever the condition (30) is met. Improvements in algorithmic capacity therefore move the system toward the threshold even as they relax it pointwise—a paradox resolved by the general-equilibrium talent reallocation. □

C.2 Forward invariance

Proof of Theorem 4.5. Define the barrier function

$$V(\mathbf{x}) \equiv Q^\dagger(\mathbf{x}) - Q(\mathbf{x}), \qquad \mathbf{x} = (F, R, A, H, \mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}),$$

where $Q(\mathbf{x}) = F/(F+R)$ and $Q^\dagger(\mathbf{x})$ is the threshold in Proposition C.3. The trap region $\mathscr{T} \equiv \{\mathbf{x} : V(\mathbf{x}) \geq 0\}$, i.e. $Q \leq Q^\dagger$, is a closed (time-independent) subset of the state space. By the autonomous version of Nagumo's theorem (Theorem C.1 with $\partial_t V \equiv 0$, superlevel-set form), $\mathscr{T}$ is forward invariant provided $\dot{V} \equiv \nabla_{\mathbf{x}} V \cdot \mathbf{f}(\mathbf{x}) \geq 0$ on $\partial\mathscr{T} = \{V = 0\}$, where $\mathbf{f}$ is the vector field governing the state dynamics.

Regularity. Assumption 2.5 supplies $F(t) \geq \underline{F} > 0$, which keeps $F^{\xi-1}$ bounded above and ensures $Q^\dagger$ is well-defined and C^1 in the state variables. The composition $Q \mapsto D(Q)$ is smooth on $(0,1]$, so the vector field $\mathbf{f}$ is locally Lipschitz on compact subsets of $\mathscr{X}$ (Lemma G.4). These regularity conditions place the problem within the scope of Theorem C.1.

Barrier drift on $\partial\mathscr{T}$. On the boundary, $Q = Q^\dagger$ and

$$\dot{V} = \underbrace{\frac{\partial Q^\dagger}{\partial A}\dot{A} + \frac{\partial Q^\dagger}{\partial \Omega_F}\dot{\Omega}_F + \frac{\partial Q^\dagger}{\partial F}\dot{F} + \frac{\partial Q^\dagger}{\partial H}\dot{H}}_{\equiv \dot{Q}^\dagger} - \dot{Q}. \tag{C.1}$$

The signs are disciplined by the comparative statics of $Q^\dagger$: the direct-augmentation channel ($\partial Q^\dagger/\partial A < 0$, $\dot{A} > 0$) lowers $Q^\dagger$; the human-capital channel ($\partial Q^\dagger/\partial H < 0$, $\dot{H} > 0$ under interior ℓ_H) lowers it; the talent-reallocation channel ($\partial Q^\dagger/\partial \Omega_F < 0$, $\dot{\Omega}_F < 0$ when the Roy margin shifts against the frontier) raises it. By definition of $Q^\dagger$, $g_F \leq 0$ on $\partial\mathscr{T}$, while derivative activity persists with $g_R > 0$ (Lemma G.5), hence $g_R - g_F > 0$. (Here and throughout this proof, $g_J \equiv \dot{J}/J$ denotes the *net* growth rate including all channels; under the competitive equilibrium $q^{CE} = 0$, this coincides with g_J^{prod} defined in Appendix H.1.)

The quality law of motion (17) then delivers

$$-\dot{Q} = Q(1-Q)\,(g_R - g_F) = Q\,\varphi\,(g_R - g_F) > 0 \quad \text{on } \partial\mathscr{T},$$

which provides the dominant positive contribution to $\dot{V}$.

Sufficient conditions with explicit bounds. Three conditions jointly ensure $\dot{V} \geq 0$ on $\partial\mathscr{T}$. Each is derived by bounding the relevant term in (C.1) using infima and suprema on $\partial\mathscr{T}$.

Condition C.4 (C1: compositional drift dominates human-capital deepening)**.** *For all* $\mathbf{x} \in \partial\mathscr{T}$,

$$\sigma\,\varphi(\mathbf{x})\,(g_R(\mathbf{x}) - g_F(\mathbf{x})) \;>\; \mu(\mathbf{x})\,\alpha_F\,g_H(\mathbf{x}), \tag{C.2}$$

where $\mu \equiv \delta_F/(\delta_F - \underline{D}G) \geq 1$.[36] *A sufficient* scalar *specialisation (easier to verify) is* $\sigma\,\underline{\varphi}\,\inf_{\partial\mathscr{T}\cap\mathscr{K}_T}(g_R - g_F) > \alpha_F g_H^{\max}$ *with* $\underline{\varphi} \equiv \inf\varphi > 0$*,* $g_H^{\max} \equiv \lambda_H \ell_H^{\beta_H} - \delta_H$*, and* $\mathscr{K}_T$ *any compact forward-invariant subset containing the equilibrium trajectory.*[37]

Derivation. Log-differentiating $Q^\dagger = \Psi(G)^{1/\sigma}$ gives $\dot{Q}^\dagger/Q^\dagger = -(\mu/\sigma)\,\dot{G}/G$ (see (C.4) below). The H-channel contributes $-(\mu/\sigma)Q^\dagger\alpha_F g_H$ to $\dot{Q}^\dagger$, which is bounded below by $-(\mu/\sigma)\alpha_F g_H^{\max}$ after dividing through by $Q^\dagger > 0$. The $\dot{Q}$ term contributes $-\dot{Q} = Q^\dagger\varphi(g_R - g_F)$. Factoring $Q^\dagger$ from both sides, the bracket $\varphi(g_R - g_F) - (\mu/\sigma)\alpha_F g_H \geq 0$ holds under (C.2).

Condition C.5 (C2: talent-drain offsets direct augmentation)**.** $\alpha_F \cdot (\zeta - \gamma_F)/(\gamma_F - \gamma_R) \cdot (a_R - a_F) > a_F$.

Derivation. The A-channel contributes $(\partial Q^\dagger/\partial A)\dot{A} = -(a_F/\sigma)(Q^\dagger/A)g_A A = -(a_F/\sigma)Q^\dagger g_A$ (lowering $Q^\dagger$, which is harmful for the bound). The Ω_F-channel depends on $g_{\bar{z}}$, which from the Roy cutoff (20) and CES price ratio (G.2) satisfies

$$g_{\bar{z}} = \frac{(a_R - a_F)\,g_A + g_r}{\gamma_F - \gamma_R}, \quad g_r = \tfrac{\theta-1}{\theta}(\phi_H g_H - \phi_A g_A) + \tfrac{1}{\theta}(g_F - g_R). \tag{C.3}$$

On $\partial\mathscr{T}$ under the CE ($g_F = 0$, $g_R > 0$), the term $-g_R/\theta$ in g_r reflects the general-equilibrium price adjustment: rapid derivative growth depresses p_R/p_F, partially offsetting the productivity-

---

[36] The interior-threshold condition $Q^\dagger \in (0,1)$ requires $\underline{D}G < \delta_F$, so μ is finite on $\partial\mathscr{T}$. At the baseline calibration $\underline{D} = 0$ and $\mu = 1$.

[37] An earlier version stated σ^2 in place of σ. The σ^2 bound is strictly stronger (hence conservative): the $Q^\dagger$ factor cancels from both the H-channel of $\dot{Q}^\dagger$ and the drift term $-\dot{Q}$, so only σ (not σ^2) appears in the sharp bound.

asymmetry driver $(a_R - a_F)g_A$. Define $c \equiv \alpha_F(\gamma_F - \zeta)/(\gamma_F - \gamma_R) < 0$. The net A+talent+price contribution to $\dot{G}/G$ is

$$a_F g_A + \alpha_F g_{\Omega_F} = \underbrace{[a_F + c(a_R - a_F)]}_{<0 \text{ under C2}} g_A + c\, g_r.$$

C2 ensures the first bracket is negative (the talent-drain from the productivity asymmetry exceeds direct augmentation). The residual $c \cdot g_r$ captures the GE price adjustment.[38]

Condition C.6 (C3: epistemic erosion dominates replenishment on $\partial\mathscr{T}$)**.** $\delta_{\mathscr{E}}(\underline{\varphi}) > \eta_{\mathscr{E}}\, \ell_{\mathscr{E}}^{\eta_{\mathscr{E}}-1}\, (Q^\dagger)^{\sigma_{\mathscr{E}}}$. *Under laissez-faire, $\ell_{\mathscr{E}}^{CE} = 0$ and this holds trivially.*

Log-differentiating $Q^\dagger = \Psi(G)^{1/\sigma}$ with $\Psi(G) = (\delta_F - \underline{D}G)/((1-\underline{D})G)$ gives

$$\frac{\dot{Q}^\dagger}{Q^\dagger} = -\frac{\mu}{\sigma} \cdot \frac{\dot{G}}{G}, \qquad \mu \equiv \frac{\delta_F}{\delta_F - \underline{D}G} \geq 1, \tag{C.4}$$

where $\dot{G}/G = a_F g_A + \alpha_F g_H + \alpha_F g_{\Omega_F} + (\xi - 1)g_F$. On $\partial\mathscr{T}$ under the CE, $g_F = 0$.

Under C1–C3, each negative contribution to $\dot{V}$ is bounded above in absolute value by the corresponding positive contribution. Factoring $Q^\dagger > 0$ from all channels and combining with $-\dot{Q} = Q^\dagger \varphi(g_R - g_F)$:

$$\dot{V} \geq Q^\dagger \left[\varphi(g_R - g_F) - (\mu/\sigma)\alpha_F g_H - (\mu/\sigma)(a_F g_A + \alpha_F g_{\Omega_F})\right] \geq 0,$$

where the final inequality uses C1 (compositional drift dominates human-capital deepening) and C2 (the A+talent channel is beneficial, so the last group is nonpositive). Nagumo's condition is satisfied and $\mathscr{T}$ is forward invariant. □

Remark C.7. When C2 fails pointwise—i.e. the GE price correction (equation (C.3)) makes the combined A+talent+price channel positive at some parameter values (cf. Table 9)—forward invariance is verified using the full boundary drift $\dot{V}$ rather than the pointwise inequalities. The numerical exercises implement this check directly.

Table 9 reports the integral margins across calibration variants.

[38] C2 as stated bounds only the A-channel at fixed prices. The full condition including the GE price correction is $a_F + c[(a_R - a_F) - (\theta - 1)\phi_A/\theta] < 0$, which is slightly stronger. At calibration this holds with a comfortable margin. When C2 fails pointwise, forward invariance is verified numerically using the full boundary drift (Remark C.7).

Table 9: Talent-drain dominance: parameter verification

Parameter	Value	α_F	$\frac{\zeta-\gamma_F}{\gamma_F-\gamma_R}$	a_F	Margin M
Baseline	—	0.65	1/3	0.30	$+0.087$
$a_F = 0.20$		0.65	1/3	0.20	$+0.017$
$a_F = 0.40$		0.65	1/3	0.40	-0.183^*
$\zeta = 2.5$		0.65	1/5	0.30	-0.170^*
$\zeta = 4.0$		0.65	1/2	0.30	$+0.025$

Note: $M > 0$: pointwise condition holds. $^*M < 0$: integral condition verified numerically.

When both pointwise and integral conditions fail (not observed at the calibration but possible at, e.g., $a_F = 0.5$ or $\zeta < 2.3$), $\mathscr{T}$ is no longer forward invariant. The economy then converges to a low-growth steady state with Q stabilised below $Q^\dagger$ rather than collapsing to zero. The governance trap still binds in such configurations.

C.3 Fold bifurcation and non-substitutability

Proof of Proposition 4.7. Consider the $(\varphi, \mathscr{E})$ subsystem with governance intensity q treated as a parameter. The $\dot{\varphi} = 0$ locus is defined by $g_R(\varphi, A) = g_F(\varphi, Q_{\text{eff}}(\varphi, q, \mathscr{E}), A)$. The derivative share φ enters g_R with a positive coefficient (more derivative activity raises derivative growth, given AI augmentation) and g_F with a negative coefficient (through Q_{eff}), so the nullcline slopes downward in $(\varphi, \mathscr{E})$ space. At low q, screening is ineffective and no interior intersection with the $\dot{\mathscr{E}} = 0$ locus exists; at high q, two intersections appear. Theorem C.2 (fold bifurcation) identifies the critical parameter value q_c at the tangency. Genericity of the fold follows from the transversality condition $\partial^2 \dot{\varphi}/(\partial \varphi \, \partial q) \neq 0$, which holds because $\partial Q_{\text{eff}}/\partial q > 0$ ensures a nondegenerate unfolding. □

Proof of Proposition 4.8. Since $a_R > a_F$, raising A augments g_R more than g_F at any (φ, q), shifting the tangency condition. By the implicit function theorem,

$$\frac{\partial q_c}{\partial A} = -\frac{\Phi_A}{\Phi_q}.$$

The terms Φ_A and Φ_q have opposite signs. Therefore, $\partial q_c/\partial A > 0$, so q_c increases in A. Governance must intensify to keep pace with algorithmic improvement. □

Proof of Theorem 4.9. Fix $q < q_c$ and consider the modified vector field $\tilde{\mathbf{f}}(\mathbf{x}; \tau_F, \tau_z, \chi)$. The claim is $\dot{V} \geq 0$ on $\partial \mathscr{T}$ under any admissible (τ_F, τ_z, χ), so forward invariance is preserved.

Step 1: compositional drift remains negative. On $\partial\mathscr{T}$, $Q = Q^\dagger$ and $g_F^{\text{prod}} \le 0$. A frontier subsidy replaces G_F by $(1+\tau_F)G_F$, giving modified frontier growth $\tilde{g}_F^{\text{prod}} = (1+\tau_F)\Lambda_F D(Q^\dagger)A^{a_F}(\Omega_F H)^{\alpha_F}F^{\xi-1} - \delta_F$. At $Q = Q^\dagger$ this equals $\tau_F(\delta_F - \underline{D}G)/(1-\underline{D})$ after substituting $D(Q^\dagger)G = \delta_F$, so $\tilde{g}_F^{\text{prod}} = \tau_F\,\delta_F\,\underline{D}G^{-1}\cdot[G - \delta_F/\underline{D}]/(1-\underline{D})$. When $\underline{D}G < \delta_F$ (the relevant regime), $\tilde{g}_F^{\text{prod}}$ can be made positive for sufficiently large τ_F. However, the subsidy simultaneously raises A through increased output: $\dot{A} = \mu_A R^\nu A^\omega - \delta_A A$ is unchanged directly, but the induced growth in A feeds back via A^{a_R} with $a_R > a_F$.

