Automation Thresholds and Regime Transitions in AI-Driven Economic Growth
Proofs Appendices

Appendix A: Theoretical Proofs

A.1 Proof of Proposition (Balanced Growth without AI)

Proposition 1. *(Balanced Growth without AI).*

Proof of Proposition in the main text

A.2 Proof of Proposition 2 (BGP with Non-Improving AI)

Proposition 2 (Balanced Growth with Non-Improving AI). *With $\beta = 0$ and $\gamma = 0$, if human researchers grow at rate n and AI capital grows at rate g_M, then:*

- *If $g_M > n$: In the long run, AI dominates and $g_A^* = \frac{\lambda g_M}{1-\phi}$*
- *If $g_M < n$: Humans dominate and $g_A^* = \frac{\lambda n}{1-\phi}$ (same as no-AI case)*
- *More generally: $g_A^* = \frac{\lambda g_R}{1-\phi}$ where $g_R = (1-\psi)n + \psi g_M$ depends on research shares*

The key insight is that AI raises *the growth rate above the no-AI benchmark $\frac{\lambda n}{1-\phi}$ provided $g_M > n$.*

Proof. With $\beta = 0$ and $\gamma = 0$, AI efficiency is constant: $\eta(M, C) = \bar{\eta}$. The idea production function is:

$$\dot{A} = \delta[L_A + \bar{\eta}M]^\lambda A^\phi$$

where total research input is $R = L_A + \bar{\eta}M$.

Step 1: Dynamics of research input

Human researchers grow at population rate: $L_A(t) = \sigma L_0 e^{nt}$ and AI capital grows by assumption: $M(t) = M_0 e^{g_M t}$.

Total research input:

$$R(t) = \sigma L_0 e^{nt} + \bar{\eta} M_0 e^{g_M t}$$

The growth rate of R is:

$$\frac{\dot{R}}{R} = \frac{\sigma L_0 e^{nt} \cdot n + \bar{\eta} M_0 e^{g_M t} \cdot g_M}{\sigma L_0 e^{nt} + \bar{\eta} M_0 e^{g_M t}}$$

Let $\psi(t) = \frac{\bar{\eta} M_0 e^{g_M t}}{\sigma L_0 e^{nt} + \bar{\eta} M_0 e^{g_M t}}$ be AI's research share. Then $1 - \psi(t) = \frac{\sigma L_0 e^{nt}}{\sigma L_0 e^{nt} + \bar{\eta} M_0 e^{g_M t}}$ is the human share.

Therefore:

$$g_R(t) = \frac{\dot{R}}{R} = (1 - \psi(t))n + \psi(t) g_M$$

Step 2: Long-run behavior of research shares

Case 1: If $g_M > n$, then as $t \to \infty$:

$$\psi(t) = \frac{\bar{\eta} M_0 e^{g_M t}}{\sigma L_0 e^{nt} + \bar{\eta} M_0 e^{g_M t}} = \frac{1}{1 + \frac{\sigma L_0}{\bar{\eta} M_0} e^{(n - g_M)t}} \to 1$$

So AI dominates and $g_R \to g_M$.

Case 2: If $g_M < n$, then $\psi(t) \to 0$ and humans dominate: $g_R \to n$.

Case 3: If $g_M = n$, then $\psi(t) \to \psi^* = \frac{\bar{\eta} M_0}{\sigma L_0 + \bar{\eta} M_0}$ (constant), giving:

$$g_R = (1 - \psi^*)n + \psi^* n = n$$

Step 3: General case with constant research growth

For the general statement, we consider the **effective long-run research growth rate**. Define the research-weighted average:

$$\bar{g}_R = \lim_{T\to\infty} \frac{1}{T} \int_0^T g_R(t)\, dt$$

However, for a balanced growth path to exist, we need g_R to asymptotically approach a constant. From Step 2, this occurs in all three cases:

- If $g_M > n$: $g_R \to g_M$
- If $g_M < n$: $g_R \to n$
- If $g_M = n$: $g_R = n$ always

Step 4: Deriving the balanced growth rate

The idea production function is:

$$\dot{A} = \delta R^\lambda A^\phi$$

Define the growth rate $g_A = \dot{A}/A$:

$$g_A = \delta R^\lambda A^{\phi-1}$$

Taking logs:

$$\ln g_A = \ln \delta + \lambda \ln R + (\phi - 1) \ln A$$

Differentiating with respect to time:

$$\frac{\dot{g}_A}{g_A} = \lambda \frac{\dot{R}}{R} + (\phi - 1)\frac{\dot{A}}{A} = \lambda g_R + (\phi - 1) g_A$$

Therefore:

$$\frac{dg_A}{dt} = \lambda g_A g_R + (\phi - 1) g_A^2$$

On a balanced growth path, g_A is constant ($\dot{g}_A = 0$), so:

$$0 = \lambda g_R + (\phi - 1) g_A^*$$

Solving for g_A^*:

$$g_A^* = \frac{\lambda g_R}{1 - \phi}$$

□

A.3 Proof of Proposition 3 Automation Threshold)

Proposition 3 (Automation Threshold). There exists a critical AI research share

$$\psi^* = \frac{1 - \phi}{\lambda(1 + \beta)}$$

such that growth is semi-endogenous for $\psi < \psi^*$ and becomes explosive for $\psi > \psi^*$.

Proof. The objective is to characterize the conditions under which the growth rate of ideas transitions from stable exponential dynamics to explosive (hyperbolic) behavior. Central to this transition is the role of AI in research production.

The AI research share is defined as

$$\psi(t) = \frac{\eta(M(t), C(t))M(t)}{L_A(t) + \eta(M(t), C(t))M(t)},$$

which measures the fraction of effective research effort contributed by AI systems. When ψ is small, research is predominantly human-driven; when ψ approaches one, AI dominates research production.

Assumption 1 (Quasi-Steady State)**.** *Along the transition path, compute and AI capital adjust such that:*

$$\frac{\dot{\eta}}{\eta} = \beta g_A + o(g_A), \quad \frac{\dot{M}}{M} = g_A + o(g_A)$$

where $o(g_A)$ denotes higher-order terms that become negligible as g_A grows. These hold when depreciation rates are small relative to growth rates.

In the regime where both human and AI researchers contribute, idea production is given by

$$\dot{A}(t) = \delta \left[L_A(t) + \eta(t)M(t)\right]^{\lambda} A(t)^{\phi},$$

where $\eta(t) = \bar{\eta}C(t)^{\beta}$ and congestion effects are suppressed for clarity. Let $g_A \equiv \dot{A}/A$ denote the growth rate of ideas. Taking logs and differentiating yields

$$\frac{\dot{g}_A}{g_A} = \frac{\lambda}{1-\phi}\frac{d}{dt}\ln[L_A + \eta M] + \frac{\phi - 1}{1-\phi}g_A.$$

The evolution of total effective research input satisfies

$$\frac{d}{dt}\ln(L_A + \eta M) = (1-\psi)n + \psi\left(\frac{\dot{\eta}}{\eta} + \frac{\dot{M}}{M}\right),$$

where population growth implies $\dot{L}_A/L_A = n$ and the definition of ψ has been used.

