Regulating AI: International Competition and Geopolitics*

Chen Lian **Andreas Schaab**

March 23, 2026

Abstract

We study optimal AI regulation in a two-country model with a winner-takes-all innovation race. An AI technology whose social costs are large enough that a closed economy would ban it can, under international competition, be not only permitted but subsidized beyond the privately optimal level. The reversal is driven by two channels. First, a domestic ecosystem channel: winning the race attracts downstream production home, raising the social but not the private return to winning. Second, a race premium: because being scooped permanently weakens the domestic ecosystem, the planner values avoiding the losing state, a cost the private innovator does not internalize. Neither channel exists in the closed economy. A global cooperative planner would still ban the technology, since both the ecosystem gains and the race premium are zero-sum from a cooperative perspective. We endogenize post-innovation AI export controls and show that the anticipation of geopolitical rivalry in the post-innovation phase reinforces the policy reversal, widening the range of social costs for which national regulators subsidize an AI that a global planner would ban.

*Lian: UC Berkeley and NBER; chen_lian@berkeley.edu. Schaab: UC Berkeley and NBER; schaab@berkeley.edu.
We thank João Guerreiro for helpful comments and suggestions.

1 Introduction

AI innovation generates large productivity gains but also imposes social costs, from labor displacement to security risks, that scale with capability. A government weighing these forces for its own economy faces a familiar trade-off: promote AI innovation when its social benefits exceed its costs, restrict it when they do not. If the social costs of a frontier AI technology are large enough, a closed economy should ban it.

But governments do not set AI policy in isolation. The White House AI Action Plan (The White House, 2025), entitled "Winning the Race," frames domestic AI regulation as a competitive handicap against China and calls for the systematic removal of regulatory barriers to maintain "unquestioned and unchallenged global technological dominance." China pursues indigenous AI development while showing ambivalence about importing U.S. models, partly to protect its nascent domestic industry. The European Union's AI Act, focused on safety, is already being amended over concerns about "falling behind." In all three cases, international competition appears to be the first-order driver of regulatory design, yet there is essentially no formal economic framework for understanding how it shapes optimal AI policy. Should a country subsidize an AI technology it would ban in the absence of foreign competition? When does trade openness turn a socially costly technology into one worth promoting? And should the winner of an AI race share frontier technology with the loser, or withhold it to preserve its competitive advantage?

We formalize these questions in a two-country model with a winner-takes-all AI innovation race. Our central result is a *policy reversal*: a technology that a closed-economy planner would ban can become one that an open-economy planner not only permits but subsidizes beyond the privately optimal level. The reversal is driven by a *domestic ecosystem channel* through which winning the AI race attracts downstream production home, raising the social return to AI innovation without raising the private return. A *race premium*, reflecting the fear of being scooped by a foreign rival, amplifies the effect. A global cooperative planner would still ban the technology: the national subsidy is entirely driven by beggar-thy-neighbor incentives that are zero-sum from a cooperative perspective. We further endogenize post-innovation AI export controls and show that the anticipation of geopolitical rivalry in the post-innovation phase reinforces the policy reversal.

Environment. We build a two-country model of international trade and AI innovation. Home and Foreign trade a continuum of differentiated varieties subject to iceberg trade costs, with free entry determining the mass of firms in each country. A single AI firm in each country produces an intermediate input used by differentiated-goods producers; the two AI firms compete through Bertrand pricing. Production in the differentiated sector features a spillover from AI that operates through two channels: a domestic ecosystem channel that scales with the local technology level, and a global frontier channel that scales with the world's best AI. AI firms engage in a winner-takes-all innovation race modeled as a Poisson game: each firm invests in R&D to increase its hazard rate

of reaching a superior technology, and the first to innovate wins. A reduced-form social cost, capturing labor displacement and other risks that scale with AI capability, rises with the frontier technology level. The government sets R&D subsidies.