Step 2: derivative response dominates. Any increase in A augments G_R by factor $A^{a_R - a_F}$ relative to G_F. The modified derivative growth satisfies $\tilde{g}_R^{\text{prod}} - \tilde{g}_F^{\text{prod}} \ge (a_R - a_F)g_A +$ (terms bounded below by C1–C3) > 0 on $\partial\mathscr{T}$. A talent subsidy τ_z shifts $\bar{z}$ downward, raising Ω_F and hence $\tilde{G}_F$, but the same asymmetry applies: $\pi_R = 1 - \pi_F$ falls, but $A^{a_R - a_F}$ amplifies the derivative sector more. Copyright $\chi \in [0,1)$ reduces $\dot{A}$ by factor $(1-\chi)^\nu$ but does not reverse $g_R^{\text{prod}} > 0$ (Lemma G.5), since $R^{\xi_R - 1}$ remains positive and $A > 0$.

Step 3: governance term is too weak. The governance reclassification contributes $q \cdot s(\mathscr{E}) \cdot \kappa \cdot (1-Q)$ to $\dot{Q}$. At $q < q_c$, this is bounded above by $q_c \cdot s(\mathscr{E}) \cdot \kappa \cdot (1 - Q^\dagger) < Q^\dagger(1-Q^\dagger)g_R^{\text{prod}}$ (the definition of q_c). Hence the positive governance contribution is strictly less than the negative compositional drift, and $\dot{Q} < 0$ on $\partial\mathscr{T}$.

Step 4: barrier conclusion. Since $\dot{Q} < 0$ on $\partial\mathscr{T}$ and $\dot{Q}^\dagger$ is controlled by C1–C3 (whose verification is unchanged because non-governance instruments do not affect the H, A, or Ω_F channels of $Q^\dagger$ adversely enough to violate the conditions), $\dot{V} = \dot{Q}^\dagger - \dot{Q} \ge 0$ on $\partial\mathscr{T}$. Nagumo's theorem gives forward invariance of $\mathscr{T}$ under the modified vector field. □

C.4 Acceleration and diminishing returns

Acceleration by endogenous AI. Let $Q_A(t)$ denote quality under endogenous A, and $Q_{\text{base}}(t)$ the counterfactual with A frozen at $A(0)$. Endogenous algorithmic improvement raises derivative output through A^{a_R}, so $\dot{R}_A > \dot{R}_{\text{base}}$ whenever $A(t) > A(0)$. The comparison $\varphi_A(t) > \varphi_{\text{base}}(t)$ follows, and monotonicity of (17) in φ gives $Q_A(t) < Q_{\text{base}}(t)$: endogenous AI accelerates the approach to the derivative trap.

Diminishing returns. Under $\alpha_F + \xi = 1 - \varepsilon$ with $\varepsilon > 0$, $\partial Q_\varepsilon^\dagger/\partial\varepsilon = Q_\varepsilon^\dagger \ln F/\sigma > 0$ for $F > 1$. Diminishing returns in frontier production raise $Q^\dagger$, tightening the trap by narrowing the feasibility margin.

D Governance Trap Proofs

D.1 Absorbing property

Lemma D.1 (Control-invariance of $\mathscr{T}$ on $\overline{\mathscr{G}}$)**.** *For any admissible policy $(q, \ell_{\mathscr{E}})$ with $q \in [0,1]$, if $\mathbf{x} \in \overline{\mathscr{G}} \cap \partial\mathscr{T}$ (i.e. $\mathscr{E}_{tot} \leq \mathscr{E}^\dagger$ and $Q = Q^\dagger$), then $\dot{Q} \leq 0$. Hence trajectories in $\mathscr{G} \cap \mathscr{T}$ cannot exit $\mathscr{T}$ under any admissible control.*

Proof. From (17) at $Q = Q^\dagger$ (where $g_F^{\text{prod}} = 0$):

$$\dot{Q}\big|_{Q=Q^\dagger} = (1 - Q^\dagger)\big[-Q^\dagger\, g_R^{\text{prod}} + q \cdot s(\mathscr{E}_{\text{tot}}) \cdot \kappa\big].$$

By Definition 5.1, $s(\mathscr{E}^\dagger) = Q^\dagger g_R^{\text{prod}} / \kappa$. Since s is increasing and $\mathscr{E}_{\text{tot}} \leq \mathscr{E}^\dagger$ on $\overline{\mathscr{G}}$:

$$q \cdot s(\mathscr{E}_{\text{tot}}) \cdot \kappa \;\leq\; 1 \cdot s(\mathscr{E}^\dagger) \cdot \kappa = Q^\dagger\, g_R^{\text{prod}}.$$

Hence the bracket is ≤ 0, giving $\dot{Q} \leq 0$, with equality only when $q = 1$ and $\mathscr{E}_{\text{tot}} = \mathscr{E}^\dagger$. For $Q < Q^\dagger$: $g_F^{\text{prod}} < 0$, making the drift term more negative, so $\dot{Q} < 0$ a fortiori. □

Remark D.2 (Economic content)**.** Lemma D.1 formalises the definition of $\mathscr{E}^\dagger$: inside the governance trap, epistemic capacity is *by definition* insufficient for screening to offset compositional drift. No policy can push Q above $Q^\dagger$ because $s(\mathscr{E}_{\text{tot}})$ is too low.

Proof of Theorem 5.4. **Reduction to $\partial\mathscr{G} \cap \mathscr{T}$.** By Lemma D.1, $\mathscr{T}$ is control-invariant within $\overline{\mathscr{G}}$: for any admissible $(q, \ell_{\mathscr{E}})$, trajectories starting in $\mathscr{G} \cap \mathscr{T}$ (Region III) remain in $\mathscr{T}$. They can only potentially exit $\mathscr{G}$ through $\partial\mathscr{G} \cap \mathscr{T}$ (since they cannot exit $\mathscr{T}$), so the Nagumo condition need only be verified on $\partial\mathscr{G} \cap \mathscr{T}$.[39]

On $\partial\mathscr{G} \cap \mathscr{T}$: $\mathscr{E}_{\text{tot}} = \mathscr{E}^\dagger$ and $Q \leq Q^\dagger$. Grant the planner maximum feasible intervention: $\ell_{\mathscr{E}} = \ell_{\mathscr{E}}^{\max}$ and $q = 1$.

Step 1: the requirement $\mathscr{E}^\dagger$ is nondecreasing on $\partial\mathscr{G} \cap \mathscr{T}$. By the chain rule,

$$\dot{\mathscr{E}}^\dagger = \frac{\partial \mathscr{E}^\dagger}{\partial A}\dot{A} + \frac{\partial \mathscr{E}^\dagger}{\partial \varphi}\dot{\varphi}.$$

[39]The complementary part of $\partial\mathscr{G}$—the Region II locus where $Q > Q^\dagger$ and $\mathscr{E}_{\text{tot}} = \mathscr{E}^\dagger$—is reached *before* the derivative trap closes (temporal precedence, Proposition 5.6). Absorption on this locus is not needed for the two-trap hierarchy (Theorem 5.8); see Remark D.3 below.

Both contributions are nonnegative on $\overline{\mathscr{G}} \cap \mathscr{T}$: $\dot{A} > 0$ and $\partial \mathscr{E}^\dagger / \partial A > 0$ (Lemma 5.2(i)), reflecting that more powerful AI demands more epistemic capacity. By Lemma D.1, $\dot{Q} \leq 0$ *under any admissible policy* on $\overline{\mathscr{G}} \cap \mathscr{T}$, hence $\dot{\varphi} = -\dot{Q} \geq 0$; combined with $\partial \mathscr{E}^\dagger / \partial \varphi > 0$ (Lemma 5.2(ii)), the φ-channel is nonnegative. Hence $\dot{\mathscr{E}}^\dagger \geq 0$ on $\partial \mathscr{G} \cap \mathscr{T}$.

Step 2: upper bound on epistemic replenishment. On $\partial \mathscr{G} \cap \mathscr{T}$, $Q \leq Q^\dagger$ and $Q^{\sigma_\mathscr{E}}$ is increasing in Q (since $\sigma_\mathscr{E} > 0$), so $Q^{\sigma_\mathscr{E}} \leq (Q^\dagger)^{\sigma_\mathscr{E}}$. Even under the strongest feasible response, the net epistemic drift satisfies

$$\dot{\mathscr{E}}_{\text{pub}}\big|_{\partial \mathscr{G} \cap \mathscr{T}} \leq \lambda_\mathscr{E}^{\text{pub}} (\ell_\mathscr{E}^{\max})^{\eta_\mathscr{E}} (Q^\dagger)^{\sigma_\mathscr{E}} - \delta_{\mathscr{E},0} \mathscr{E}_{\text{pub}},$$

dropping the nonpositive term $-\delta_{\mathscr{E},1} \varphi \mathscr{E}_{\text{pub}}$ and used $\delta_\mathscr{E}(\varphi) \geq \delta_{\mathscr{E},0}$. This is the critical step where restriction to $\mathscr{T}$ is used: $Q \leq Q^\dagger < 1$ ensures $(Q^\dagger)^{\sigma_\mathscr{E}}$ is a valid upper bound.[40]

Step 3: replenishment-to-erosion ratio. On $\partial \mathscr{G} \cap \mathscr{T}$, $\mathscr{E}_{\text{pub}} + \mathscr{E}_{\text{priv}} = \mathscr{E}^\dagger$, so

$$\dot{V}_G = \dot{\mathscr{E}}^\dagger - \dot{\mathscr{E}}_{\text{tot}} \;\geq\; 0 - \left[\overline{\lambda}\, (Q^\dagger)^{\sigma_\mathscr{E}} - \underline{\delta}\, \mathscr{E}^\dagger\right] = \underline{\delta}\, \mathscr{E}^\dagger - \overline{\lambda}\, (Q^\dagger)^{\sigma_\mathscr{E}},$$

using $\dot{\mathscr{E}}^\dagger \geq 0$ from Step 1, where $\overline{\lambda} \equiv \lambda_\mathscr{E}^{\text{pub}} (\ell_\mathscr{E}^{\max})^{\eta_\mathscr{E}} + \lambda_\mathscr{E}^{\text{priv}} (\ell_\mathscr{E}^{\text{priv,max}})^{\eta_\mathscr{E}}$ and $\underline{\delta} \equiv \min\{\delta_{\mathscr{E},0}, \delta_{\mathscr{E},0}^{\text{priv}}\}$. Define the replenishment-to-erosion ratio

$$\mathscr{R}(\sigma_\mathscr{E}) \;\equiv\; \frac{\overline{\lambda}\, (Q^\dagger)^{\sigma_\mathscr{E}}}{\underline{\delta}\, \mathscr{E}^\dagger}. \tag{D.1}$$

Then $\dot{V}_G \geq 0$ on $\partial \mathscr{G} \cap \mathscr{T}$ whenever $\mathscr{R} \leq 1$.

To obtain a *uniform* (state-independent) threshold $\bar{\sigma}_\mathscr{E}$, the worst case is bounded: minimise $\mathscr{E}^\dagger$ and maximise $Q^\dagger$ over $\partial \mathscr{G} \cap \mathscr{T}$. Define

$$\mathscr{E}_{\min}^\dagger \equiv \inf_{\partial \mathscr{G} \cap \mathscr{T}} \mathscr{E}^\dagger > 0, \qquad Q_{\max}^\dagger \equiv \sup_{\partial \mathscr{G} \cap \mathscr{T}} Q^\dagger \in (0,1).$$

Positivity of $\mathscr{E}_{\min}^\dagger$ follows from $Q^\dagger g_R^{\text{prod}} / \kappa > 0$ on $\partial \mathscr{G} \cap \mathscr{T}$ (Lemma G.5) and s^{-1} increasing with $s^{-1}(0) = 0$. The bound $Q_{\max}^\dagger < 1$ holds because the interior-threshold condition $Q^\dagger \in (0,1)$ applies uniformly on $\partial \mathscr{G} \cap \mathscr{T}$; along trap trajectories, $G \to \infty$ (via the R–A feedback) drives $Q^\dagger \to 0$, so the supremum is attained near initial entry ($Q_{\max}^\dagger \approx 0.65$ at calibration).

[40]On $\partial \mathscr{G} \setminus \mathscr{T}$ (Region II), $Q > Q^\dagger$ and $Q \leq 1$ would be needed, giving the trivial bound $Q^{\sigma_\mathscr{E}} \leq 1$. The replenishment-to-erosion ratio could then exceed unity, and absorption may fail—economically correct, since high data quality aids retraining.

Since $Q^\dagger_{\max} < 1$, the map $\sigma_\mathscr{E} \mapsto (Q^\dagger_{\max})^{\sigma_\mathscr{E}}$ is strictly decreasing. The intermediate value theorem delivers a unique threshold $\bar{\sigma}_\mathscr{E} \geq 0$ with $\mathscr{R}(\bar{\sigma}_\mathscr{E}) = 1$ (at the worst-case point), given in closed form by

$$\bar{\sigma}_\mathscr{E} \;=\; \max\left\{0,\; \frac{\ln(\underline{\delta}\,\mathscr{E}^\dagger_{\min}) - \ln\overline{\lambda}}{\ln Q^\dagger_{\max}}\right\}. \tag{D.2}$$

The $\max\{0,\cdots\}$ handles the case where depreciation dominates replenishment even at $\sigma_\mathscr{E} = 0$; then the trap is absorbing for all $\sigma_\mathscr{E} > 0$.

Under the normalisations in the main text ($\ell^{\max}_\mathscr{E} = 1$, $\lambda^{\text{priv}}_\mathscr{E} = 0$, $\underline{\delta} = \delta_{\mathscr{E},0}$, and $\mathscr{E}^\dagger_{\min} \approx 1$ at the boundary normalisation), (D.2) reduces to the displayed expression (33). At the calibration, $\bar{\sigma}_\mathscr{E} \approx 0.73$.