Since $\eta = \bar{\eta}C^{\beta}$ and compute accumulates according to $\dot{C} = \theta s_M Y$, it follows that

$$\frac{\dot{\eta}}{\eta} = \beta\frac{\dot{C}}{C}.$$

Along balanced growth paths and in their neighborhood, output is proportional to the stock of ideas, so $Y = \Omega A$ for some constant Ω. Standard arguments then imply that the ratio C/A converges to a quasi–steady state, yielding

$$\frac{\dot{\eta}}{\eta} \simeq \beta g_A.$$

Similarly, AI capital accumulation implies $\dot{M}/M \simeq g_M \simeq g_Y \simeq g_A$ once growth becomes sufficiently rapid.

Substituting these expressions into the growth-rate equation gives

$$\frac{dg_A}{dt} = \frac{\lambda(1-\psi)n}{1-\phi}g_A + \frac{g_A^2}{1-\phi}\big[\lambda\psi(1+\beta) + \phi - 1\big]. \tag{1}$$

Equation (1) has two components. The first term is linear in g_A and reflects standard semi-endogenous growth driven by population expansion. The second term is quadratic in g_A and captures recursive feedback from AI-driven research. For sufficiently large g_A, the quadratic term dominates the dynamics.

Explosive growth requires the coefficient of g_A^2 to be positive:

$$\kappa(\psi) = \lambda\psi(1+\beta) + \phi - 1 > 0$$

Solving this condition for $\kappa(\psi^*) = 0$ yields the critical threshold

$$\psi^* = \frac{1-\phi}{\lambda(1+\beta)}.$$

When $\psi > \psi^*$, the quadratic term in (1) is positive and growth accelerates without bound, generating hyperbolic dynamics and a finite-time singularity. When $\psi < \psi^*$, the quadratic term is negative, offsetting the linear component and ensuring convergence to a stable semi-endogenous growth path. This establishes the result. □

Corollary 4 (Impossibility of Explosive Growth)**.** *If $\psi^* > 1$, explosive growth cannot occur for any feasible AI research share $\psi \in [0, 1]$. This happens when:*

$$\frac{1-\phi}{\lambda(1+\beta)} > 1 \iff \lambda(1+\beta) < 1-\phi$$

Stability near threshold: For $\psi < \psi^*$, the quadratic term in (1) is negative, forcing $dg_A/dt < 0$ when g_A is large. The economy converges to a standard BGP. For $\psi > \psi^*$, the positive quadratic term dominates, making the BGP unstable and triggering divergence.

The threshold ψ^* balances two forces:

- **Stabilizing force:** Diminishing returns to knowledge stock $(1-\phi)$ slow growth
- **Amplifying force:** AI's contribution to research $(\lambda(1+\beta))$ accelerates growth

When AI's share ψ crosses ψ^*, amplification dominates, creating positive feedback. This threshold has a transparent interpretation: explosive growth requires AI's amplification effect, $\lambda\psi(1+\beta)$, to outweigh the drag from diminishing returns to ideas, $1-\phi$. Parameter configurations with strong diminishing returns or weak scaling laws may imply $\psi^* > 1$, in which case explosive growth is infeasible. Conversely, higher research returns or stronger AI scaling substantially lower the threshold, making explosive dynamics possible at empirically plausible AI research shares.

□

A.4 Proof of Proposition 4 (Hyperbolic Growth)

Proposition 5 (**Hyperbolic Growth**)**.** *Suppose $\beta > 0$, $\phi < 1$, and the economy maintains constant investment share $s_M > 0$. If AI's research share satisfies $\psi > \psi^* = \frac{1-\phi}{\lambda(1+\beta)}$, then growth follows a hyperbolic trajectory:*

$$g_A(t) = \frac{g_0}{1-\kappa(\psi)g_0(t-t_0)}, \quad \kappa(\psi) = \frac{\lambda\psi(1+\beta)+\phi-1}{1-\phi}$$

Proof. **Step 1: AI-Dominated Regime.** When $\psi \approx 1$, AI dominates research production. The idea production function becomes:

$$\dot{A} = \delta[\bar{\eta}C^{\beta}M]^{\lambda}A^{\phi}$$

Step 2: Compute-Ideas Relationship. From $\dot{C} = \theta s_M Y$ and $Y = \Omega A$ (constant capital-output ratio), we have:

$$C(t) = \theta s_M \Omega \int_0^t A(s)ds$$

For large t and slowly varying g_A, the integral is dominated by recent values:

$$\int_0^t A(s)ds \approx \frac{A(t)}{g_A(t)}$$

Thus:

$$C(t) \approx \frac{\theta s_M \Omega}{g_A(t)} A(t)$$

Step 3: AI Capital Relationship. AI capital accumulates according to $\dot{M} = \xi s_M Y - \delta_M M$. For large growth rates, depreciation is negligible, giving:

$$M(t) \propto Y(t) = \Omega A(t)$$

Let $M(t) = \nu A(t)$ where ν is the proportionality constant.

Step 4: Substitute into Production Function. Substituting $M = \nu A$ and $C = (\theta s_M \Omega / g_A)A$:

$$\dot{A} = \delta \left[\bar{\eta}\left(\frac{\theta s_M \Omega}{g_A} A\right)^{\beta} (\nu A)\right]^{\lambda} A^{\phi}$$

$$\dot{A} = \delta \bar{\eta}^{\lambda} \nu^{\lambda} (\theta s_M \Omega)^{\beta\lambda} g_A^{-\beta\lambda} A^{\beta\lambda+\lambda+\phi}$$

Step 5: Derive Growth Rate Equation. Define $g_A = \dot{A}/A$:

$$g_A = \delta\bar{\eta}^{\lambda}\nu^{\lambda}(\theta s_M \Omega)^{\beta\lambda} g_A^{-\beta\lambda} A^{\beta\lambda+\lambda+\phi-1}$$

Taking logs:

$$\ln g_A = \ln(\delta\bar{\eta}^{\lambda}\nu^{\lambda}(\theta s_M \Omega)^{\beta\lambda}) - \beta\lambda \ln g_A + (\beta\lambda + \lambda + \phi - 1)\ln A$$

Differentiating with respect to time:

$$\frac{\dot{g}_A}{g_A} = -\beta\lambda\frac{\dot{g}_A}{g_A} + (\beta\lambda + \lambda + \phi - 1)g_A$$

Solving for $\dot{g}_A$:

$$\frac{\dot{g}_A}{g_A}(1 + \beta\lambda) = (\lambda(1+\beta) + \phi - 1)g_A$$

$$\frac{dg_A}{dt} = \frac{\lambda(1+\beta) + \phi - 1}{1 + \beta\lambda} g_A^2$$

Step 6: Generalize to $\psi \in (0, 1)$. For general research share ψ, the effective research input is $R = L_A + \eta M$. The growth dynamics become:

$$\frac{dg_A}{dt} = \frac{\lambda(1-\psi)n}{1-\phi} g_A + \frac{g_A^2}{1-\phi}[\lambda\psi(1+\beta) + \phi - 1]$$