The policy reversal. Consider first a closed economy. The planner weighs the benefits of AI innovation against its social cost. When social costs are large enough to outweigh these benefits, the planner bans AI innovation entirely. Our central result is a policy reversal: international competiton can reverse this ban into a subsidy. The reversal is driven by two channels that international competition creates.

The first is the domestic ecosystem channel. When Home wins the AI race, its differentiated-goods producers benefit from a stronger domestic AI spillover, lowering their production costs. This draws more firms into the Home economy, raising the number of Home varieties at the expense of Foreign. Because domestic varieties enter the consumer price index without trade costs, a larger Home ecosystem translates directly into a lower cost of living. This effect is stronger when trade is more open. Lower trade costs make the equilibrium ecosystem more responsive to technology differences, so winning triggers a larger reallocation of firms toward Home. Under autarky, firm locations are pinned regardless of who innovates, and the channel shuts down. The private AI firm does not internalize this ecosystem externality: its profits depend on the technology gap, not on trade costs.

The second is the race premium. The AI innovation game is winner-takes-all: the first country to reach the frontier wins permanently, and the loser cannot catch up. If the foreign country wins, the domestic ecosystem weakens as production shifts abroad and the cost of living rises, with no possibility of reversal. The planner therefore values not only the gain from winning but also the avoidance of this permanent losing state. When the foreign rival innovates aggressively, the urgency intensifies. This channel is entirely absent in the closed economy, where there is no rival.

Both channels raise the social return to AI innovation without raising the private return. Both components of the consumer-price gain are increasing in trade openness, so the social-cost threshold above which the planner wants less AI innovation than the market rises as trade becomes more open. The policy reversal occurs when social costs lie above the autarky ban threshold but below this open-economy threshold. We further provide a sufficient condition, independent of trade costs, under which this reversal region is non-empty for every level of trade openness.

Cooperative innovation policy. We next ask what a global cooperative planner would choose. A planner who internalizes both countries' welfare strips out the two forces that drive the national policy reversal. The ecosystem channel is near zero-sum: innovation reshuffles where firms locate rather than expanding total production, so the winner's ecosystem gain is largely offset by the loser's loss. The race premium has no bite: since the global planner is indifferent to which country wins, the losing state carries no welfare cost. The global planner also internalizes social costs

in both countries, since both access the frontier under AI trade. Competing national planners therefore always over-invest relative to the cooperative optimum. In the range of parameters where the closed-economy planner would ban innovation but the national planner subsidizes it, the cooperative planner also bans: the entire innovation incentive in this region is driven by the ecosystem externality and the race premium, forces that are purely redistributive from a global perspective.

AI export controls and the geoeconomic arms race. We also endogenize AI trade policy after the innovation race. Once a country wins, its frontier technology becomes a strategic asset. The winner can impose export controls to preserve its competitive advantage in the goods market; the loser can reject imports to avoid social costs and preserve a catch-up option through domestic R&D. The equilibrium depends on how each side values frontier access relative to its outside option, and three regimes can arise: loser-initiated autarky, AI trade with fee extraction by the winner, or a winner-initiated export ban.

The winner-initiated export ban creates the sharpest feedback to first-stage innovation incentives. Under the ban, the loser forfeits the global frontier spillover, widening the effective productivity gap and shifting more production toward the winner. Private and social returns to innovation then move in opposite directions. The winning AI firm now serves only the domestic market, halving its post-innovation profits, so the private return falls. But the amplified ecosystem gain delivers a larger consumer-price improvement that the firm does not capture, so the social return rises. The anticipation of post-innovation strategic competition therefore widens the wedge between optimal and laissez-faire policy, expanding the range of social costs for which the policy reversal holds. When AI trade occurs instead, with the winner extracting a fee for frontier access, the fiscal rent raises the social return to winning without affecting the private return, also reinforcing the reversal.