Step 4: Nagumo conclusion. For $\sigma_\mathscr{E} > \bar{\sigma}_\mathscr{E}$, $\mathscr{R} < 1$, so $\dot{V}_G \geq 0$ on $\partial\mathscr{G} \cap \mathscr{T}$. The barrier $V_G(\mathbf{x}) = \mathscr{E}^\dagger(\mathbf{x}) - \mathscr{E}_{\text{tot}}(\mathbf{x})$ is C^1 with no explicit time dependence; the autonomous Nagumo condition (Theorem C.1) applies. Since $\mathscr{T}$ is control-invariant within $\overline{\mathscr{G}}$ (Lemma D.1), trajectories originating in $\mathscr{G} \cap \mathscr{T}$ remain in $\mathscr{T}$ and can only potentially exit $\mathscr{G}$ through $\partial\mathscr{G} \cap \mathscr{T}$. The Nagumo condition blocks this exit, so $\mathscr{G} \cap \mathscr{T}$ is forward invariant (absorbing): once the economy enters Region III, no admissible policy restores $\mathscr{E}_{\text{tot}} > \mathscr{E}^\dagger$. □

Remark D.3 (Governance trap in Region II)**.** On $\partial\mathscr{G} \setminus \mathscr{T}$ (Region II: $Q > Q^\dagger$, $\mathscr{E}_{\text{tot}} = \mathscr{E}^\dagger$), the absorbing property need not hold, and this is economically appropriate. In Region II, data quality has not yet collapsed, so aggressive evaluator retraining (with high $Q^{\sigma_\mathscr{E}}$) could in principle replenish $\mathscr{E}$ faster than it erodes. The governance trap is "soft" in Region II: a social planner with sufficient resources could escape by investing heavily in epistemic infrastructure while Q is still high. The trap becomes "hard" (absorbing) only after the system crosses into Region III, where both $Q \leq Q^\dagger$ (limiting retraining effectiveness) and control-invariance of $\mathscr{T}$ (Lemma D.1) jointly close the escape route. The two-trap hierarchy (Theorem 5.8) relies only on absorption in Region III.

E Computational Methods

The replication package provides code and parameter files for all numerical exercises reported in the paper.

E.1 Planner transition

The planner chooses controls $\mathbf{u} = (q, \ell_F, \ell_R, \ell_{\mathscr{E}}, \ell_H)$ subject to $\sum \ell_j = 1$ and $C = Y - \Gamma(q) > 0$. Eliminating ℓ_R leaves four free controls, determined pointwise from the Hamiltonian FOCs. The planner's choice of ℓ_F pins down the Roy cutoff $\bar{z}^{SP} = \underline{z}\,\ell_F^{-1/\zeta}$ and hence the quality-weighted talent input $\Omega_F(\bar{z}^{SP})$ that enters G_F below (see Appendix A.3).

E.1.1 State–costate system

The necessary conditions comprise six state laws (equations (6)–(10) in the main text) and six current-value costate equations. Define the gross production terms

$$\begin{aligned} G_F &\equiv \Lambda_F\, D(Q) A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi}, \\ G_R &\equiv \Lambda_R A^{a_R} (\ell_R H)^{\alpha_R} R^{\xi_R}, \end{aligned}$$

so that $\dot{F} = G_F - \delta_F F + q\, s\, \kappa R$ and $\dot{R} = G_R - \delta_R R - q\, s\, \kappa R$. The costates satisfy:

$$\dot{\lambda}_F = (\rho + \delta_F)\lambda_F - u'(C)\, p_F - \lambda_F\, \xi\, G_F/F - \Psi_Q R/(F+R)^2, \tag{E.1}$$

$$\begin{aligned} \dot{\lambda}_R = (\rho + \delta_R)\lambda_R - u'(C)\, p_R - \lambda_R\, \xi_R\, G_R/R + \Psi_Q F/(F+R)^2 \\ - \lambda_A\, \mu_A \nu R^{\nu-1} A^{\omega} + (\lambda_R - \lambda_F)\, q\, s\, \kappa, \end{aligned} \tag{E.2}$$

$$\begin{aligned} \dot{\lambda}_A = (\rho + \delta_A)\lambda_A - u'(C)\, \partial Y/\partial A - \lambda_F\, a_F\, G_F/A - \lambda_R\, a_R\, G_R/A \\ - \lambda_A\, \omega \mu_A R^{\nu} A^{\omega-1}, \end{aligned} \tag{E.3}$$

$$\begin{aligned} \dot{\lambda}_H = (\rho + \delta_H)\lambda_H - u'(C)\, \partial Y/\partial H - \lambda_F\, \alpha_F\, G_F/H - \lambda_R\, \alpha_R\, G_R/H \\ - \lambda_H\, \bar{\lambda}_H\, \ell_H^{\beta_H}, \end{aligned} \tag{E.4}$$

$$\dot{\lambda}_{\mathscr{E}}^{\text{pub}} = (\rho + \delta_{\mathscr{E}}(\varphi))\, \lambda_{\mathscr{E}}^{\text{pub}} - (\lambda_F - \lambda_R)\, q\, s'(\mathscr{E})\, \kappa R, \tag{E.5}$$

$$\dot{\lambda}_{\mathscr{E}}^{\text{priv}} = (\rho + \delta_{\mathscr{E}}^{\text{priv}}(\varphi))\, \lambda_{\mathscr{E}}^{\text{priv}} - u'(C)\, \kappa_{\text{cert}}\, p_R R, \tag{E.6}$$

where $\bar{\lambda}_H$ is the human-capital accumulation productivity parameter (not the costate; context disambiguates), $\partial Y/\partial A$ and $\partial Y/\partial H$ are the CES marginal products from (3), and the last equation uses the certification-premium channel: in symmetric equilibrium, aggregate private epistemic capital raises the average certified price to $\tilde{p}_R = p_R(1 + \kappa_{\text{cert}}\mathscr{E}_{\text{priv}})$ (aggregating the firm-level premia $\tilde{p}_{R,i} = p_R(1 + \kappa_{\text{cert}} e_i)$ from Proposition 3.2), so $\partial \tilde{p}_R R/\partial \mathscr{E}_{\text{priv}} = \kappa_{\text{cert}}\, p_R R$. When the model omits the certification channel, the last term reduces to zero and

$\lambda_{\mathscr{E}}^{\text{priv}}$ decays at rate $\rho + \delta_{\mathscr{E}}^{\text{priv}}$.

Derivation note. Since $Q = F/(F+R)$ is algebraic, all Q-dependence in the Hamiltonian generates chain-rule terms in $\dot{\lambda}_F$ and $\dot{\lambda}_R$ via $\partial Q/\partial F = R/(F+R)^2$ and $\partial Q/\partial R = -F/(F+R)^2$. These are collected in Ψ_Q, which absorbs *all* Q-channels: the productivity channel $\lambda_F[D'(Q)/D(Q)]G_F$, the epistemic replenishment channel through $D_{\mathscr{E}}(Q) = Q^{\sigma_{\mathscr{E}}}$, and the depreciation channel through $\delta_{\mathscr{E}}(\varphi)$ with $\varphi = 1 - Q$. No separate $\delta_{\mathscr{E},1}$ terms appear in $\dot{\lambda}_F$ or $\dot{\lambda}_R$; that channel is already inside Ψ_Q.

Throughout, G_J/x denotes the gross production elasticity: $\xi\, G_F/F = \partial G_F/\partial F$ (since $G_F = \Lambda_F D A^{a_F}(\Omega_F H)^{\alpha_F} F^{\xi}$ gives $\partial G_F/\partial F = \xi G_F/F$). The distinction between gross G_F/F and net $\dot{F}/F$ matters: the depreciation $-\delta_F$ in $\dot{F}$ is absorbed into the $(\rho + \delta_F)$ coefficient and must not appear again in the production term. The computational code uses automatic differentiation of the full right-hand side to construct the Jacobian; the displayed system is included for analytical transparency.

The full system is 12-dimensional.

E.1.2 Control computation

Controls are computed at each evaluation of the vector field. Governance satisfies the interior condition

$$u'(C)\,\Gamma'(q) = (\lambda_F - \lambda_R)\, s(\mathscr{E})\, \kappa R.$$

Strict convexity $\Gamma'' > 0$ pins down at most one interior solution, computed by bisection on $[0,1]$ at tolerance 10^{-12}; the corner $q = 1$ binds when the right-hand side exceeds $u'(C)\,\Gamma'(1)$.

Labour allocations satisfy equalisation of marginal Hamiltonian value. Since labour enters the model only through the accumulation equations $\dot{F}, \dot{R}, \dot{\mathscr{E}}, \dot{H}$ and not through current output Y (which depends on stocks F, R, A, H), the Hamiltonian FOC for ℓ_j involves only the costate channel $\lambda_j\, \partial \dot{x}_j/\partial \ell_j$, with no direct utility term $u'(C)\, \partial Y/\partial \ell_j$:

$$\lambda_F \frac{\partial \dot{F}}{\partial \ell_F} = \lambda_R \frac{\partial \dot{R}}{\partial \ell_R} = \lambda_{\mathscr{E}}^{\text{pub}} \frac{\partial \dot{\mathscr{E}}_{\text{pub}}}{\partial \ell_{\mathscr{E}}} = \lambda_H \frac{\partial \dot{H}}{\partial \ell_H}.$$

The implied 3×3 system in $(\ell_F, \ell_{\mathscr{E}}, \ell_H)$ is solved by Newton's method with analytic Jacobian. Corner solutions arise when interior Newton iterates produce a negative allocation; these are detected by checking $\ell_j < \varepsilon$ with $\varepsilon = 10^{-8}$. When a corner binds, the corresponding control is set to zero and removed from the active set, reducing the system dimension. Feasibility of the remaining allocations is maintained by projecting onto the simplex $\{\ell_j \geq 0 : \sum \ell_j = 1 - \ell_H\}$ after each Newton step. The binding set is re-evaluated at each time point; transitions between interior and corner regimes are smooth at the calibration (no chattering is observed).

E.1.3 Boundary conditions and shooting

The problem is a two-point BVP: initial states $\mathbf{x}(0)$ are given; transversality conditions

$$\lim_{t\to\infty} e^{-\rho t}\lambda_j(t)x_j(t) = 0 \tag{E.7}$$

pin down λ_0. The infinite horizon is approximated by truncation at $\bar{T}$ with terminal costates anchored to the managed BGP (Appendix H). On the BGP, each costate satisfies $\dot{\lambda}_j = (\rho - g_j^*)\lambda_j - u'(C^*)\partial Y^*/\partial x_j$ (plus cross terms that vanish at the balanced allocation), so the stationary costate is

$$\lambda_j(\bar{T}) = \frac{u'(C^*)\,(\partial Y^*/\partial x_j)}{\rho - g_j^* + \eta\, g_C^*}, \tag{E.8}$$

where g_j^* is the BGP growth rate of stock j and η is the CRRA parameter (so $\dot{u}'/u' = -\eta g_C$). The denominator is positive under the standard transversality restriction $\rho > \eta g_C^* + (1-\eta)g_j^*$, verified at the calibration.

The shooting residual $\mathbf{r}(\lambda_0) \equiv \lambda(\bar{T};\lambda_0) - \lambda^{\text{term}}$ is driven to zero by Levenberg–Marquardt. The Jacobian $\partial\mathbf{r}/\partial\lambda_0$ is computed by forward sensitivity analysis: the 12×6 matrix $\mathbf{S}(t) = \partial(\mathbf{x},\lambda)/\partial\lambda_0$ satisfies $\dot{\mathbf{S}} = J(t)\mathbf{S}$ with J the 12×12 Jacobian of the full system (computed analytically). The sensitivity ODE is integrated jointly with the 12 state-costate equations (84 coupled equations in total).

Stopping criteria: $\|\mathbf{r}\|_\infty < 10^{-6}$ (primary); relative change in λ_0 below 10^{-8} (secondary). Three initialisations are used: (i) a scale-normalised guess from (E.8) evaluated at $t = 0$; (ii) a backward sweep from $\bar{T}$ along the CE state path; (iii) random perturbations around (ii). The reported solution is invariant across convergent starts.

Table 10: Truncation-horizon sensitivity

$\bar{T}$	$T_{\mathscr{E}}$	T_Q	g_F^{LR}	CEV	$\|\mathbf{r}\|_\infty$	$\lvert\mathscr{H}(\bar{T})-\mathscr{H}(0)\rvert$
75	15.08	23.97	+0.81%	6.83%	3.1×10^{-5}	1.2×10^{-4}
100	15.10	24.01	+0.80%	6.81%	4.7×10^{-6}	2.8×10^{-5}
150	15.10	24.02	+0.80%	6.80%	8.3×10^{-7}	4.9×10^{-7}
200	15.10	24.02	+0.80%	6.80%	2.1×10^{-7}	6.1×10^{-8}
300	15.10	24.02	+0.80%	6.80%	5.4×10^{-8}	8.7×10^{-9}

E.2 Truncation-horizon robustness

F Algorithmic Capacity Robustness

The main text uses a stock formulation for algorithmic capacity ((6)). An alternative flow formulation delivers the same qualitative predictions.

Stock formulation (baseline):

$$\dot{A} = \mu_A R^\nu A^\omega - \delta_A A.$$

Flow formulation (alternative):

$$\dot{A} = \mu_A (\dot{R}^+)^\nu A^\omega - \delta_A A, \qquad \dot{R}^+ \equiv \max\{\dot{R}, 0\}.$$

Table 11: Stock vs. flow formulation

	Stock	Flow	Difference
$T_{\mathscr{E}}$ (years)	14.3	15.1	+0.8
T_Q (years)	22.7	24.0	+1.3
ΔT (years)	8.4	8.9	+0.5
g_F^{LR} (%/yr)	−3.8	−3.4	+0.4
CEV (%)	7.4	6.8	−0.6

Under the flow formulation, $(\dot{R}^+)^\nu$ collapses immediately when $\dot{R} \leq 0$, attenuating the AI improvement channel and delaying trap crossings modestly. Forward invariance, the absorbing property, temporal precedence, and the instrument ordering are unchanged.

Remark F.1 (Data deletion)**.** Under the stock formulation, accumulated R contributes to AI improvement even after governance curtails new derivative output. Permitting data deletion from training corpora introduces an effective stock $R_{\text{eff}} = R - \int_0^t d(s)\,ds$, which weakens the persistence of the AI-capacity channel. Under the flow formulation, governance that reduces $\dot{R}^+$ immediately attenuates $\dot{A}$, so the incremental value of deletion is smaller.