In the AI-dominated regime ($\psi \to 1$), the linear term becomes negligible and we obtain:

$$\frac{dg_A}{dt} = \kappa(\psi)g_A^2, \quad \kappa(\psi) = \frac{\lambda\psi(1+\beta) + \phi - 1}{1-\phi}$$

Step 7: Solve Differential Equation. Solving $\dot{g}_A = \kappa g_A^2$ with initial condition $g_A(t_0) = g_0$:

$$g_A(t) = \frac{g_0}{1 - \kappa(\psi)g_0(t - t_0)}$$

The solution exhibits a finite-time singularity at $T = t_0 + 1/(\kappa(\psi)g_0)$ when $\kappa(\psi) > 0$, which occurs precisely when $\psi > \psi^*$. □

Corollary 6 (Impossibility of Explosive Growth)**.** *If $\psi^* > 1$, explosive growth cannot occur for any feasible AI research share $\psi \in [0, 1]$. This happens when:*

$$\frac{1-\phi}{\lambda(1+\beta)} > 1 \iff \lambda(1+\beta) < 1 - \phi$$

□

A.5 Proof of Proposition 5 (Complementarity Effects))

Proposition 7 (**Complementarity Effects**)**.** *With CES complementarity ($\rho < 1$), the automation threshold satisfies $\psi^*_{CES} > \psi^*_{baseline}$.*

Proof. When human and AI researchers are imperfect substitutes, we model effective research input using a CES aggregator:

$$R = \left[L_A^{\rho} + (\eta M)^{\rho}\right]^{1/\rho}, \quad \rho \in (-\infty, 1]$$

The elasticity of substitution is $\sigma = \frac{1}{1-\rho}$. Lower ρ implies stronger complementarity.

Step 1: Derive marginal product of AI research.

The marginal product of AI research capital is:

$$\frac{\partial R}{\partial M} = \frac{1}{\rho}\left[L_A^{\rho} + (\eta M)^{\rho}\right]^{\frac{1-\rho}{\rho}} \cdot \rho(\eta M)^{\rho-1}\eta = \eta\left(\frac{R}{\eta M}\right)^{1-\rho}$$

Step 2: Compare to baseline case.

In the baseline model with perfect substitutes ($\rho = 1$), we have $\frac{\partial R}{\partial M} = \eta$. The ratio of marginal products is:

$$\frac{\partial R/\partial M|_{CES}}{\partial R/\partial M|_{baseline}} = \left(\frac{R}{\eta M}\right)^{1-\rho} = \psi^{-(1-\rho)}$$

where $\psi = \frac{\eta M}{L_A + \eta M}$ is the AI research share.

Since $\rho < 1$ implies $(1-\rho) > 0$, this ratio is **less than 1** for any $\psi \in (0, 1)$. Thus, complementarity **reduces** AI's marginal contribution to research for a given research share.

Step 3: Derive effective research elasticity.

The key term in the growth dynamics is the elasticity of idea production with respect to AI research:

$$\varepsilon_{AI} \equiv \frac{\partial \dot{A}}{\partial M}\frac{M}{\dot{A}} = \lambda \frac{\partial R}{\partial M}\frac{M}{R}$$

For the CES case:

$$\frac{\partial R}{\partial M}\frac{M}{R} = \eta \left(\frac{R}{\eta M}\right)^{1-\rho} \cdot \frac{M}{R} = \left(\frac{\eta M}{R}\right)^{\rho} = \psi^{\rho}$$

Step 4: Threshold condition with complementarity.

The acceleration condition from Proposition 4 generalizes to:

$$\varepsilon_{AI}(1+\beta) > 1 - \phi$$

Substituting $\varepsilon_{AI} = \lambda\psi^{\rho}$:

$$\lambda\psi^{\rho}(1+\beta) > 1 - \phi$$

Solving for the critical threshold:

$$\psi^*_{CES} = \left[\frac{1-\phi}{\lambda(1+\beta)}\right]^{1/\rho}$$

Step 5: Compare to baseline.

The baseline threshold (perfect substitutes) is:

$$\psi^*_{baseline} = \frac{1-\phi}{\lambda(1+\beta)}$$

Since $\rho < 1$ implies $1/\rho > 1$, we have:

$$\psi^*_{CES} = (\psi^*_{baseline})^{1/\rho} > \psi^*_{baseline}$$

□

Appendix B: Extended Model Proofs

B.1 Proof of Proposition 6 (Energy-Bounded Growth)

Proposition 8 (Energy-Bounded Growth)**.** *Suppose energy availability grows at rate $g_E < \infty$ and efficiency improvements satisfy $\epsilon(t) = \max\{\epsilon_{Landauer}, \epsilon_0 e^{-\mu t}\}$ where $\mu > 0$ initially but $\mu \to 0$ as $t \to \infty$. Then the long-run growth rate of ideas satisfies:*

$$\limsup_{t\to\infty} g_A(t) \leq \frac{\lambda(1+\beta)(g_E + \mu(t))}{1-\phi}$$

In particular, as efficiency improvements exhaust ($\mu \to 0$):

$$\lim_{t\to\infty} g_A(t) = \frac{\lambda(1+\beta)g_E}{1-\phi}$$

Proof. The proof establishes that once energy constraints bind, the explosive feedback mechanism from Proposition 3 saturates. Growth then depends solely on exogenous rates g_E and μ, which are bounded.

Step 1: The energy constraint binds in finite time

Compute accumulation is governed by:

$$\dot{C}(t) = \min\left\{\theta s_M Y(t), \frac{E_{\max}(t)}{\epsilon(t)}\right\} \tag{2}$$

From Proposition 3, in the unconstrained regime with $\psi > \psi^*$, output grows hyperbolically: $Y(t) \sim (T-t)^{-1/\kappa}$ as $t \to T$. Meanwhile, energy-limited compute grows at most exponentially:

$$\frac{E_{\max}(t)}{\epsilon(t)} = \frac{E_0 e^{g_E t}}{\max\{\epsilon_{\text{Landauer}}, \epsilon_0 e^{-\mu t}\}} \leq \frac{E_0 e^{g_E t}}{\epsilon_{\text{Landauer}}}$$

Since hyperbolic growth eventually dominates exponential growth, there exists finite $t_1 < T$ such that for all $t > t_1$, the energy constraint binds:

$$\dot{C}(t) = \frac{E_{\max}(t)}{\epsilon(t)} \tag{3}$$

Step 2: Growth rate of energy-constrained compute

Lemma A (Compute Growth Under Energy Constraints)**.** *When equation (3) holds:*

- *If $\epsilon(t) = \epsilon_0 e^{-\mu t}$ (before Landauer limit): $g_C(t) = g_E + \mu$*
- *If $\epsilon(t) = \epsilon_{Landauer}$ (at Landauer limit): $g_C(t) = g_E$*

Proof. **Phase 1:** With $\epsilon(t) = \epsilon_0 e^{-\mu t}$, equation (3) gives:

$$\dot{C}(t) = \frac{E_0 e^{g_E t}}{\epsilon_0 e^{-\mu t}} = \frac{E_0}{\epsilon_0} e^{(g_E+\mu)t}$$