Related literature. Our paper relates first to the literature on the economics of automation and AI regulation. Acemoglu and Restrepo (2018) and Acemoglu and Restrepo (2019) develop task-based frameworks in which automation raises productivity while displacing labor, and Acemoglu (2024) argues more broadly that market incentives may not direct innovation toward socially efficient technologies. On the policy side, Guerreiro et al. (2023) study optimal AI regulation under uncertainty about risks, showing that a two-stage regulatory approval process—sandbox testing followed by conditional deployment—dominates Pigouvian taxation when developer beliefs about harms are heterogeneous and private information. Gans (2025) studies how a planner should steer the direction of AI innovation when technologies may cause harm, showing that it can be optimal to invest more in a risky path for its option value. Guerreiro et al. (2022) and Thuemmel (2023) study the optimal taxation of robots, characterizing conditions under which automation should be taxed rather than encouraged, and Costinot and Werning (2023) analyze optimal technology

regulation in a second-best environment with applications to robots and trade. These papers mostly ask how a single planner regulates AI or automation without emphasizing strategic rivalry across countries. We ask a different question: how does optimal regulation change when governments compete? Our main result is that international competition can flip the sign of optimal policy: a technology that would be banned under autarky may instead be subsidized when governments compete.

Second, our paper connects to the literature on trade, innovation, and strategic industrial policy. Classic strategic-trade papers such as Brander and Spencer (1985) and Eaton and Grossman (1986) show how governments may intervene under imperfect competition to shift profits or improve national welfare. The domestic ecosystem channel in our model is closely related to the home market effect in Krugman (1980) and to the agglomeration forces in Krugman (1991): in both settings, a cost advantage attracts more firms, which lowers the local price index because domestic varieties avoid trade costs. In our model, the cost advantage is triggered by a discrete AI innovation rather than by demand-side circular causation, and the resulting firm reallocation is purely redistributive since free entry pins the total mass of varieties. Akcigit et al. (2018) study innovation policy when countries compete technologically in open economies, and Goldfarb and Trefler (2018) analyze how scale economies and data fragmentation shape international competition in AI. Bai et al. (2025) analyze optimal dynamic policy under technological rivalry in an open economy. Relative to these papers, we highlight a distinct mechanism: the key wedge in our model comes from the fact that winning the AI race strengthens the domestic production ecosystem, which feeds into the cost of living, while the private innovator does not internalize that ecosystem externality. This gives rise to an open-economy policy reversal that is absent under autarky.

Finally, our paper contributes to the emerging literature on geoeconomics and export controls. Clayton et al. (2026) define geoeconomics as the use of economic relationships for geopolitical and economic ends, and Clayton et al. (2025) model how hegemonic powers leverage control of chokepoint sectors to coerce trading partners, generating endogenous fragmentation. This logic maps directly into our model, where the winner of the AI race can leverage its frontier technology to impose export controls that widen the competitive gap against the loser. Mattoo et al. (2024) analyze how geopolitical rivalry reshapes trade cooperation and raises unilateral policy distortions, and Becko and O'Connor (2025) show that the anticipation of future geopolitical conflict distorts current trade relationships. Liu et al. (2024) characterize export restrictions as a tool of trade policy, and Liu et al. (2025) show that export controls can stimulate innovation in sanctioned economies. Chatterji and Murray (2025) argue that the economics of innovation has shifted from a paradigm of diffusion toward one of control over critical technologies, and Athey and Scott Morton (2025) study the open-economy welfare effects of upstream AI market power. Our contribution is to endogenize AI export controls inside an innovation-race model and show how anticipated restrictions feed back into ex ante innovation policy: the national case for subsidies grows stronger precisely as the global case for a ban remains.

2 Model

We study a two-country model of international trade and AI innovation. Two countries, Home and Foreign, trade a freely available numeraire good and a continuum of differentiated varieties subject to iceberg trade costs. A single AI firm in each country produces an intermediate input used by differentiated-goods producers and competes in a global market. The AI firms engage in a winner-takes-all innovation race that endogenously determines technology levels. The key economic forces are a productivity spillover from AI to differentiated production—operating through both a domestic ecosystem channel and a global frontier channel—a labor social cost that rises with AI capability, and a domestic ecosystem channel through which trade openness amplifies the social return to innovation.