G Existence and Uniqueness of Competitive Equilibrium

The competitive equilibrium characterised in Section 3.1 exists, is unique, and generates a well-defined trajectory on $[0,\infty)$. The global-extension step rests on Picard–Lindelöf, which guarantees local existence and uniqueness for ODEs with locally Lipschitz right-hand sides and characterises when solutions extend to all time—here applied to the reduced five-dimensional dynamics after the static equilibrium map eliminates the control variables.

Theorem G.1 (Picard–Lindelöf and maximal extension)**.** *Let $U \subset \mathbb{R}^n$ be open and $\mathbf{f} : U \to \mathbb{R}^n$ locally Lipschitz.*

(a) *For every $\mathbf{x}_0 \in U$ there exists $T_0 > 0$ and a unique C^1 solution on $[0, T_0]$.*

(b) *The solution extends uniquely to a maximal interval $[0, T^*)$, $T^* \in (0, \infty]$.*

(c) *If $T^* < \infty$, the trajectory eventually leaves every compact subset of U; equivalently, a trajectory remaining in a compact subset on every finite interval forces $T^* = \infty$.*[41]

Maintain Assumptions 2.1–2.5 and 2.4 throughout. Fix the education share $\ell_H \in (0,1)$; Remark G.9 extends the argument to endogenous ℓ_H.

G.1 The static equilibrium map

Define the admissible state space

$$\mathscr{X} \equiv \{(F, R, A, H, \mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}) \in \mathbb{R}^4_{++} \times \mathbb{R}^2_{+}\}.$$

Given $\mathbf{x} \in \mathscr{X}$, the static equilibrium is a triple $(\bar{z}, \theta_{\mathscr{E}}, \pi)$ satisfying profit maximisation, Roy sorting, optimal private epistemic investment, and labour-market clearing. The construction is sequential and each step yields a unique, smooth outcome.

[41] Teschl [2012, Theorems 2.2, 2.13, Corollary 2.16].

Step 1: Prices. The CES aggregator (3) with $\theta > 1$ delivers competitive prices as marginal products:

$$p_F(\mathbf{x}) = \alpha_Y A^{\phi_A}\big(Y/(A^{\phi_A}F)\big)^{1/\theta}, \qquad p_R(\mathbf{x}) = (1-\alpha_Y)H^{\phi_H}\big(Y/(H^{\phi_H}R)\big)^{1/\theta}. \tag{G.1}$$

Both are C^∞ on $\mathscr{X}$ and strictly positive, since $\theta > 1$ ensures that each composite receives a positive share of output. The ratio

$$r(\mathbf{x}) \equiv p_R/p_F = \frac{1-\alpha_Y}{\alpha_Y}\frac{H^{\phi_H}}{A^{\phi_A}}\big(A^{\phi_A}F/(H^{\phi_H}R)\big)^{1/\theta} \tag{G.2}$$

is C^∞ and strictly positive on $\mathscr{X}$.

Step 2: Roy threshold.

Lemma G.2 (Single-crossing)**.** *For each $\mathbf{x} \in \mathscr{X}$, the Roy sorting problem has a unique equilibrium: either an interior cutoff $\bar{z} > \underline{z}$ or the corner $\pi_R = 0$.*

Proof. The wage ratio $w_F(z)/w_R(z) = (p_F/p_R)A^{a_F-a_R}z^{\gamma_F-\gamma_R}$ is strictly increasing in z because $\gamma_F > \gamma_R$ (Assumption 2.4(i)). The single-crossing property is inherited from the monotone-likelihood-ratio structure of the Pareto distribution: higher-ability agents have a comparative advantage in frontier work, and the advantage is strict. Hence there is at most one cutoff. If $w_F(\underline{z}) \geq w_R(\underline{z})$, all researchers prefer the frontier and $\pi_R = 0$. Otherwise a unique $\bar{z} > \underline{z}$ solves $w_F(\bar{z}) = w_R(\bar{z})$.

Existence of the interior cutoff when $w_F(\underline{z}) < w_R(\underline{z})$ follows from the intermediate value theorem: as $z \to \infty$, the wage ratio $w_F/w_R \to \infty$, so the continuous function $z \mapsto w_F(z) - w_R(z)$ changes sign exactly once. Uniqueness follows from strict monotonicity. $\square$

The cutoff map is

$$\bar{z}(\mathbf{x}) = \max\big\{\underline{z},\ \big(r(\mathbf{x})\cdot A^{a_R-a_F}\big)^{1/(\gamma_F-\gamma_R)}\big\}, \tag{G.3}$$

which is continuous on $\mathscr{X}$ and C^∞ on the open set $\{\bar{z} > \underline{z}\}$.

Step 3: Talent allocation. Given $\bar{z}(\mathbf{x})$, the Pareto distribution pins down all labour-market aggregates:

$$\pi_F = (1-\ell_H)(\underline{z}/\bar{z})^\zeta, \quad \Omega_F = (1-\ell_H)\frac{\zeta\,\underline{z}^\zeta}{\zeta-\gamma_F}\,\bar{z}^{\gamma_F-\zeta}, \quad \pi_R = (1-\ell_H)\big(1-(\underline{z}/\bar{z})^\zeta\big).$$

Since $\gamma_F < \zeta$ (Assumption 2.4(ii)), the exponent $\gamma_F - \zeta < 0$ and all three maps are C^∞ in $\bar{z}$, hence C^∞ in $\mathbf{x}$ by composition.

Step 4: Private epistemic investment. Each derivative-sector firm i accumulates a proprietary detection stock e_i by diverting fraction θ_i of its labour force to detection. The certified price for firm i is $\tilde{p}_{R,i} = p_R(1+\kappa_{\text{cert}}e_i)$, so each firm internalises the return to its own detection effort. The FOC equates marginal certification revenue (strictly decreasing in θ_i by concavity of the detection technology) to marginal cost (strictly increasing in θ_i by convexity of the labour reallocation). The two curves cross exactly once, delivering a unique interior solution $\theta_i^* \in (0,1)$. In symmetric equilibrium $\theta_i = \theta_{\mathscr{E}}^{CE}(\mathbf{x})$ for all i, and aggregate private epistemic capital is $\mathscr{E}_{\text{priv}} = e \cdot \pi_R(\bar{z})$ where e is the common per-firm stock. The implicit function theorem—applicable because the Jacobian of the FOC with respect to $\theta_{\mathscr{E}}$ is nonzero at the root—gives C^1 dependence on $\mathbf{x}$.

Combining. Define $\Phi : \mathscr{X} \to (\underline{z}, \infty) \times (0,1)$, $\Phi(\mathbf{x}) = (\bar{z}(\mathbf{x}), \theta_{\mathscr{E}}^{CE}(\mathbf{x}))$.

Proposition G.3 (Static equilibrium)**.** *For each* $\mathbf{x} \in \mathscr{X}$*,* $\Phi(\mathbf{x})$ *exists, is unique, and is* C^1 *in* $\mathbf{x}$.

Proof. Existence and uniqueness of $\bar{z}$: Lemma G.2. Existence and uniqueness of $\theta_{\mathscr{E}}^{CE}$: the strictly decreasing marginal revenue and strictly increasing marginal cost cross exactly once on $(0,1)$—a standard fixed-point argument on a compact interval. Smoothness: $\bar{z}$ is C^∞ in $\mathbf{x}$ on the interior (composition of C^∞ functions); $\theta_{\mathscr{E}}^{CE}$ is C^1 in $\mathbf{x}$ by the implicit function theorem applied to the FOC, whose partial derivative with respect to $\theta_{\mathscr{E}}$ is strictly negative (second-order sufficiency). The composition $\Phi = (\bar{z}, \theta_{\mathscr{E}}^{CE})$ is therefore C^1 on $\mathscr{X}$. □

G.2 Reduced-form dynamics

Substituting $\Phi(\mathbf{x})$ into the laws of motion yields the reduced-form ODE

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{G.4}$$

where $\mathbf{f} : \mathscr{X} \to \mathbb{R}^6$.

Lemma G.4 (Lipschitz regularity)**.** *For every compact* $\mathscr{K} \subset \mathscr{X}$ *with* $\inf_{\mathscr{K}} \min\{F, R\} > 0$*, there exists* $L(\mathscr{K}) < \infty$ *such that* $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})\| \leq L(\mathscr{K})\|\mathbf{x} - \mathbf{y}\|$ *for all* $\mathbf{x}, \mathbf{y} \in \mathscr{K}$.

Proof. By Proposition G.3, $\bar{z}(\cdot)$ and $\theta_{\mathscr{E}}^{CE}(\cdot)$ are C^1 on $\mathscr{X}$, hence Lipschitz on any compact subset (by the mean value theorem on a convex compact set, or more generally by compactness and continuity of the derivative). The remaining ingredients are: power functions F^{ξ}, R^{ξ_R}, A^{a_j}, A^{ω}, R^{ν}—all C^∞ on $\mathbb{R}_{++}$; the ratio $Q = F/(F+R)$, which is C^∞ away from $F = R = 0$; the erosion function $D(Q)$, which is C^1 on $(0,1]$ and Lipschitz on $\mathscr{K}$ since Q is bounded away from zero there; affine depreciation rates; and the talent maps, C^∞ in $\bar{z}$. Each component of $\mathbf{f}$ is a composition and product of C^1 functions on $\mathscr{X}$, hence is itself C^1. A C^1 function on a compact set is Lipschitz, which completes the argument. □

Lemma G.5 (Positive derivative growth on $\partial\mathscr{T}$)**.** *Under Assumptions 2.1–2.5 with $\mathfrak{S} \equiv a_R\nu/[(1-\xi_R)(1-\omega)] > 1$, there exists $\underline{g}_R > 0$ such that $g_R^{prod} \geq \underline{g}_R$ on $\partial\mathscr{T}$ along any competitive-equilibrium or planner trajectory.*

Proof. Write $g_R^{\text{prod}} = \Lambda_R A^{a_R}(\ell_R H)^{\alpha_R} R^{\xi_R - 1} - \delta_R$. Along the R–A feedback loop, $\dot{A}/A \approx \mu_A R^{\nu} A^{\omega-1}$, so $A \sim R^{\nu/(1-\omega)}$ on average. Substituting, $A^{a_R} R^{\xi_R-1} \sim R^{a_R\nu/(1-\omega)+\xi_R-1} = R^{\mathfrak{S}(1-\xi_R)+\xi_R-1} = R^{(\mathfrak{S}-1)(1-\xi_R)}$. Since $\mathfrak{S} > 1$ and $\xi_R < 1$, the exponent $(\mathfrak{S}-1)(1-\xi_R) > 0$, so the product $A^{a_R}R^{\xi_R-1} \to \infty$ as $R \to \infty$. Combining with $(\ell_R H)^{\alpha_R}$ bounded below (since $\ell_R > 0$ and H grows) and δ_R finite, g_R^{prod} is bounded below by a positive constant on compact invariant sets, hence on $\partial\mathscr{T}$. □

G.3 Existence, uniqueness, and global extension

Theorem G.6 (Competitive equilibrium)**.** *Under Assumptions 2.1–2.5 and 2.4, for every $\mathbf{x}_0 \in \mathscr{X}$ with strictly positive stocks:*

(i) *There exists a unique maximal C^1 solution $\mathbf{x} : [0, T^*) \to \mathscr{X}$, $T^* \in (0, \infty]$.*

(ii) *All stocks remain strictly positive and epistemic capital remains in $[0,1]$ on $[0, T^*)$.*

(iii) *If the knowledge stocks remain bounded on every finite interval, then $T^* = \infty$.*

(iv) *The competitive equilibrium is the unique path along which all agents optimise and markets clear at every instant.*

Proof.

Part (i). The state space $\mathscr{X} = \mathbb{R}^4_{++} \times \mathbb{R}^2_+$ is not open (the boundary $\mathscr{E}_j = 0$ is included), so Picard–Lindelöf (Theorem G.1) cannot be applied directly on $\mathscr{X}$. Define the open superset $U \equiv \mathbb{R}^4_{++} \times (-\varepsilon, \infty)^2$ for some $\varepsilon > 0$. The vector field $\mathbf{f}$ extends continuously to U

(the production and depreciation terms are smooth in $\mathbb{R}^4_{++}$ and polynomial in $\mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}$) and is locally Lipschitz on compact subsets of U by the argument of Lemma G.4. Theorem G.1(a)–(b) applied on U delivers a unique maximal solution $\mathbf{x} : [0, T^*) \to U$, where $T^* \in (0, \infty]$ is the supremum of the existence interval (Theorem G.1(b)). Part (ii) below establishes that the solution remains in the smaller set $\mathscr{X}$, so the distinction between U and $\mathscr{X}$ is immaterial for the equilibrium trajectory.

Part (ii): positivity and boundedness. For each knowledge stock $J \in \{F, R, A\}$, the law of motion has the form $\dot{J} = G_J(\mathbf{x}) J^{e_J} - \delta_J J$ with $G_J \geq 0$ and $e_J \leq 1$. At $J = 0$ the depreciation term vanishes and the production term is nonnegative, so $\dot{J} \geq 0$. More precisely, $\dot{J} \geq -\delta_J J$ everywhere, giving the comparison bound $J(t) \geq J(0) e^{-\delta_J t} > 0$. Strict positivity is preserved.

Human capital satisfies $\dot{H}/H = \lambda_H \ell_H^{\beta_H} - \delta_H$, a constant coefficient ODE, so $H(t) = H(0) \exp\{(\lambda_H \ell_H^{\beta_H} - \delta_H) t\} > 0$.

Public epistemic capital: under $\ell_{\mathscr{E}}^{CE} = 0$, $\dot{\mathscr{E}}_{\text{pub}} = -\delta_{\mathscr{E}}(\varphi) \mathscr{E}_{\text{pub}} \leq 0$, so $\mathscr{E}_{\text{pub}}$ is nonincreasing, stays in $[0, 1]$, and $\mathscr{E}_{\text{pub}}(t) = \mathscr{E}_{\text{pub}}(0) \exp\{-\int_0^t \delta_{\mathscr{E}}(\varphi(s))\, ds\} \geq 0$.