Integrating:

$$C(t) = C_0 + \frac{E_0}{\epsilon_0}\int_0^t e^{(g_E+\mu)s} ds = C_0 + \frac{E_0}{\epsilon_0(g_E+\mu)}[e^{(g_E+\mu)t} - 1]$$

For large t: $C(t) \sim \frac{E_0}{\epsilon_0(g_E+\mu)} e^{(g_E+\mu)t}$

Therefore: $g_C(t) = \frac{\dot{C}(t)}{C(t)} = g_E + \mu$

Phase 2: With $\epsilon(t) = \epsilon_{\text{Landauer}}$ (constant):

$$\dot{C}(t) = \frac{E_0 e^{g_E t}}{\epsilon_{\text{Landauer}}}$$

Integrating:

$$C(t) = C_0 + \frac{E_0}{\epsilon_{\text{Landauer}}} \int_0^t e^{g_E s} ds = C_0 + \frac{E_0}{\epsilon_{\text{Landauer}} g_E} (e^{g_E t} - 1)$$

For large t: $C(t) \sim \frac{E_0}{\epsilon_{\text{Landauer}} g_E} e^{g_E t}$
Therefore: $g_C(t) = \frac{\dot{C}(t)}{C(t)} = g_E$ □

Step 3: Dynamics of idea growth under energy constraints
In the AI-dominated regime ($\psi \approx 1$), the idea production function is $\dot{A} = \delta[\bar{\eta} C^\beta M]^\lambda A^\phi$ where AI efficiency is $\eta = \bar{\eta} C^\beta$ and AI capital satisfies $M \propto A$. A key assumption (formalized below) is that AI capital accumulation is also energy-constrained, so $g_M \leq g_C$.

Assumption 2 (Energy-Constrained AI Capital)**.** *When energy constraints bind, AI capital growth satisfies $g_M(t) \leq g_C(t)$. In the limiting case, $g_M(t) \to g_C(t)$.*

Taking logarithms and differentiating the idea production function:

$$\frac{d}{dt} \ln \dot{A} = \lambda \beta g_C + \lambda g_M + \phi g_A$$

Since $\frac{d}{dt} \ln \dot{A} = \frac{\dot{g}_A}{g_A} + g_A$, we obtain:

$$\frac{\dot{g}_A}{g_A} = \lambda \beta g_C + \lambda g_M + (\phi - 1) g_A$$

Multiplying by g_A:

$$\frac{dg_A}{dt} = \lambda \beta g_C g_A + \lambda g_M g_A + (\phi - 1) g_A^2 \tag{4}$$

Under energy constraints with $g_M \leq g_C$ (Assumption 1), the limiting dynamics are:

$$\frac{dg_A}{dt} = \lambda (1 + \beta) g_C g_A + (\phi - 1) g_A^2 \tag{5}$$

Step 4: Steady-state growth bound
Once energy constraints bind, g_C becomes constant (either $g_E + \mu$ or g_E). At steady state, $\frac{dg_A}{dt} = 0$:

$$0 = \lambda(1 + \beta) g_C g_A^* + (\phi - 1)(g_A^*)^2$$

Factoring out g_A^* and solving:

$$g_A^* = \frac{\lambda(1 + \beta) g_C}{1 - \phi}$$

Substituting $g_C = g_E + \mu$ (Phase 1) or $g_C = g_E$ (Phase 2):

$$g_A^* = \frac{\lambda(1 + \beta)(g_E + \mu)}{1 - \phi}$$

As $\mu \to 0$:

$$\lim_{t \to \infty} g_A(t) = \frac{\lambda(1 + \beta) g_E}{1 - \phi}$$

With baseline parameters ($\lambda = 0.5$, $\beta = 0.08$, $\phi = 0.5$, $g_E = 0.03$):

$$g_{\max} = \frac{0.5 \times 1.08 \times 0.03}{0.5} = 0.0324 = 3.24\%$$

This establishes the long-run bound on growth rates under energy constraints. □

Remark on transitional dynamics

At $t = t_1$ when the energy constraint binds, the economy transitions smoothly from hyperbolic to energy-limited dynamics. The growth rate $g_A(t)$ peaks at t_1 and then asymptotically converges to g_A^* as derived above.

□

B Appendixe C

C.0 Proof of Proposition 7 (Vanishing Labor Share Under Full Automation)

Proposition 9. *(Vanishing Labor Share Under Full Automation).*
Proof of Proposition in the main text

C.1 Proof of Proposition 8 (Wage Polarization Under Sequential Automation)

Assumption 3 (Sequential Automation)**.** *The rate of improvement in AI effectiveness varies across skill groups:*

$$\frac{d \ln \bar{\eta}_M}{dt} > \frac{d \ln \bar{\eta}_L}{dt} > \frac{d \ln \bar{\eta}_H}{dt} \tag{6}$$

with potentially $\frac{d \ln \bar{\eta}_H}{dt} < 0$ *if high-skill tasks become relatively harder to automate as AI capabilities advance.*

Proposition 10 (Wage Polarization Under Sequential Automation)**.** *Under Assumption 3, wage dynamics satisfy:*

1. ***Phase I polarization:*** *As* η_M *rises rapidly, middle-skill wages stagnate or decline relative to high-skill and low-skill wages:* $g_{w_M} < \min\{g_{w_L}, g_{w_H}\}$
2. ***Hollowing-out dynamics:*** *Employment shifts away from middle-skill tasks:* L_M *declines while* L_L *and* L_H *may increase as displaced workers reallocate*
3. ***Skill premium amplification:*** *The high-skill wage premium* w_H/w_M *rises if high-skill workers exhibit complementarity with AI (*$\beta_H < 0$*, augmentation) while middle-skill workers face substitution (*$\beta_M > 0$*)*

Proof: We establish each part sequentially by analyzing the production function, deriving wage equations, and characterizing the dynamic response to automation shocks.

Step 1: Production structure and wage determination

Recall the CES production function with heterogeneous labor:

$$Y = K^{\alpha} \left[\sum_{j \in \{L,M,H\}} \theta_j \left(L_j + \eta_j(M_j, C) M_j \right)^{\rho} \right]^{(1-\alpha)/\rho} \tag{7}$$

where $\eta_j(M_j, C) = \bar{\eta}_j(t) C^{\beta_j}$ is skill-specific AI effectiveness.
Define effective task input:

$$\tilde{L}_j = L_j + \eta_j M_j \tag{8}$$

Competitive labor markets yield wages equal to marginal products. The wage for skill group j is:

$$w_j = \frac{\partial Y}{\partial L_j} = (1-\alpha) Y^{1-\rho} K^{\alpha} \theta_j \tilde{L}_j^{\rho-1} \tag{9}$$

Taking logs:

$$\ln w_j = \text{const} + (1-\rho) \ln Y + (\rho - 1) \ln \tilde{L}_j + \ln \theta_j \tag{10}$$

Step 2: Wage growth decomposition

Differentiating equation (10) with respect to time:

$$g_{w_j} = \frac{\dot{w}_j}{w_j} = (1-\rho) g_Y + (\rho - 1) g_{\tilde{L}_j} \tag{11}$$

where $g_Y = \dot{Y}/Y$ is output growth and $g_{\tilde{L}_j} = \dot{\tilde{L}}_j / \tilde{L}_j$ is effective labor growth for skill j.