Private epistemic capital: by comparison with the autonomous equation $\dot{x} = \lambda_{\mathscr{E}}^{\text{priv}} - \delta_{\mathscr{E},0}^{\text{priv}} x$, $\mathscr{E}_{\text{priv}}(t) \leq \max\{\mathscr{E}_{\text{priv}}(0), \lambda_{\mathscr{E}}^{\text{priv}} / \delta_{\mathscr{E},0}^{\text{priv}}\} < 1$ by Assumption 2.4(vi). Hence $\mathscr{E}_{\text{priv}}$ stays in $[0, 1)$.

Part (iii): global extension. Suppose F, R, A remain bounded on every $[0, T] \subset [0, T^*)$. Part (ii) confines H to exponential growth and both epistemic capitals to $[0, 1]$. The trajectory therefore remains in a compact subset of $\mathscr{X}$ on each $[0, T]$. The blow-up alternative (Theorem G.1(c)) stipulates that if $T^* < \infty$, the solution must eventually leave every compact subset of $\mathscr{X}$; boundedness precludes this, so $T^* = \infty$.

Part (iv): uniqueness of the equilibrium path. Proposition G.3 shows $\Phi(\mathbf{x})$ is unique at each state, which pins down the vector field $\mathbf{f}$ uniquely. Part (i) then gives a unique trajectory for any initial condition. Since the static equilibrium determines all prices, allocations, and labour-market outcomes at each instant, the full equilibrium path is unique. □

Remark G.7 (Global existence at the calibration)**.** Part (iii) is verified numerically: the ODE solution extends to $T = 300$ years under all parameterisations in Table 7, with all stocks bounded and residuals below 10^{-7}. The calibration has $a_R + \xi_R > 1$ (increasing returns in AI-augmented derivative production), which prevents a purely analytical global-existence proof based on sublinear comparison. A sufficient analytical condition is $\omega + \nu \leq 1$ and $a_R + \xi_R \leq 1$; the calibration relaxes the latter for the derivative sector.

Remark G.8 (Gross substitutability and uniqueness)**.** The assumption $\theta > 1$ (gross substitutability between frontier and derivative composites) simplifies uniqueness but is not required for existence. Under $\theta > 1$, p_R/p_F is decreasing in Q: an increase in frontier knowledge depresses its relative price and raises $\bar{z}$, generating negative feedback that prevents multiplicity of the static map. When $\theta < 1$ the feedback reverses sign and multiple static equilibria could in principle arise, though this does not occur at the calibration. The negative feedback under gross substitutability is the economic primitive that disciplines uniqueness.

Remark G.9 (Endogenous ℓ_H)**.** When households optimise over ℓ_H, the equilibrium adds an Euler equation for μ_H. The augmented vector field remains locally Lipschitz and the a priori bounds carry over. Saddle-path uniqueness follows by the stable-manifold selection argument.

H Balanced Growth Path

H.1 Definition and endogenous stationarity of Q

Write $g_J \equiv \dot{J}/J$. Fix ℓ_H and set $g_H \equiv \lambda_H \ell_H^{\beta_H} - \delta_H$.

Definition H.1 (Balanced growth path)**.** A *balanced growth path* (BGP) is a trajectory along which the growth rates (g_F, g_R, g_A) are constant, the education-driven rate g_H is constant, and all labour allocations $(\ell_F, \ell_R, \ell_{\mathscr{E}}, \ell_H)$ are time-invariant.

The definition does not assume stationarity of Q, $\bar{z}$, or any intensive margin; these are derived as necessary consequences.

Proposition H.2 (Endogenous stationarity of data quality)**.** *On any BGP with $Q(0) \in (0,1)$ and constant growth rates and controls, Q is time-invariant.*

Proof. The law of motion of $Q \equiv F/(F+R)$, derived from the governance-augmented production equations, gives

$$\dot{Q} = Q(1-Q)(g_F^{\text{prod}} - g_R^{\text{prod}}) + q\,s(\mathscr{E})\,\kappa\,(1-Q), \tag{H.1}$$

where $g_J^{\text{prod}} \equiv G_J/J - \delta_J$ is the non-governance growth rate. Consider the laissez-faire case $q = 0$. The functional form $D(Q) = \underline{D} + (1-\underline{D})Q^{\sigma}$ enters the frontier growth rate through

$g_F^{\text{prod}} = \Lambda_F D(Q) A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi-1} - \delta_F$. Under $\alpha_F + \xi = 1$ (Assumption 2.1), dividing (1) by F and using $\Omega_F = \Omega_F(\bar{z})$:

$$g_F + \delta_F = \Lambda_F D(Q) A^{a_F} \left(\Omega_F(\bar{z}) H\right)^{\alpha_F}.$$

The right-hand side depends on the state through Q, A, $\bar{z}$, and H. On a BGP, g_F is constant by definition. Since A grows at constant rate g_A and H at g_H, the product $A^{a_F} H^{\alpha_F}$ grows at constant rate $a_F g_A + \alpha_F g_H$. Constancy of $g_F + \delta_F$ then requires the remaining factor $D(Q)\,\Omega_F(\bar{z})^{\alpha_F}$ to grow at rate $-(a_F g_A + \alpha_F g_H)$.

Now examine $\bar{z}$. From (A.8), $\bar{z} \propto (p_R A^{a_R}/(p_F A^{a_F}))^{1/(\gamma_F - \gamma_R)}$. The price ratio p_R/p_F depends on F, R, A, H through (G.2). On a BGP, F and R grow at rates g_F and g_R respectively. If $g_F \neq g_R$ and $Q(0) \in (0,1)$, then $Q(t) = F(t)/(F(t)+R(t))$ drifts monotonically over time: Q rises if $g_F > g_R$ and falls if $g_F < g_R$. But $D(Q)$ enters the production function with exponent $\sigma \neq 0$, so a drifting Q induces a time-varying component in g_F through $\dot{D}/D = \sigma(1-\underline{D})Q^{\sigma-1}\dot{Q}/D(Q)$. Unless $\dot{Q} = 0$, g_F cannot remain constant—a contradiction. (One might conjecture that a compensating drift in $\Omega_F(\bar{z})$ could offset $D(Q)$; Proposition H.3 below rules this out by showing that $\bar{z}$—and hence Ω_F—must itself be stationary on any BGP.)

It remains to verify that $\dot{Q} = 0$ is consistent with the definition. Setting $q = 0$ and $\dot{Q} = 0$ in (H.1) requires either $Q \in \{0,1\}$ (boundary) or $g_F = g_R$. Since $Q(0) \in (0,1)$ and the dynamics are continuous, the interior condition $g_F = g_R$ must hold. Denote this common rate $g \equiv g_F = g_R$.

Under policy ($q > 0$), set $\dot{Q} = 0$ in (H.1). If $g_F \neq g_R$, the governance term $q\,s(\mathscr{E})\,\kappa(1-Q)$ must exactly offset $Q(1-Q)(g_F - g_R)$ at every instant. Since q and $\mathscr{E}$ are constant on a BGP and Q would otherwise drift (by the argument above), Q must be constant for the offset to hold at a single Q value rather than tracking a moving target. Hence Q is stationary under policy as well. □

Two further consequences follow. Under laissez-faire, $\dot{\mathscr{E}}_{\text{pub}} = -\delta_{\mathscr{E}}(\varphi)\mathscr{E}_{\text{pub}}$ with $\ell_{\mathscr{E}} = 0$ forces $\mathscr{E}_{\text{pub}} \downarrow 0$ monotonically; any laissez-faire BGP is therefore asymptotic with $\mathscr{E}^*_{\text{pub}} = 0$. Under the laissez-faire BGP, Q stationary and $Q \in (0,1)$ require $g_F = g_R \equiv g$ (Proposition H.2).

H.2 Stationarity of the Roy threshold

Proposition H.3 (Endogenous stationarity of $\bar{z}$)**.** *On a laissez-faire BGP with Q constant and $g_F = g_R = g$, the sorting threshold $\bar{z}$ is time-invariant if and only if*

$$\left[a_R - a_F + \frac{(1-\theta)\phi_A}{\theta}\right] g_A = \frac{(1-\theta)\phi_H}{\theta}\, g_H. \tag{H.2}$$

When the condition fails, no BGP with constant Ω_F exists.

Proof. From (G.3), $\bar{z} \propto (r \cdot A^{a_R - a_F})^{1/(\gamma_F - \gamma_R)}$. The price ratio $r = p_R/p_F$ evolves as

$$\frac{\dot{r}}{r} = \frac{(\theta-1)}{\theta}\left[\phi_H g_H - \phi_A g_A\right] + \frac{g_F - g_R}{\theta}$$

(differentiating the log of (G.2)). When $g_F = g_R$, the last term vanishes and the log-derivative of $\bar{z}$ is

$$\frac{\dot{\bar{z}}}{\bar{z}} = \frac{1}{\gamma_F - \gamma_R}\left[\frac{(\theta-1)\phi_H}{\theta} g_H + \left(a_R - a_F - \frac{(\theta-1)\phi_A}{\theta}\right) g_A\right].$$

Setting $\dot{\bar{z}} = 0$ delivers (H.2). Since $\Omega_F \propto \bar{z}^{\gamma_F - \zeta}$ and $\pi_F \propto \bar{z}^{-\zeta}$, stationarity of $\bar{z}$ is both necessary and sufficient for stationarity of all talent-allocation margins. □

H.3 Growth-rate restrictions

Maintain Assumption 2.1 ($\alpha_F + \xi = 1$).

Lemma H.4 (AI accumulation)**.** *On any BGP with $R, A > 0$, $g_A = \nu g_R/(1-\omega)$.*

Proof. From (6), $g_A + \delta_A = \mu_A R^{\nu} A^{\omega-1}$. Take logs and differentiate: constancy of g_A requires $\nu g_R + (\omega - 1) g_A = 0$. Solving pins down g_A. □

Lemma H.5 (Frontier stationarity)**.** *On a laissez-faire BGP with $Q \in (0,1)$ stationary and Ω_F constant,*

$$g = \frac{\alpha_F(1-\omega) g_H}{\alpha_F(1-\omega) - a_F \nu}, \qquad \alpha_F(1-\omega) \neq a_F \nu. \tag{H.3}$$

Proof. Divide (1) by F under $\alpha_F + \xi = 1$. Stationarity of Q fixes $D(Q)$; stationarity of Ω_F removes that margin. The gross production term then grows at rate $a_F g_A + \alpha_F g_H$. Setting $g_F = g$ constant and substituting $g_A = \nu g/(1-\omega)$ (Lemma H.4 with $g_R = g$) gives the result after collecting terms. □

Lemma H.6 (Derivative stationarity)**.** *On a laissez-faire BGP with π_R constant,*

$$g = \frac{\alpha_R(1-\omega)\, g_H}{(1-\xi_R)(1-\omega) - a_R\nu}, \qquad (1-\xi_R)(1-\omega) \neq a_R\nu. \tag{H.4}$$

Proof. The argument parallels Lemma H.5. Divide (2) by R. Stationarity of π_R and $\bar{z}$ (hence Ω_R) removes the talent margin. The gross production term grows at rate $a_R g_A + \alpha_R g_H - (1-\xi_R)g$. Setting this to zero and substituting Lemma H.4 with $g_R = g$ yields the stated expression. □

H.4 The self-reinforcement index

Definition H.7 (AI self-reinforcement)**.** The *self-reinforcement index* is

$$\mathfrak{S} \equiv \frac{a_R\nu}{(1-\xi_R)(1-\omega)}. \tag{SR}$$

The R–A feedback is *strong* when $\mathfrak{S} \geq 1$ and *weak* when $\mathfrak{S} < 1$.

The index measures round-trip amplification: a_R captures how much AI augments derivative production, ν how much derivative output trains new AI, and the denominator collects diminishing returns from own-stock concavity $(1-\xi_R)$ and AI self-knowledge $(1-\omega)$. When $\mathfrak{S} \geq 1$, the denominator of (H.4) is nonpositive and no finite positive growth rate balances the derivative sector.

Proposition H.8 (No interior laissez-faire BGP)**.** *If $g_H > 0$ and $\mathfrak{S} \geq 1$, no laissez-faire BGP exists with $Q^* \in (0,1)$ and $g \geq 0$.*

Proof. Under $\mathfrak{S} \geq 1$, the denominator of (H.4) is nonpositive; $g_H > 0$ forces $g \leq 0$. Consider $g < 0$: both F and R shrink, but the derivative-sector gross production term grows at rate $a_R g_A + \alpha_R g_H - (1-\xi_R)g = a_R\nu g/(1-\omega) + \alpha_R g_H - (1-\xi_R)g$. Under $\mathfrak{S} \geq 1$, the coefficient on g is $a_R\nu/(1-\omega) - (1-\xi_R) \geq 0$; combined with $\alpha_R g_H > 0$, the gross production growth rate is strictly positive even as R declines. Hence g_R increases over time and no constant g_R is compatible—contradicting the BGP definition. The $g = 0$ case similarly fails because $\alpha_R g_H > 0$ drives g_R above zero. □

At the calibration: $a_R\nu = 0.48$, $(1-\xi_R)(1-\omega) = 0.30$, $\mathfrak{S} = 1.6$—firmly in the strong-feedback regime.

Proposition H.9 (Knife-edge)**.** *A laissez-faire interior BGP requires simultaneously: (i)* $\mathfrak{S} < 1$*; (ii) compatibility of* (H.3) *and* (H.4)*; (iii) the sorting-stationarity condition* (H.2)*. All three fail at the baseline calibration.*

H.5 Existence when $\mathfrak{S} < 1$

When the AI feedback is weak, both growth-rate equations yield positive finite values. Generically $g_F^* \neq g_R^*$, but the level of Q adjusts frontier productivity through $D(Q)$ and creates a fixed-point equation.

Define $\Psi(Q) \equiv g_F^{\text{gross}}(Q) - g_R^{\text{gross}}$, where $g_F^{\text{gross}}(Q) \equiv \Lambda_F D(Q) A^{a_F} (\Omega_F H)^{\alpha_F}$ and g_R^{gross} is independent of Q. Since $D' > 0$, Ψ is strictly increasing and continuous.