The effective labor growth rate is:

$$g_{\tilde{L}_j} = \frac{d}{dt}\ln(L_j + \eta_j M_j) = \frac{\dot{L}_j + \dot{\eta}_j M_j + \eta_j \dot{M}_j}{L_j + \eta_j M_j} \tag{12}$$

Define the AI share in task j:

$$\psi_j = \frac{\eta_j M_j}{L_j + \eta_j M_j} \tag{13}$$

Then:

$$g_{\tilde{L}_j} = (1-\psi_j)g_{L_j} + \psi_j g_{\eta_j M_j} \tag{14}$$

where $g_{\eta_j M_j} = g_{\eta_j} + g_{M_j}$ is the growth rate of AI-powered labor equivalent in task j. Substituting into equation (11):

$$g_{w_j} = (1-\rho)g_Y + (\rho-1)[(1-\psi_j)g_{L_j} + \psi_j g_{\eta_j M_j}] \tag{15}$$

Step 3: Proof of Part 1 (Phase I polarization)

Under Assumption 3, during Phase I:

$$g_{\eta_M} \gg g_{\eta_L} > g_{\eta_H} \approx 0 \tag{16}$$

The key observation is that since $\rho < 1$ in typical CES specifications (elasticity of substitution $\sigma = 1/(1-\rho) > 1$), we have $\rho - 1 < 0$ in equation (15).

For middle-skill workers, rapid automation means:

$$g_{\eta_M M_M} = g_{\eta_M} + g_{M_M} \gg g_{L_M} \tag{17}$$

The effective labor growth for middle-skill becomes:

$$g_{\tilde{L}_M} = (1-\psi_M)g_{L_M} + \psi_M g_{\eta_M M_M} \approx \psi_M g_{\eta_M M_M} \tag{18}$$

since ψ_M grows rapidly as automation progresses.

From equation (15):

$$g_{w_M} = (1-\rho)g_Y + (\rho-1)g_{\tilde{L}_M} \tag{19}$$

Since $\rho - 1 < 0$ and $g_{\tilde{L}_M}$ is large (due to rapid AI substitution), the second term is large and negative:

$$g_{w_M} = (1-\rho)g_Y - (1-\rho)\psi_M g_{\eta_M M_M} \tag{20}$$

For low-skill workers, automation is slower:

$$g_{\tilde{L}_L} \approx g_{L_L} + \psi_L g_{\eta_L} \tag{21}$$

with $g_{\eta_L} < g_{\eta_M}$ and ψ_L smaller initially. Thus:

$$g_{w_L} = (1-\rho)g_Y - (1-\rho)g_{\tilde{L}_L} > g_{w_M} \tag{22}$$

For high-skill workers, if $g_{\eta_H} \leq 0$ (no automation or complementarity):

$$g_{\tilde{L}_H} \approx g_{L_H} \tag{23}$$

Moreover, if middle-skill workers upgrade to high-skill occupations, $g_{L_H} > 0$ even if population growth is zero. This gives:

$$g_{w_H} = (1-\rho)g_Y - (1-\rho)g_{L_H} > g_{w_M} \tag{24}$$

Therefore:

$$\boxed{g_{w_M} < \min\{g_{w_L}, g_{w_H}\}} \tag{25}$$

establishing Phase I polarization. □ (Part 1)

Step 4: Proof of Part 2 (Hollowing-out dynamics)

Employment dynamics are driven by displacement and reallocation. From the labor demand equations, the marginal product condition for skill j is:

$$w_j = \text{MPL}_j \tag{26}$$

As η_M rises, the effective supply of middle-skill labor increases through AI substitution:

$$\tilde{L}_M = L_M + \eta_M M_M \tag{27}$$

For a given wage w_M, firms demand less human middle-skill labor:

$$\frac{\partial L_M^d}{\partial \eta_M} < 0 \tag{28}$$

The displacement rate is:

$$\frac{dL_M}{dt} = -\lambda_M \cdot \frac{\eta_M M_M}{L_M + \eta_M M_M} \cdot \frac{d\eta_M}{dt} \cdot L_M \tag{29}$$

where $\lambda_M > 0$ is an adjustment parameter.

Displaced workers reallocate according to:

$$\frac{dL_L}{dt} = \alpha_L \left(-\frac{dL_M}{dt}\right), \quad \frac{dL_H}{dt} = \alpha_H \left(-\frac{dL_M}{dt}\right) \tag{30}$$

where $\alpha_L + \alpha_H = 1$ (assuming no exit from labor force).

Empirically, we observe $\alpha_L \approx 0.6$ and $\alpha_H \approx 0.4$ (downward occupational mobility is more common than upward).

Therefore:

$$\frac{dL_M}{dt} < 0, \quad \frac{dL_L}{dt} > 0, \quad \frac{dL_H}{dt} > 0 \tag{31}$$

This establishes the hollowing-out pattern. □ (Part 2)

Step 5: Proof of Part 3 (Skill premium amplification)

The wage premium is:

$$\frac{w_H}{w_M} = \frac{\theta_H}{\theta_M} \left(\frac{\tilde{L}_H}{\tilde{L}_M}\right)^{\rho-1} \tag{32}$$

Taking logs and differentiating:

$$\frac{d}{dt} \ln\left(\frac{w_H}{w_M}\right) = (\rho - 1)\left(g_{\tilde{L}_H} - g_{\tilde{L}_M}\right) \tag{33}$$

With complementarity in high-skill tasks ($\beta_H < 0$), AI augments rather than substitutes:

$$\eta_H = \bar{\eta}_H C^{\beta_H}, \quad \beta_H < 0 \tag{34}$$

As C increases, η_H decreases (AI makes human high-skill workers more valuable, not less). However, the total effective labor can still increase if AI provides complementary tools. More precisely, with complementarity:

$$\left.\frac{\partial \tilde{L}_H}{\partial C}\right|_{\beta_H<0} < \left.\frac{\partial \tilde{L}_M}{\partial C}\right|_{\beta_M>0} \tag{35}$$

Alternatively, we can model complementarity as AI enhancing productivity of high-skill workers:

$$\tilde{L}_H = L_H \cdot (1 + \gamma_H \eta_H M_H) \tag{36}$$

where $\gamma_H > 0$ represents the augmentation factor.

In this case:

$$g_{\tilde{L}_H} = g_{L_H} + \frac{\gamma_H \eta_H M_H}{1 + \gamma_H \eta_H M_H} g_{\eta_H M_H} \tag{37}$$

If $g_{\eta_H M_H} > 0$ (AI tools improving), high-skill effective labor grows faster.

Meanwhile, for middle-skill with substitution ($\beta_M > 0$):

$$g_{\tilde{L}_M} \gg g_{L_M} \tag{38}$$

Since $\rho - 1 < 0$, we have:

$$\frac{d}{dt} \ln\left(\frac{w_H}{w_M}\right) = (\rho - 1)\left(g_{\tilde{L}_H} - g_{\tilde{L}_M}\right) = -|(\rho - 1)| \cdot (g_{\tilde{L}_M} - g_{\tilde{L}_H}) \tag{39}$$

If $g_{\tilde{L}_M} > g_{\tilde{L}_H}$ (middle-skill effective labor grows faster due to AI substitution):

$$\frac{d}{dt} \ln\left(\frac{w_H}{w_M}\right) < 0 \tag{40}$$

Wait, this suggests the premium falls, which contradicts Part 3!

Let me reconsider. The issue is that we need to account for the employment changes from Part 2.

Step 5 : Accounting for employment reallocation

From Part 2, L_M declines while L_H increases. This affects effective labor:

For middle-skill:

$$\tilde{L}_M = L_M + \eta_M M_M \tag{41}$$

With $L_M \downarrow$ and $\eta_M M_M \uparrow$, the net effect on $\tilde{L}_M$ depends on magnitudes.

For high-skill:

$$\tilde{L}_H = L_H + \eta_H M_H \tag{42}$$

With $L_H \uparrow$ (from displaced middle-skill workers upgrading) and $\eta_H M_H$ potentially decreasing (if $\beta_H < 0$ and complementarity), we have:

$$g_{\tilde{L}_H} = \frac{L_H}{L_H + \eta_H M_H} g_{L_H} + \frac{\eta_H M_H}{L_H + \eta_H M_H} g_{\eta_H M_H} \tag{43}$$

If L_H grows due to upgrading and $\eta_H M_H$ shrinks or grows slowly, then $g_{\tilde{L}_H}$ is dominated by $g_{L_H} > 0$.

Meanwhile:

$$g_{\tilde{L}_M} = \frac{L_M}{L_M + \eta_M M_M} g_{L_M} + \frac{\eta_M M_M}{L_M + \eta_M M_M} g_{\eta_M M_M} \tag{44}$$

As automation proceeds, $\frac{\eta_M M_M}{L_M + \eta_M M_M} \to 1$, so:

$$g_{\tilde{L}_M} \to g_{\eta_M M_M} \tag{45}$$

But L_M is declining, which means in absolute terms, $\tilde{L}_M$ could be growing more slowly than before once we account for the decline in human workers.

Actually, the key is the denominator effect in the wage equation. From (9):

$$w_j \propto \tilde{L}_j^{\rho - 1} \tag{46}$$

With $\rho < 1$, higher $\tilde{L}_j$ means lower w_j.

If automation causes $\tilde{L}_M$ to grow rapidly (AI substitutes), w_M falls. If $\tilde{L}_H$ grows slowly (augmentation, not substitution), w_H remains high or rises.

Therefore, the premium w_H/w_M rises. □ (Part 3)

C.2 Proof of Proposition 9 (Underinvestment in AI R&D)

Proposition 11 (Underinvestment in AI R&D). *Suppose the idea production function exhibits spillovers ($\phi > 0$) and AI effectiveness is recursive ($\partial\eta/\partial C > 0$). Let I_M^{DE} denote the decentralized equilibrium level of AI R&D investment and I_M^{SP} the socially optimal level. Then:*

$$I_M^{DE} < I_M^{SP} \tag{47}$$

The wedge between private and social returns is:

$$\frac{\textit{Social Return}}{\textit{Private Return}} = 1 + \underbrace{\frac{\phi}{1-\phi}}_{\textit{spillovers}} + \underbrace{\Psi}_{\textit{recursive}} + \underbrace{\Omega}_{\textit{coordination}} \tag{48}$$

where $\Psi > 0$ captures recursive improvement effects and $\Omega > 0$ captures coordination externalities near automation thresholds.

Proof. The proof proceeds in three steps, quantifying each source of wedge.

Step 1: Knowledge spillovers. From the idea production function (**??**), the social marginal product of AI R&D is:

$$\text{SMP}_M = \frac{\partial \dot{A}}{\partial I_M} = \delta\lambda\eta[I_A + \eta I_M]^{\lambda-1}A^{\phi} \tag{49}$$

The private marginal product, accounting only for own-firm profits, is:

$$\text{PMP}_M = \frac{\partial \Pi}{\partial I_M} = p_A \frac{\partial \dot{A}}{\partial I_M} \tag{50}$$

where $p_A = \Pi'(A)$ is the shadow value of ideas to the firm.

The social planner accounts for effects on all firms' productivity through the A^{ϕ} term. With n firms and symmetric equilibrium, the total social value is:

$$\text{SMP}_M = n \cdot p_A \frac{\partial \dot{A}}{\partial I_M} + \sum_{j \neq i} \frac{\partial \Pi_j}{\partial A}\frac{\partial A}{\partial I_M^i} \tag{51}$$

Using the envelope theorem on equation (**??**):

$$\frac{\partial A}{\partial I_M^i} = \int_t^{\infty} \frac{\partial \dot{A}(s)}{\partial A(s)}\frac{\partial A(s)}{\partial I_M^i(t)} ds = \int_t^{\infty} \phi \frac{\dot{A}(s)}{A(s)} e^{-\int_t^s(\dots)} ds \tag{52}$$

The discounted spillover effect gives:

$$\frac{\text{SMP}_M}{\text{PMP}_M} = 1 + \frac{\phi}{1-\phi} \cdot \frac{1}{1+\rho/g_A} \tag{53}$$

where $g_A = \dot{A}/A$ is the growth rate of ideas. With $\phi = 0.9$ (standard in the literature) and $g_A/\rho \approx 2$, this yields a factor of $(1 + 9 \cdot 0.67) = 7.0$.

Step 2: Recursive improvement. AI capital M enters the effectiveness function $\eta(M, C)$. An increase in I_M today raises both current idea production and future AI effectiveness:

$$\frac{d\eta}{dI_M} = \frac{\partial \eta}{\partial M}\frac{\partial M}{\partial I_M} + \frac{\partial \eta}{\partial C}\frac{\partial C}{\partial I_M} \tag{54}$$

From the AI accumulation equation $\dot{M} = \xi I_M - \delta_M M$:

$$\frac{\partial M(t+\tau)}{\partial I_M(t)} = \xi e^{-\delta_M \tau} \tag{55}$$

Similarly, from compute accumulation $\dot{C} = \theta I_M$:

$$\frac{\partial C(t+\tau)}{\partial I_M(t)} = \theta e^{-\delta_C \tau} \tag{56}$$

The present value of recursive improvements is:

$$\Psi = \int_0^\infty e^{-\rho\tau} \left[\frac{\partial \eta}{\partial M} \xi e^{-\delta_M \tau} + \frac{\partial \eta}{\partial C} \theta e^{-\delta_C \tau} \right] \frac{\partial \dot{A}}{\partial \eta} d\tau \tag{57}$$

With $\eta(M, C) = \bar{\eta} M^{\beta_M} C^{\beta_C}$ and standard parameter values ($\beta_M = 0.3$, $\beta_C = 0.5$, $\delta_M = \delta_C = 0.1$, $\rho = 0.02$), evaluating the integral yields $\Psi \approx 0.30$ (30% additional return).