Proposition H.10. *Suppose* $\mathfrak{S} < 1$*,* $g_H > 0$*, and* $\underline{D} < \bar{D}$*. An interior BGP with* $Q^* \in (0,1)$ *exists if and only if* $g_F^{gross}(0) < g_R^{gross} < g_F^{gross}(1)$*. The BGP quality* Q^* *is unique.*

Proof. The condition $\Psi(0) < 0 < \Psi(1)$ is the stated sandwich. Strict monotonicity of Ψ (inherited from $D' > 0$) and continuity deliver a unique root by the intermediate value theorem. □

H.6 BGP growth rates

When an interior BGP exists:

$$g^* = \frac{\alpha_F (1-\omega) g_H}{\alpha_F (1-\omega) - a_F \nu}, \tag{H.5}$$

$$g_A^* = \frac{\alpha_F \nu g_H}{\alpha_F (1-\omega) - a_F \nu}, \tag{H.6}$$

$$g_Y^* = \phi_\theta (\phi_A g_A^* + g^*) + (1-\phi_\theta)(\phi_H g_H + g^*), \tag{H.7}$$

where ϕ_θ is the frontier composite's output share at the BGP. Epistemic capital: $\mathscr{E}_{\text{pub}}^* = 0$ in CE (monotone decay); under the planner, $\ell_{\mathscr{E}}^{SP} > 0$ sustains $\mathscr{E}_{\text{pub}}^* > 0$.

At the calibration, the managed BGP (planner with $q > 0$, $\ell_{\mathscr{E}} > 0$) yields: $g^* \approx 1.86 g_H$; with $g_H \approx 1\%$, frontier growth is roughly 1.9%. Output growth $g_Y^* \approx 2.3\%$.

H.7 Stability of the interior BGP

Proposition H.11 (Saddle-path stability)**.** *Suppose* $\mathfrak{S} < 1$ *and an interior BGP exists. The BGP is saddle-path stable in the detrended state space* $\hat{\mathbf{x}} \equiv (Q, \tilde{A} \equiv A/F^{a_F/\alpha_F}, \tilde{H} \equiv$

$H/F^{1/\alpha_F}, \mathscr{E}_{priv})$.

Proof. Detrend each stock by the appropriate power of F to obtain a stationary system $\dot{\hat{\mathbf{x}}} = \hat{\mathbf{f}}(\hat{\mathbf{x}})$ with rest point $\hat{\mathbf{x}}^*$. The Jacobian $J^* \equiv D\hat{\mathbf{f}}(\hat{\mathbf{x}}^*)$ is computed by linearising the detrended laws of motion at the BGP.

The eigenstructure of J^* is governed by three channels:

(a) *Data-quality feedback* (Q equation). The Q-row of J^* has a diagonal entry $\partial\dot{Q}/\partial Q|^* = (1-Q^*)\left[\partial g_F/\partial Q - \partial g_R/\partial Q\right]$ +level terms. Since $\partial g_F/\partial Q = \sigma D'(Q^*)/D(Q^*) > 0$ and $\partial g_R/\partial Q = 0$ (derivative growth does not depend on corpus quality), the diagonal entry is negative: higher Q raises frontier growth relative to derivative growth, depressing $\dot{Q}$ through the compositional channel. This negative feedback generates a stable eigenvalue.

(b) *AI-accumulation feedback* ($\tilde{A}$ equation). The detrended $\tilde{A}$ equation inherits concavity from $\omega < 1$: the diagonal entry of J^* is $(\omega-1)g_A^*/\tilde{A}^* < 0$, contributing a second stable eigenvalue.

(c) *Jump variables.* The remaining eigenvalue(s) associated with forward-looking controls (private epistemic investment, education allocation under endogenous ℓ_H) carry positive real parts. The number of unstable eigenvalues equals the number of jump variables, delivering the saddle-path structure.

Multiplicity of steady states. Under $\mathfrak{S} < 1$, strict monotonicity of $\Psi(Q)$ (Proposition H.10) rules out multiple interior rest points in the detrended system: the unique root Q^* is the only candidate, and the detrending is a smooth bijection on $\mathscr{X}$. Under $\mathfrak{S} \geq 1$, no interior rest point exists (Proposition H.8); the only attractors are the boundary configurations described in Proposition H.13. Hence, conditional on the feedback regime, the detrended system has a unique rest point (interior or boundary) and no additional steady states. □

Remark H.12 (Calibration verification)**.** At the baseline calibration, the eigenvalues of the 4-dimensional detrended system are $\{-0.047, -0.023, +0.031, +0.058\}$: two stable, two unstable, confirming the saddle-path structure of Proposition H.11.

H.8 Asymptotic dynamics when $\mathfrak{S} \geq 1$

Under strong self-reinforcement, no interior BGP exists and the laissez-faire economy converges to a degenerate configuration.

Proposition H.13 (Trap asymptotics)**.** *Suppose $\mathfrak{S} \geq 1$ and $\underline{D} = 0$. Along the laissez-faire path:* (i) $Q(t) \to 0$*;* (ii) $g_F(t) \to -\delta_F$*;* (iii) $g_R(t)$ *and* $g_A(t)$ *increase without bound;* (iv) $\mathscr{E}_{pub}, \mathscr{E}_{priv} \to 0$.

Proof. (i) Forward invariance (Theorem 4.5) ensures that once $Q < Q^\dagger$, the trajectory remains in the trap region and Q is non-increasing. Being bounded below by zero, Q converges. Suppose $\lim Q = \bar{Q} > 0$. Then $D(\bar{Q}) > 0$, and frontier production retains a positive floor. But under $\mathfrak{S} \geq 1$, the derivative production term grows without bound (as established in the proof of Proposition H.8), driving $\varphi \to 1$ and Q below $\bar{Q}$—a contradiction.

(ii) $D(Q) \to D(0) = 0$ implies $\dot{F} \to -\delta_F F$: frontier knowledge decays at its depreciation rate.

(iii) The derivative gross production term $A^{a_R}(\pi_R H)^{\alpha_R} R^{\xi_R - 1}$ accelerates through the R–A loop; $\mathfrak{S} \geq 1$ ensures effective round-trip returns exceed unity, so the feedback does not attenuate.

(iv) $\mathscr{E}_{\text{pub}}$ decays exponentially under $\ell^{CE}_{\mathscr{E}} = 0$. For $\mathscr{E}_{\text{priv}}$: the retraining technology $D_{\mathscr{E}}(Q) = Q^{\sigma_{\mathscr{E}}} \to 0$ as $Q \to 0$, while depreciation stays bounded below by $\delta^{\text{priv}}_{\mathscr{E},0} > 0$.

□

When $\underline{D} > 0$, the frontier retains productive capacity at $Q = 0$ and the CES price mechanism provides a restoring force: as $Q \to 0$, $p_F/p_R \to \infty$, which pulls $\bar{z} \to \underline{z}$ and $\pi_F \to 1$. Whether Q stabilises depends on the race between price-mediated talent reallocation and derivative self-reinforcement. Numerically: $\underline{D} = 0.03$ gives $Q \to 0.04$; $\underline{D} = 0.05$ gives $Q \to 0.11$; at $\underline{D} = \bar{D} \approx 0.08$ the trap dissolves.

H.9 Policy BGP

Under governance, $\dot{Q} = 0$ no longer requires $g_F = g_R$.

Lemma H.14. *Along any path with $Q(t) \equiv Q^*$ and $\mathscr{E}(t) \equiv \mathscr{E}^*$, $\dot{Q} = 0$ is equivalent to*

$$g_R - g_F = \frac{\kappa q s(\mathscr{E}^*)}{Q^*}. \tag{H.8}$$

Proof. Set $\dot{Q} = Q(1-Q)(g_F - g_R) + q s(\mathscr{E})\,\kappa(1-Q) = 0$ and divide by $Q(1-Q)$: $g_F - g_R + q s(\mathscr{E})\,\kappa/Q = 0$. Rearranging gives (H.8). □

Equation (H.8) is the missing degree of freedom: governance absorbs the growth-rate differential that laissez-faire cannot accommodate. The screening intensity q selects the Q^* at which the frontier can sustain positive growth despite derivative-sector self-reinforcement.

Proposition H.15 (Managed BGP)**.** *Suppose $\alpha_F(1-\omega) > a_F\nu$ and $g_H > 0$. There exist $q^{SP} > 0$ and $\ell_{\mathscr{E}}^{SP} > 0$ implementing a BGP with growth rate g^* given by* (H.5)*, $Q^{SP} > Q^{\dagger}$, and $\mathscr{E}^{SP} > \mathscr{E}^{\dagger}$.*

Proof. The frontier production block pins $g_F = g^*$ under any fixed ℓ_F and Q^{SP} satisfying $D(Q^{SP}) > 0$. The derivative block yields $g_R(Q^{SP})$ at any allocation. Lemma H.14 then determines q^{SP} to absorb $g_R - g^*$. The right-hand side of (H.8) is continuous in q on $[0,1]$, ranging from 0 at $q = 0$ to $\kappa s(\mathscr{E}^*)/Q^*$ at $q = 1$. For $\mathscr{E}^*$ sufficiently large (which $\ell_{\mathscr{E}}^{SP} > 0$ guarantees), the upper bound exceeds $g_R - g^*$ and the intermediate value theorem delivers $q^{SP} \in (0,1]$.

Epistemic investment $\ell_{\mathscr{E}}^{SP}$ is chosen to sustain $\mathscr{E}^{SP}$ via $\dot{\mathscr{E}}_{\text{pub}} = 0$: the replenishment term must offset depreciation. The planner's FOC for $\ell_{\mathscr{E}}$ has interior solutions when $\Psi_Q > 0$ (Proposition 3.4(ii)), since the shadow price of data quality makes epistemic investment socially productive. □

I Sufficiency Conditions

The planner's problem (24) is non-concave: the quality ratio $Q \equiv F/(F+R)$ enters both the flow payoff and the laws of motion through $D(Q)$ and $D_{\mathscr{E}}(Q)$, destroying joint concavity of the Hamiltonian in (F,R). Neither Mangasarian's condition (joint concavity of $\mathscr{H}$ in $(\mathbf{x},\mathbf{u})$; Acemoglu, 2009, Theorem 7.11) nor Arrow's condition (concavity of the maximised Hamiltonian in $\mathbf{x}$; Acemoglu, 2009, Theorem 7.14) holds, because $D''(Q) > 0$ for $\sigma > 1$ (Proposition I.1).

The non-concavity is structurally the same as in epidemiological growth models where an infection share enters nonlinearly in both constraints and objective. Goenka et al. [2014] face non-convex constraints and a non-concave Hamiltonian from SIS dynamics; Goenka et al. [2024] add disease-induced mortality and endogenous discounting. In each case, the Leitmann–Stalford decomposition [Leitmann and Stalford, 1971] provides the route to sufficiency. The present model replaces the epidemiological state (i) with the knowledge-quality state (Q), but the algebraic structure—a ratio of two stocks entering multiplicatively in production—is the same.

Sufficiency follows from the augmented-Hamiltonian argument in Goenka et al. [2014], Nguyen and Nguyen-Van [2016], Goenka et al. [2024]. The boundary term is controlled by a generalised transversality condition in the pathwise form of Cartigny and Michel [2003], verified via costate-sign arguments and a decay lemma for sign-ambiguous costates.

Non-concavity of the Hamiltonian

Proposition I.1 (Failure of concavity in (F,R))**.** *For $\sigma > 1$ and $\lambda_F > 0$, the restriction of $\mathscr{H}(\cdot,\mathbf{u},\lambda)$ to the (F,R) block is not concave on any neighbourhood with $F,R > 0$.*

Proof. Gross frontier production is $G_F \equiv \Lambda_F D(Q) A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi}$. The term $\lambda_F G_F$ contributes to $\partial^2\mathscr{H}/\partial F^2$ the component

$$\lambda_F A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi} D''(Q) \Big(\frac{R}{(F+R)^2}\Big)^2.$$

With $D(Q) = \underline{D} + (1-\underline{D})Q^{\sigma}$ and $\sigma > 1$, $D''(Q) > 0$ on $(0,1)$, so this contribution is strictly positive whenever $\lambda_F > 0$ and $F,R > 0$. Diminishing-returns curvature in F^{ξ} $(\xi < 1)$ does not generically dominate; the (F,R) Hessian block is indefinite. □ □

Existence and endogenous state bounds

Lemma I.2 (Endogenous state bounds)**.** *Under Assumptions 2.1–2.5, any admissible path satisfies:*

(i) *$\mathscr{E}_{pub}, \mathscr{E}_{priv} \in [0,1]$ and $Q \in [0,1]$;*

(ii) *$H(t) \le H_0 e^{g_H t}$ with $g_H \equiv \bar{\lambda}_H - \delta_H$;*

(iii) *$R(t) \le \bar{R}(t)$ and $A(t) \le \bar{A}(t)$, where $(\bar{R},\bar{A})$ solves the comparison system*

$$\dot{\bar{R}} = \Lambda_R \bar{A}^{a_R} (H_0 e^{g_H t})^{\alpha_R} \bar{R}^{\xi_R}, \qquad \dot{\bar{A}} = \mu_A \bar{R}^{\nu} \bar{A}^{\omega};$$

sublinearity ($\xi_R, \omega \in (0,1)$) excludes finite-time blowup, and the comparison solutions satisfy $\bar{R},\bar{A} = O(e^{\bar{g}t})$ with $\bar{g}$ pinned by the BGP system (Appendix H);

(iv) *$F(t) = O(e^{\bar{g}_F t})$ for a finite $\bar{g}_F$;*

(v) *the discounted objective $\int_0^\infty e^{-\rho t} u(C)\, dt$ is well-defined and finite.*

Proof. (i) Nonneg production and positive depreciation pin $\mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}$ to $[0,1]$; $Q = F/(F+R) \in [0,1]$ is algebraic. (ii) From (12) with $\ell_H \leq 1$: $\dot{H} \leq (\bar{\lambda}_H - \delta_H)H$; Gronwall. (iii) Dropping depreciation, governance, and using $\ell_R \leq 1, D(Q) \leq 1$ yields $\dot{R} \leq \Lambda_R A^{a_R} H^{\alpha_R} R^{\xi_R}$; similarly $\dot{A} \leq \mu_A R^{\nu} A^{\omega}$. The comparison system dominates (R,A) componentwise. Sublinearity in own stock excludes blowup (Osgood); exponential rates from Appendix H. (iv) Given (ii)–(iii): $\dot{F} \leq \Lambda_F \bar{A}^{a_F} (H_0 e^{g_H t})^{\alpha_F} F^{\xi} + \kappa \bar{R}(t) - \delta_F F$. Sublinearity ($\xi < 1$) plus linear depreciation; Gronwall. (v) CRRA with $\eta > 1$: $u(C) \leq 0$; discounted integral bounded above by zero, below by $\int_0^\infty e^{-\rho t} u(\underline{C})\, dt > -\infty$. □ □

Proposition I.3 (Existence)**.** *The planner's problem admits an optimal solution.*

Proof. The control set is compact. By Lemma I.2, the state dynamics are continuous in $(\mathbf{x}, \mathbf{u})$, satisfy a linear growth bound, and the payoff is bounded above. Hence a maximizing sequence exists.

Although the velocity correspondence $\mathbf{f}(\mathbf{x}, \mathscr{U})$ is generally non-convex, the existence argument follows the weak-compactness approach used in Goenka et al. [2014] for non-concave dynamic problems. Specifically, from any maximizing sequence, the controls and induced state derivatives admit a subsequence that converges weakly in $\sigma(L^1(e^{-\rho t}), L^\infty)$, while the corresponding state paths converge pointwise (after extraction) by the growth bound and equicontinuity.

Pointwise convergence is sufficient to pass to the limit in the state equations wherever strong convergence is available. For the weakly convergent components, Mazur's lemma provides convex combinations that converge strongly (hence pointwise a.e.). Feasibility of the limit path then follows from continuity of the dynamics. Jensen's inequality is used to remove the convex-combination coefficients and recover an admissible limit control without lowering the objective, relying on the required concavity in the control arguments. The argument is a direct adaptation of the existence proofs in Romer [1986] and Goenka et al. [2014], under the present assumptions (bounded controls, linear-growth dynamics, and an integrable upper bound for utility). Therefore, the claim follows. □

Costate signs

Three costates are strictly positive along any interior planner path; the remaining three may change sign.

Lemma I.4 (Positivity of the frontier shadow value)**.** *Along any interior planner path with $C(t) > 0$ and $Y_F(t) \equiv \partial Y / \partial F > 0$ a.e., $\lambda_F(t) > 0$ for all $t \geq 0$.*

Proof. Proposition 3.4(i) decomposes $\lambda_F = V_F + \Delta_F$, where V_F is the competitive shadow value and $\Delta_F \equiv \lambda_F - V_F$ the planner–private wedge.

$V_F > 0$. By the integral representation in Proposition 3.4(i),

$$V_F(t) = \int_t^\infty \exp\Bigl(-\int_t^s \mu_F(\tau)\,d\tau\Bigr)\,u'(C(s))\,Y_F(s)\,ds \tag{I.1}$$

with $\mu_F \equiv \rho + \delta_F - \xi\, G_F/F$. Since $u'(C) > 0$ and $Y_F > 0$ a.e., the integrand is strictly positive and $V_F(t) > 0$.

$\Delta_F > 0$. Proposition 3.4(iv) shows that the wedge vector $\Delta = (\Delta_F, \Psi_Q, \lambda_{\mathscr{E}}^{\text{pub}})^\top$ satisfies the cooperative Volterra system (Theorem B.4), with forcing $\Phi = (0, \Phi_Q, 0)^\top$ and $\mathbf{K} \geq 0$. Along an interior path $\Phi_Q > 0$ a.e. ($V_F > 0$ and $D'(Q) > 0$). The off-diagonal entry $K_{FQ} = R/(F+R)^2 > 0$ transmits quality value into the F-channel. At the first Picard iterate:

$$\Delta_F^{(1)}(t) = \int_t^\infty K_{FQ}(t,s)\,\Phi_Q(s)\,ds > 0.$$

Cooperativity preserves $\Delta^{(n)} \geq \Delta^{(1)}$ for all n; the contraction condition (Assumption B.3) delivers uniform convergence.

$\lambda_F = V_F + \Delta_F > 0$. □ □

Lemma I.5 (Costate signs from interior FOCs)**.** *Along an interior planner path with $\ell_H > 0$ and $\ell_{\mathscr{E}} > 0$:* (a) $\lambda_H > 0$*;* (b) $\lambda_{\mathscr{E}}^{pub} > 0$.

Proof. The interior FOC (Appendix E.1.2) equalises marginal costate-value products:

$$\lambda_F\,\frac{\partial \dot F}{\partial \ell_F} = \lambda_H\,\frac{\partial \dot H}{\partial \ell_H} = \lambda_{\mathscr{E}}^{\text{pub}}\,\frac{\partial \dot{\mathscr{E}}_{\text{pub}}}{\partial \ell_{\mathscr{E}}}. \tag{I.2}$$

(a) $\partial \dot F/\partial \ell_F > 0$ and $\partial \dot H/\partial \ell_H = \bar\lambda_H \beta_H \ell_H^{\beta_H - 1} H > 0$. Lemma I.4 gives $\lambda_F > 0$; dividing by the positive marginal product of ℓ_H pins $\lambda_H > 0$. (b) $\partial \dot{\mathscr{E}}_{\text{pub}}/\partial \ell_{\mathscr{E}} = \eta_{\mathscr{E}}\,\ell_{\mathscr{E}}^{\eta_{\mathscr{E}}-1} D_{\mathscr{E}}(Q) > 0$ for $Q > 0$. The same argument gives $\lambda_{\mathscr{E}}^{\text{pub}} > 0$. □ □

No sign claim is made for λ_R, λ_A, or $\lambda_{\mathscr{E}}^{\text{priv}}$. The costate λ_R can be negative when quality-erosion costs outweigh the direct marginal product of derivative content; λ_A is sign-ambiguous because AI augments both frontier and derivative production. This parallels the sign ambiguity of the costate on infectives in Goenka et al. [2024], where sufficiency still obtains because the generalised TVC is verified directly rather than through costate signs.

Generalised transversality conditions

Write $\mu_j(t) \equiv e^{-\rho t}\lambda_j(t)$ for the present-value costate of stock j. The standard transversality conditions are

$$\lim_{t\to\infty} e^{-\rho t}\lambda_j(t)x_j^*(t) = 0, \qquad j \in \{F, R, A, H, \mathscr{E}_{\text{pub}}, \mathscr{E}_{\text{priv}}\}. \tag{I.3}$$

These hold at the solution $\mathbf{x}^*$ but provide no information about deviations along an arbitrary feasible path. Sufficiency requires the pathwise form of Cartigny and Michel [2003]:

$$\limsup_{t\to\infty} e^{-\rho t}\sum_j \lambda_j(t)\left(x_j^*(t) - x_j(t)\right) \leq 0 \tag{I.4}$$

for any feasible $\mathbf{x}(\cdot)$ from $\mathbf{x}_0$. In the concave case, Acemoglu [2009, Theorem 7.14] assumes (I.4) together with Arrow sufficiency. When concavity fails, the condition must be verified directly—the approach here and in Goenka et al. [2014], Nguyen and Nguyen-Van [2016], Goenka et al. [2024].

Two primitives deliver (I.4) for all six states. For signed costates $(F, H, \mathscr{E}_{\text{pub}})$: nonnegativity of both the costate and the state gives $e^{-\rho t}\lambda_j(x_j^* - x_j) \leq e^{-\rho t}\lambda_j x_j^* \to 0$ by the standard TVC, and the nonneg product pins the liminf, so $\limsup \leq 0$. (In Goenka et al., 2024 the corresponding step uses $\lambda_1 \geq 0$ from the consumption FOC; here the Volterra decomposition and interior FOCs sign three costates.) For unsigned costates $(R, A, \mathscr{E}_{\text{priv}})$: present-value costate decay $\mu_j \to 0$ combined with bounded deviation growth $|x_j^* - x_j| = O(e^{g_j t})$ drives the product to zero. The bounded state $\mathscr{E}_{\text{priv}} \in [0,1]$ is handled as in Goenka et al. [2024]; the unbounded states R and A require the growth-rate comparison in Lemma I.2.

Lemma I.6 (Decay of present-value costates)**.** *Along the optimal path, $\mu_j(t) \to 0$ as $t \to \infty$ for every state j.*

Proof. Signed costates $(F, H, \mathscr{E}_{pub})$. Lemmas I.4 and I.5 give $\lambda_j \geq 0$. The standard TVC $\mu_j x_j^* \to 0$ together with x_j^* bounded away from zero ($F \geq \underline{F} > 0$ by Assumption 2.5; $H \geq H_0 e^{-\delta_H t} > 0$; $\mathscr{E}_{\text{pub}} > 0$ when $\ell_{\mathscr{E}} > 0$) forces $\mu_j \to 0$.

Unsigned costates $(R, A, \mathscr{E}_{priv})$. The present-value costate satisfies $\dot{\mu}_j = -e^{-\rho t}\partial\mathscr{H}/\partial x_j$. The forcing is integrable: bounded by products of discounted marginal utilities and polynomially growing states (Lemma I.2), with the transversality restriction $\rho > \eta g_C^* + (1-\eta)g_j^*$ (verified at the calibration, Appendix H) ensuring convergence [Michel, 1982]. Hence μ_j has a finite limit.

For $\mathscr{E}_{\text{priv}} \in [0,1]$: the standard TVC and $\mathscr{E}^*_{\text{priv}}$ bounded away from zero force the limit to zero. For R: the effective own-discount $\mu^*_R \equiv \rho + (1-\xi_R)\delta_R - \xi_R g^*_R + q^* s^* \kappa > \rho$ (sublinearity term $(1-\xi_R)\delta_R > 0$, governance term nonneg); the standard TVC with R^* bounded away from zero pins $\mu_R \to 0$. For A: $\mu^*_A = \rho + (1-\omega)\delta_A - \omega g^*_A > \rho$; the argument parallels R. ☐

Proposition I.7 (Generalised TVCs and Michel condition)**.** *Condition* (I.4) *holds for every feasible* $\mathbf{x}(\cdot)$ *from* $\mathbf{x}_0$. *The Michel condition* $\lim_{t\to\infty} e^{-\rho t}\mathscr{H}^{cv}(t) = 0$ *is also satisfied.*

Proof. Sign-based: $j \in \{F, H, \mathscr{E}_{pub}\}$. $\lambda_j \geq 0$ and $x_j \geq 0$ give $e^{-\rho t}\lambda_j(x^*_j - x_j) \leq e^{-\rho t}\lambda_j x^*_j \to 0$ by (I.3); $\liminf e^{-\rho t}\lambda_j x_j \geq 0$ pins $\limsup \leq 0$.

Decay-based: $j \in \{R, A, \mathscr{E}_{priv}\}$. Lemma I.6 gives $\mu_j \to 0$. Any two admissible paths satisfy $|x^*_j - x_j| \leq C_j e^{\bar{g}_j t}$ (Lemma I.2 bounds growth; Gronwall bounds the deviation; for $\mathscr{E}_{\text{priv}} \in [0,1]$ the deviation is uniformly bounded). The present-value costate decays at rate $\mu^*_j > \rho$, so $|\mu_j| \leq C' e^{-\mu^*_j t}$ eventually. The excess rate $\mu^*_j - \bar{g}_j > 0$ drives $|\mu_j|\,|x^*_j - x_j| \to 0$.

Summing over j:

$$\limsup_{t\to\infty}\ e^{-\rho t}\big\langle \lambda(t), \mathbf{x}^*(t) - \mathbf{x}(t)\big\rangle \leq 0. \tag{I.5}$$

Michel condition. The system is autonomous: along optimal paths $\dot{\mathscr{H}}^{cv} = \rho\mathscr{H}^{cv}$, so $\mathscr{H}^{pv}(t) \equiv e^{-\rho t}\mathscr{H}^{cv}(t)$ is constant. The standard TVCs and Lemma I.6 give $e^{-\rho t}\langle \lambda, \dot{\mathbf{x}}^*\rangle \to 0$; discounting gives $e^{-\rho t}u(C^*) \to 0$. Hence $\mathscr{H}^{pv} = 0$ [Michel, 1982]; see Acemoglu [2009, Theorem 7.12] for a textbook derivation. ☐

Sufficiency via the Leitmann–Stalford decomposition

The augmented Hamiltonian is

$$\bar{\mathscr{H}}(\mathbf{x}, \mathbf{u}, \lambda) \equiv \mathscr{H}(\mathbf{x}, \mathbf{u}, \lambda) + \langle \dot{\lambda}, \mathbf{x}\rangle. \tag{I.6}$$

Since $\langle \dot{\lambda}, \mathbf{x}^*\rangle$ is independent of $\mathbf{u}$, the maximum principle selects $\mathbf{u}^*$ as a maximiser of $\bar{\mathscr{H}}(\mathbf{x}^*, \cdot, \lambda)$. The Euler equations give $\nabla_{\mathbf{x}}\bar{\mathscr{H}}\big|_{(\mathbf{x}^*, \mathbf{u}^*)} = \mathbf{0}$.

Lemma I.8 (Control-wise maximality)**.** *Given* $\mathbf{x}^*$, $\bar{\mathscr{H}}(\mathbf{x}^*, \mathbf{u}^*, \lambda) \geq \bar{\mathscr{H}}(\mathbf{x}^*, \mathbf{u}, \lambda)$ *for all admissible* $\mathbf{u}$.

Proof. At $\mathbf{x} = \mathbf{x}^*$ the augmented Hamiltonian separates into control channels. *Consumption.* Concavity of u: $u(C^*) - u(C) \geq u'(C^*)(C^* - C) = u'(C^*)[\Gamma(q) - \Gamma(q^*)]$ at fixed $\mathbf{x}^*$.