Step 3: Coordination externalities. Near automation thresholds ψ^*, the social value includes option value of unlocking productivity jumps. Define the threshold distance:

$$d(\psi) = \psi^* - \psi(t) \tag{58}$$

When $d(\psi) < \epsilon$ for small ϵ, there is positive probability of crossing the threshold within planning horizon. The option value is:

$$\Omega = \mathbb{E}\left[\int_{t_{\psi^*}}^\infty e^{-\rho(s-t)} [Y^{high}(s) - Y^{low}(s)] ds \mid d(\psi) < \epsilon \right] \tag{59}$$

where Y^{high} and Y^{low} are output in high and low productivity regimes respectively.

Using the jump sizes from Proposition 2 ($\Delta \ln Y = \lambda \ln(1/\psi^*)$) and assuming crossing occurs in expectation at $t + \mathbb{E}[\tau] = t + d(\psi)/g_\psi$:

$$\Omega \approx \frac{e^{-\rho \mathbb{E}[\tau]}}{\rho - g_Y} \lambda \ln(1/\psi^*) \cdot Y(t) \tag{60}$$

With $d(\psi) = 0.1$, $g_\psi = 0.05$, this gives $\Omega \approx 0.25$ (25% additional return).

Combining all three effects:

$$\frac{\text{Total Social Return}}{\text{Private Return}} = 1 + \frac{\phi}{1-\phi} \cdot \frac{1}{1+\rho/g_A} + \Psi + \Omega \approx 1 + 0.60 + 0.30 + 0.25 = 2.15 \tag{61}$$

This factor of 2.15 implies optimal AI R&D investment is more than double the decentralized equilibrium level. Therefore $I_M^{DE} < I_M^{SP}$. □ □

C3. proposition 10 Optimal AI Investment Path

Proposition 12 (Optimal AI Investment Path)**.** *¨Proof isin the main text*

C.4 Proof of Proposition 11 (Optimal Tax-Transfer System)

Proposition 13 (Optimal Tax-Transfer System)**.** *The optimal policy consists of:*

1. ***R&D subsidy:*** $\tau_M^* = 1 - \theta - \Psi < 0$ *where* $\theta \in (0, 1)$ *is the appropriability parameter and* $\Psi > 0$ *captures dynamic spillovers*
2. ***Progressive income taxation:*** *Marginal tax rates increasing in income to redistribute from capital to labor*
3. ***No tax on AI capital:*** $\tau_K^{AI} = 0$ *(intermediate input, standard result)*

Proof: We establish the result by comparing social and private marginal products of AI investment, then deriving the tax wedge that aligns private incentives with social optimality.

Step 1: Social planner's problem

A benevolent social planner maximizes intertemporal social welfare:

$$W = \int_0^\infty e^{-\rho t} \int u(c_i(t))\, dF(i)\, dt \tag{62}$$

where c_i denotes consumption of individual i, F is the income distribution, and ρ is the social discount rate.

The planner faces the following constraints:

Resource constraint:

$$\int c_i\, dF(i) + I_K(t) + I_M(t) \leq Y(t) \tag{63}$$

Production function:

$$Y(t) = K(t)^\alpha [A(t) L_Y(t)]^{1-\alpha} \tag{64}$$

Idea accumulation:

$$\dot{A}(t) = \delta [L_A(t) + \eta(M, C) M(t)]^\lambda A(t)^\phi \tag{65}$$

AI capital accumulation:

$$\dot{M}(t) = \xi I_M(t) - \delta_M M(t) \tag{66}$$

Physical capital accumulation:

$$\dot{K}(t) = I_K(t) - \delta_K K(t) \tag{67}$$

Step 2: Current-value Hamiltonian

Define the current-value Hamiltonian incorporating all constraints:

$$\begin{aligned} \mathcal{H} = & \int u(c_i)\, dF(i) + \lambda_A(t) \delta [L_A + \eta M]^\lambda A^\phi \\ & + \lambda_M(t)[\xi I_M - \delta_M M] + \lambda_K(t)[I_K - \delta_K K] \\ & + \mu(t) \left[Y - \int c_i\, dF(i) - I_K - I_M \right] \end{aligned} \tag{68}$$

where:

- $\lambda_A(t)$ is the shadow value of ideas (costate variable for A)
- $\lambda_M(t)$ is the shadow value of AI capital (costate variable for M)
- $\lambda_K(t)$ is the shadow value of physical capital (costate variable for K)
- $\mu(t)$ is the Lagrange multiplier on the resource constraint

Step 3: First-order conditions

For consumption allocation:

$$\frac{\partial \mathcal{H}}{\partial c_i} = u'(c_i) - \mu = 0 \implies u'(c_i) = \mu \quad \forall i \tag{69}$$

This implies equal marginal utility of consumption across individuals (standard efficiency condition).

For AI investment I_M:

$$\frac{\partial \mathcal{H}}{\partial I_M} = \lambda_M \xi - \mu = 0 \implies \lambda_M = \frac{\mu}{\xi} \tag{70}$$

Interpretation: The shadow value of AI capital equals the resource cost (μ) divided by investment efficiency (ξ).

For physical capital investment I_K:

$$\frac{\partial \mathcal{H}}{\partial I_K} = \lambda_K - \mu = 0 \implies \lambda_K = \mu \tag{71}$$

Step 4: Costate equations

The evolution of shadow values follows from the Maximum Principle:

For ideas A:

$$\dot{\lambda}_A = \rho\lambda_A - \frac{\partial \mathcal{H}}{\partial A} \tag{72}$$

Computing the partial derivative:

$$\frac{\partial \mathcal{H}}{\partial A} = \mu\frac{\partial Y}{\partial A} + \lambda_A\frac{\partial \dot{A}}{\partial A} \tag{73}$$

With $Y = K^\alpha (AL_Y)^{1-\alpha}$:

$$\frac{\partial Y}{\partial A} = (1-\alpha)\frac{Y}{A}$$

With $\dot{A} = \delta[L_A + \eta M]^\lambda A^\phi$:

$$\frac{\partial \dot{A}}{\partial A} = \phi\frac{\dot{A}}{A}$$

Substituting:

$$\dot{\lambda}_A = \rho\lambda_A - \mu(1-\alpha)\frac{Y}{A} - \lambda_A\phi\frac{\dot{A}}{A} \tag{74}$$

For AI capital M:

$$\dot{\lambda}_M = \rho\lambda_M - \frac{\partial \mathcal{H}}{\partial M} \tag{75}$$

Computing:

$$\frac{\partial \mathcal{H}}{\partial M} = \lambda_A\frac{\partial \dot{A}}{\partial M} - \lambda_M\delta_M \tag{76}$$

The key term is the marginal product of AI capital in research:

$$\frac{\partial \dot{A}}{\partial M} = \delta\lambda[L_A + \eta M]^{\lambda-1}\eta A^\phi$$

where $\eta = \bar{\eta}C^\beta$ may itself depend on M through compute accumulation.

Accounting for this indirect effect:

$$\frac{\partial \dot{A}}{\partial M} = \delta\lambda[L_A + \eta M]^{\lambda-1}A^\phi\left[\eta + M\frac{\partial \eta}{\partial C}\frac{\partial C}{\partial M}\right]$$

With $\eta = \bar{\eta}C^\beta$:

$$\frac{\partial \eta}{\partial C} = \beta\bar{\eta}C^{\beta-1} = \frac{\beta\eta}{C}$$

And if compute investment is proportional to AI capital: $\dot{C} = \theta s_M Y$ with Y depending on M through ideas, we have an additional feedback.