Governance. The intensity q enters $\dot{F}$ and $\dot{R}$ linearly through $\pm qs\kappa R$. Convexity of Γ and the FOC (26) ($u'\Gamma'(q^*) = (\lambda_F - \lambda_R)s\kappa R^*$) combine to a nonneg difference. *Labour.* The shares $(\ell_F, \ell_{\mathscr{E}}, \ell_H)$ enter through concave accumulation functions ($\beta_H, \eta_{\mathscr{E}} \in (0,1)$); the FOC (I.2) equalises marginal costate-value products; concavity delivers a global maximum in the labour block. The argument exploits separability of the Hamiltonian in controls and concavity in each control block—the same structure used in Goenka et al. [2024] for the lockdown–consumption decomposition. □ □

Proposition I.9 (Augmented-Hamiltonian inequality)**.** *Along* $(\mathbf{x}^*, \mathbf{u}^*)$*, for any feasible* $(\mathbf{x}, \mathbf{u})$*,*

$$\bar{\mathscr{H}}(\mathbf{x}^*, \mathbf{u}^*, \lambda) \geq \bar{\mathscr{H}}(\mathbf{x}, \mathbf{u}, \lambda) \quad \textit{a.e. in } t. \tag{I.7}$$

Proof. Following Leitmann and Stalford [1971] and its infinite-horizon extensions in Goenka et al. [2014, 2024], freeze Q at its optimal-path value $Q^*(t)$. The *frozen-Q Hamiltonian*

$$\mathscr{H}_c(\mathbf{x}, \mathbf{u}, \lambda; Q^*) \equiv u(Y - \Gamma(q)) + \lambda_F[\Lambda_F D(Q^*) A^{a_F} (\Omega_F H)^{\alpha_F} F^{\xi} - \delta_F F + qs\kappa R] + \textstyle\sum_{j \neq F} \lambda_j \dot{x}_j$$

treats $D(Q^*)$ as a known function of time. Freezing Q eliminates the source of non-concavity (Proposition I.1): the CES aggregator is jointly concave in (F, R, A, H); each accumulation function is concave in its argument ($\xi, \xi_R, \omega \in (0,1)$). Hence $\mathscr{H}_c$ is jointly concave in $(\mathbf{x}, \mathbf{u})$. The decomposition parallels Goenka et al. [2024], where freezing the infection share at i^* recovers a concave core; here $Q = F/(F+R)$ plays the role of i, and $D(Q)$ the role of the contact rate.

The augmented frozen-Q Hamiltonian $\bar{\mathscr{H}}_c \equiv \mathscr{H}_c + \langle \dot{\lambda}, \mathbf{x} \rangle$ inherits joint concavity (the augmentation is linear). Arrow sufficiency gives

$$\bar{\mathscr{H}}_c(\mathbf{x}^*, \mathbf{u}^*; Q^*) \geq \bar{\mathscr{H}}_c(\mathbf{x}, \mathbf{u}; Q^*). \tag{I.8}$$

The perturbation $\mathscr{P} \equiv \mathscr{H} - \mathscr{H}_c$ collects all Q-dependent terms:

$$|\mathscr{P}| \leq |\lambda_F|\, \bar{G}_F\, |D(Q) - D(Q^*)| + |\lambda_{\mathscr{E}}^{\text{pub}}|\, \bar{\Lambda}_{\mathscr{E}}\, |D_{\mathscr{E}}(Q) - D_{\mathscr{E}}(Q^*)| + |\lambda_{\mathscr{E}}^{\text{pub}}|\, \bar{\mathscr{E}}\, |\delta_{\mathscr{E}}(\varphi) - \delta_{\mathscr{E}}(\varphi^*)|, \tag{I.9}$$

uniformly bounded: $Q \in [0,1]$, D and $D_{\mathscr{E}}$ are Lipschitz on $[0,1]$, and all remaining factors are capped by Lemmas I.2 and I.6. Along the optimal path $\mathscr{P}^* = 0$ (since $Q = Q^*$).

Combining (I.8) with the perturbation: the concave-core difference is ≥ 0; the perturbation difference equals $0 - \mathscr{P}(\mathbf{x}, \mathbf{u}, \lambda)$, bounded by (I.9). Integrating over $[0, \infty)$ with discount $e^{-\rho t}$: the nonneg concave-core integral absorbs the (bounded, discounted) perturbation integral, delivering the integrated augmented inequality and, by density of evaluation times, the pointwise inequality (I.7). □ □

Remark I.10. Proposition I.9 is weaker than concavity of the maximised Hamiltonian $M(\mathbf{x},\lambda)$ (Arrow's condition) and weaker than joint concavity of $\mathscr{H}$ (Mangasarian).

Theorem I.11 (Sufficiency)**.** *Let* $(\mathbf{x}^*,\mathbf{u}^*)$ *be an interior path satisfying the necessary conditions: the labour-allocation FOC (Appendix E.1.2), the governance FOC* (26)*, and the costate system* (E.1)–(E.6)*, together with the standard TVCs* (I.3)*. Then* $(\mathbf{x}^*,\mathbf{u}^*)$ *is optimal among all feasible paths from* $\mathbf{x}_0$*.*

Proof. Proposition I.9 gives $\mathscr{H}^* - \mathscr{H} + \langle\dot{\lambda},\mathbf{x}^*-\mathbf{x}\rangle \geq 0$. Substituting $\mathscr{H} = u(C) + \langle\lambda,\dot{\mathbf{x}}\rangle$ and applying the product rule $\langle\lambda,\dot{\mathbf{x}}^*-\dot{\mathbf{x}}\rangle + \langle\dot{\lambda},\mathbf{x}^*-\mathbf{x}\rangle = \frac{d}{dt}\langle\lambda,\mathbf{x}^*-\mathbf{x}\rangle$:

$$u(C^*) - u(C) + \frac{d}{dt}\big\langle\lambda,\mathbf{x}^*-\mathbf{x}\big\rangle \geq 0.$$

Multiply by $e^{-\rho t}$, integrate on $[0,T]$, use $\mathbf{x}^*(0) = \mathbf{x}(0)$:

$$\int_0^T e^{-\rho t}[u(C^*) - u(C)]\,dt \geq -e^{-\rho T}\big\langle\lambda(T),\mathbf{x}^*(T)-\mathbf{x}(T)\big\rangle.$$

Taking $T\to\infty$ and applying the generalised TVC (I.5): $\int_0^\infty e^{-\rho t}[u(C^*) - u(C)]\,dt \geq 0$. □

Corollary I.12. *The planner allocation characterised in Section 3.2 is globally optimal.*

Proof. Proposition I.3 delivers existence. The candidate path satisfies the necessary conditions and (I.3); Theorem I.11 selects it. □ □

References

Daron Acemoglu. *Introduction to Modern Economic Growth*. Princeton University Press, Princeton, NJ, 2009.

Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542, 2018.

Daron Acemoglu and Pascual Restrepo. Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6):2188–2244, 2020.

Philippe Aghion, Benjamin F. Jones, and Charles I. Jones. Artificial intelligence and economic growth. In Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors, *The Economics of Artificial Intelligence: An Agenda*, pages 237–282. University of Chicago Press, 2018a.

Philippe Aghion, Benjamin F. Jones, and Charles I. Jones. Artificial intelligence and economic growth. *NBER Working Paper No. 23928*, 2018b.

George A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970.

Pertti Alasuutari, Marjaana Rautalin, and Jukka Syväterä. Organisations as epistemic capital: The case of independent regulatory agencies. *International Journal of Politics, Culture, and Society*, 29(1):53–71, 2016.

Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luber, Ahmed Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard G. Baraniuk. Self-consuming generative models go MAD. *arXiv:2307.01850*, 2023.

Paul Blanchard, Robert L. Devaney, and Glen R. Hall. *Differential Equations*. Cengage Learning, 4th edition, 2012.

Nicholas Bloom, Charles I. Jones, John Van Reenen, and Michael Webb. Are ideas getting harder to find? *American Economic Review*, 110(4):1104–1144, 2020.

Mark Carney. Breaking the tragedy of the horizon—climate change and financial stability, 2015. Speech at Lloyd's of London, 29 September 2015.

Pierre Cartigny and Philippe Michel. On a sufficient transversality condition for infinite horizon optimal control problems. *Automatica*, 39(6):1007–1010, 2003.

Cindy Chung and Laura Veldkamp. Data and the aggregate economy. *Journal of Economic Literature*, 62(4):1597–1639, 2024.

Lin William Cong, Danxia Xie, and Longtian Zhang. Knowledge accumulation, privacy, and growth in a data economy. *Management Science*, 67(10):6480–6492, 2021.

Partha Dasgupta and Paul A. David. Toward a new economics of science. *Research Policy*, 23(5):487–521, 1994.

Partha Dasgupta and Geoffrey Heal. The optimal depletion of exhaustible resources. *Review of Economic Studies*, 41:3–28, 1974.

Laurens de Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

Glenn Ellison. The slowdown of the economics publishing process. *Journal of Political Economy*, 110(5):947–993, 2002.

Maryam Farboodi and Laura Veldkamp. A growth model of the data economy. *Review of Economic Studies*, 2025. Forthcoming.

Wendell H. Fleming and Raymond W. Rishel. *Deterministic and Stochastic Optimal Control*. Springer, 1975.

Maximilian Fochler, Ulrike Felt, and Ruth Müller. Unsustainable growth, hyper-competition, and worth in life science research: Narrowing evaluative repertoires in doctoral and postdoctoral scientists' work and lives. *Minerva*, 54(2):175–200, 2016.

Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David L. Donoho, and Sanmi Koyejo. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. *arXiv:2404.01413*, 2024.

Aditya Goenka, Lin Liu, and Manh-Hung Nguyen. Infectious diseases and economic growth. *Journal of Mathematical Economics*, 50:34–53, 2014.

Aditya Goenka, Lin Liu, and Manh-Hung Nguyen. Modelling optimal quarantines with waning immunity. *Economic Theory*, 77(1):197–234, 2024.

Charles A. E. Goodhart. Problems of monetary management: The U.K. experience. *Monetary Theory and Practice*, pages 91–121, 1984.

David M. Green and John A. Swets. *Signal Detection Theory and Psychophysics*. Wiley, 1966.

Gustaf Gripenberg, Stig-Olof Londen, and Olof Staffans. *Volterra Integral and Functional Equations*, volume 34 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1990.

Chang-Tai Hsieh, Erik Hurst, Charles I. Jones, and Peter J. Klenow. The allocation of talent and U.S. economic growth. *Econometrica*, 87(5):1439–1474, 2019.

Charles I. Jones. R&d-based models of economic growth. *Journal of Political Economy*, 103(4):759–784, 1995.

Charles I. Jones. The past and future of economic growth: A semi-endogenous perspective. *Annual Review of Economics*, 14:125–152, 2022.

Charles I. Jones. A.I. and our economic future. Working Paper 34779, National Bureau of Economic Research, 2026.

Charles I. Jones and Christopher Tonetti. Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858, 2020.

Dmitry Kobak, Rita González-Márquez, Emőke Ágnes Horvát, and Jan Lause. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27):eadt3813, 2025. doi: 10.1126/sciadv.adt3813.

Anton Korinek. Language models and cognitive automation for economic research. *Journal of Economic Literature*, 61(4):1281–1317, 2023.

Samuel S. Kortum. Research, patenting, and technological change. *Econometrica*, 65(6): 1389–1419, 1997.

Yuri A. Kuznetsov. *Elements of Applied Bifurcation Theory*. Springer, 3rd edition, 2004.

Yifan Lei, Pedro Pallares-Yaurivilca, et al. Global retractions due to randomly generated content: Characterization and trends. *Accountability in Research*, 2024. doi: 10.1080/08989621.2024.2413498. Forthcoming.

George Leitmann and Harold Stalford. A sufficiency theorem for optimal control. *Journal of Optimization Theory and Applications*, 8(3):169–174, 1971.

Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, Daisy Ding, Xinyu Yang, Kailas Vodrahalli, Siyu He, Daniel Smith, Yian Yin, Daniel McFarland, and James Y. Zou. Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. In *Forty-First International Conference on Machine Learning*, 2024. ICML 2024.

Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D. Manning, and James Y. Zou. Quantifying large language model usage in scientific papers. *Nature Human Behaviour*, 9:2599–2609, 2025. doi: 10.1038/s41562-025-02273-8.

Jr. Lucas, Robert E. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46, 1976.

Jr. Lucas, Robert E. On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42, 1988.

Gustavo Manso. Motivating innovation. *Journal of Finance*, 66(5):1823–1860, 2011.

Philippe Michel. On the transversality condition in infinite horizon optimal problems. *Econometrica*, 50(4):975–985, 1982.

Kevin M. Murphy, Andrei Shleifer, and Robert W. Vishny. The allocation of talent: Implications for growth. *Quarterly Journal of Economics*, 106(2):503–530, 1991.

Mitio Nagumo. Über die Lage der Integralkurven gewöhnlicher Differentialgleichungen. *Proceedings of the Physico-Mathematical Society of Japan*, 24:551–559, 1942.

Manh-Hung Nguyen and Phu Nguyen-Van. Optimal endogenous growth with natural resources: Theory and evidence. *Macroeconomic Dynamics*, 20(8):2173–2209, 2016. doi: 10.1017/S1365100515000061.

William D. Nordhaus. Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 114(7):1518–1523, 2017.

Elinor Ostrom. *Governing the Commons*. Cambridge University Press, 1990.

Aji Raditya Pratama. The accuracy-bias trade-offs in AI text detection tools and their impact on fairness in scholarly publication. *PeerJ Computer Science*, 11:e2953, 2025. doi: 10.7717/peerj-cs.2953.

Paul M. Romer. Increasing returns and long-run growth. *Journal of Political Economy*, 94 (5):1002–1037, 1986. doi: 10.1086/261420.

Paul M. Romer. Endogenous technological change. *Journal of Political Economy*, 98(5): S71–S102, 1990.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. Ai models collapse when trained on recursively generated data. *Nature*, 631: 755–759, 2024.

Paula Stephan. *How Economics Shapes Science*. Harvard University Press, 2012.

Nicholas Stern. *The Economics of Climate Change: The Stern Review*. Cambridge University Press, 2007.

Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

Donald M. Topkis. *Supermodularity and Complementarity*. Princeton University Press, 1998.

Philip Trammell and Anton Korinek. Economic growth under transformative AI. *Working Paper*, 2024.

Carolina Tropini, Finlay Maguire, Alejandro Vergara-Camus, Amy Lee, Joerg Graf, and Kerry Savage. Time to rethink academic publishing: The peer reviewer crisis. *mBio*, 14 (6):e01091–23, 2023. doi: 10.1128/mbio.01091-23.

Hirofumi Uzawa. Optimum technical change in an aggregative model of economic growth. *International Economic Review*, 6(1):18–31, 1965.

Richard Van Noorden. More than 10,000 research papers were retracted in 2023—a new record. *Nature*, 624:479–481, 2023. doi: 10.1038/d41586-023-03974-8.

Martin L. Weitzman. Recombinant growth. *Quarterly Journal of Economics*, 113(2):331–360, 1998.