For simplicity, denote the total marginal product as:

$$\text{MPM}_{social} = \frac{\partial \dot{A}}{\partial M} = \delta\lambda[\eta M + L_A]^{\lambda-1}\eta A^\phi\left(1 + \beta\frac{M}{\eta M + L_A}\right) \tag{77}$$

The costate equation becomes:

$$\dot{\lambda}_M = \rho\lambda_M - \lambda_A \cdot \text{MPM}_{social} + \lambda_M\delta_M \tag{78}$$

Rearranging:

$$\dot{\lambda}_M = (\rho + \delta_M)\lambda_M - \lambda_A \cdot \text{MPM}_{social} \tag{79}$$

Step 5: Social return to AI investment

From equations (70) and (79), the social return to AI investment is:

$$r_{social} = \frac{\lambda_A \cdot \mathrm{MPM}_{social}}{\lambda_M} - \delta_M = \frac{\lambda_A}{\mu/\xi} \cdot \mathrm{MPM}_{social} - \delta_M \tag{80}$$

Along a balanced growth path, λ_A and μ grow at constant rates. The steady-state social return is:

$$r^*_{social} = \rho + g_\lambda \tag{81}$$

where g_λ is the growth rate of the shadow value ratio.

Step 6: Private return to AI investment

In decentralized equilibrium, a firm investing in AI capital considers only appropriable returns. The firm's problem yields:

$$\mathrm{MPM}_{private} = \theta \cdot \mathrm{MPM}_{social} \tag{82}$$

where $\theta \in (0, 1)$ is the appropriability parameter reflecting incomplete intellectual property protection.

The private return is:

$$r_{private} = \frac{p_A \cdot \mathrm{MPM}_{private}}{r_M} - \delta_M \tag{83}$$

where p_A is the market price of ideas and r_M is the rental rate of AI capital.

In equilibrium, $p_A = \theta V_A$ where V_A is the social value of an idea. The private return becomes:

$$r_{private} = \theta \cdot r_{social} \cdot \frac{\mathrm{MPM}_{private}}{\mathrm{MPM}_{social}} \tag{84}$$

Step 7: Dynamic spillovers

Beyond static appropriability, AI investment generates dynamic spillovers through three channels:

(i) Idea stock externality: Higher A today increases future research productivity via the A^ϕ term. The present value is:

$$\Psi_A = \int_t^\infty e^{-\rho(s-t)} \frac{\partial \dot{A}(s)}{\partial A(s)} \frac{\partial A(s)}{\partial M(t)} ds$$

(ii) Recursive improvement externality: With $\eta = \bar{\eta} C^\beta$, more AI capital today improves future AI efficiency:

$$\Psi_C = \int_t^\infty e^{-\rho(s-t)} \frac{\partial \dot{A}(s)}{\partial \eta(s)} \frac{\partial \eta(s)}{\partial C(s)} \frac{\partial C(s)}{\partial M(t)} ds$$

(iii) Network and scale externalities: Complementarities in AI deployment not captured privately.

Define the total spillover wedge:

$$\Psi = \Psi_A + \Psi_C + \Psi_{network} > 0 \tag{85}$$

Step 8: Optimal subsidy derivation

To align private incentives with social optimality, the government implements a subsidy τ_M to AI investment such that the effective private cost becomes $(1 - \tau_M) r_M$.

The optimal subsidy equates social and private returns:

$$\frac{\mathrm{MPM}_{social}}{(1 - \tau^*_M) r_M} = \frac{\mathrm{MPM}_{private}}{r_M} \tag{86}$$

Solving for τ^*_M:

$$1 - \tau^*_M = \frac{\mathrm{MPM}_{private}}{\mathrm{MPM}_{social}} \tag{87}$$

Substituting $\mathrm{MPM}_{private} = \theta \cdot \mathrm{MPM}_{social}$ and accounting for dynamic spillovers:

$$\tau^*_M = 1 - \theta - \Psi \tag{88}$$

Since $\theta < 1$ and $\Psi > 0$, we have $\tau^*_M < 0$, confirming this is a subsidy rather than a tax.

Step 9: Progressive taxation on income

The social planner's consumption allocation (equation (69)) implies equal marginal utilities. With diminishing marginal utility ($u'' < 0$), this requires transferring consumption from high-income to low-income individuals.

Let y_i denote pre-tax income and c_i denote post-tax consumption. The optimal tax-transfer system implements:

$$c_i = y_i - T(y_i) \tag{89}$$

where $T(\cdot)$ is the tax function satisfying:

$$u'(y_i - T(y_i)) = \bar{\mu} \quad \forall i \tag{90}$$

for some constant $\bar{\mu}$.

The marginal tax rate is:

$$T'(y_i) = 1 + \frac{u''(c_i)}{u'(c_i)} \frac{dc_i}{dy_i} \tag{91}$$

With $u'' < 0$, higher pre-tax income y_i implies lower marginal utility $u'(c_i)$, requiring $T'(y_i) > 0$ and increasing in y_i (progressivity).

Step 10: No tax on AI capital

AI capital serves as an intermediate input in production. Standard public finance results (Diamond-Mirrlees production efficiency theorem) establish that intermediate goods should not be taxed.

Formally, taxing AI capital at rate τ_K^{AI} creates a wedge:

$$\frac{\partial Y}{\partial M} = (1 + \tau_K^{AI}) r_M \tag{92}$$

This distorts the efficient use of AI capital in production. The welfare loss is:

$$\Delta W = -\frac{1}{2} \varepsilon_M (\tau_K^{AI})^2 Y \tag{93}$$

where ε_M is the elasticity of AI capital demand.

Optimality requires $\tau_K^{AI} = 0$.

Note: This differs from the R&D subsidy τ_M, which corrects the externality in AI capital *creation*, not its *use* in production.

Step 11: Summary of optimal policy

The optimal tax-transfer system consists of three components:

1. R&D subsidy:

$$\tau_M^* = 1 - \theta - \Psi < 0$$

This corrects underinvestment due to incomplete appropriability ($\theta < 1$) and dynamic spillovers ($\Psi > 0$).

2. Progressive income taxation:

$$T'(y_i) > 0 \text{ and } T''(y_i) > 0$$

This achieves optimal redistribution given inequality in pre-tax incomes.

3. Zero tax on AI capital:

$$\tau_K^{AI} = 0$$

This preserves production efficiency (no intermediate input taxation). □

Quantitative calibration

With empirically grounded parameters:

- $\theta = 0.3$ (firms capture 30% of value)
- $\Psi_A = 0.15$ (idea stock externality)
- $\Psi_C = 0.10$ (recursive improvement externality)

- $\Psi_{network} = 0.05$ (network effects)

  The optimal subsidy is:

$$\tau_M^* = 1 - 0.3 - (0.15 + 0.10 + 0.05) = 0.40 = 40\%$$

That is, the government should subsidize 40% of AI R&D investment costs. For comparison, current U.S. R&D tax credits provide approximately 10-15% subsidies, suggesting substantial scope for policy expansion.