Time is continuous and indexed by $t \in [0, \infty)$. We focus our exposition on Home and use asterisks for Foreign variables.

2.1 Households

Preferences. There is a representative household in each country. The Home household has preferences

$$\int_0^\infty e^{-\rho t} u_t \, dt \qquad \text{with} \qquad u_t = a_t + \beta \log c_t, \tag{1}$$

where $\rho > 0$ is the discount rate, a_t is consumption of the numeraire good, and c_t is a CES composite of differentiated varieties,

$$c_t = \left(\int_0^{N_t} c_{H,t}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega + \int_0^{N_t^*} c_{F,t}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}. \tag{2}$$

Here $c_{H,t}(\omega)$ denotes Home consumption of Home-produced variety $\omega \in [0, N_t]$, $c_{F,t}(\omega)$ Home consumption of Foreign-produced variety $\omega \in [0, N_t^*]$, and $\sigma > 1$ is the elasticity of substitution across varieties. We denote by N_t and N_t^* the mass of differentiated-variety firms operative in Home and Foreign at date t. These masses capture the size of each country's production ecosystem: a higher N_t means more firms choose to locate in Home, reflecting a more attractive domestic production base. As we show in Section 3, shifts in N_t and N_t^* in response to AI innovation outcomes are the key channel through which trade openness amplifies the social return to winning the AI race. The quasi-linear, log specification in (1) pins aggregate expenditure on differentiated goods at β and eliminate wealth effects, allowing the analysis to isolate how AI productivity spillovers and trade openness shape innovation incentives.[1]

[1] With quasi-linear preferences, intertemporal allocation is straightforward: the numeraire good serves as a linear savings technology, so there is no need for financial markets or intertemporal budget constraints beyond the period budget constraint stated below.

Budget constraint and trade costs. Normalize the numeraire price to 1. Let $\tilde{p}_{j,t}(\omega)$ denote the *delivered* price in Home of variety ω produced in country j. The Home household's period budget constraint is

$$a_t + \int_0^{N_t} \tilde{p}_{H,t}(\omega)\, c_{H,t}(\omega)\, d\omega + \int_0^{N_t^*} \tilde{p}_{F,t}(\omega)\, c_{F,t}(\omega)\, d\omega = Y_t, \tag{3}$$

where income $Y_t = w_t L + \Pi_t + \Pi_t^{\text{AI}} + T_t$ comprises wage income from a fixed labor endowment L, aggregate profits from differentiated-good firms Π_t, profits from the AI firm Π_t^{AI}, and a lump-sum transfer T_t from the government.

We denote the producer price of variety ω in Home by $p_t(\omega)$ and in Foreign by $p_t^*(\omega)$. The numeraire good is freely tradable. Trading differentiated goods, however, is subject to a symmetric iceberg trade cost $\tau > 1$: to deliver one unit of a variety from Foreign to Home, τ units must be shipped. The delivered price of a domestic variety therefore coincides with its producer price, while imports carry the iceberg markup,

$$\tilde{p}_{F,t}(\omega) = \tau\, p_t^*(\omega) \qquad \text{and} \qquad \tilde{p}_{H,t}^*(\omega) = \tau\, p_t(\omega). \tag{4}$$

2.2 Production

There are two kinds of final goods: the numeraire good and a continuum of differentiated goods.

Numeraire good. The numeraire good is produced by a perfectly competitive firm using a linear technology, $y_t^a = \ell_t^a$, where ℓ_t^a denotes labor allocated to agriculture. Since the numeraire price is normalized to 1, profit maximization implies $w_t = 1$ whenever agriculture is active. We think of agriculture as the sector of the economy unaffected by AI.

Differentiated goods. Each country hosts a continuum of monopolistically competitive firms. We denote by N_t and N_t^* the mass of such firms operative in Home and Foreign; these masses measure the scale of each country's domestic production ecosystem. Each firm produces a single differentiated variety using labor and AI,

$$y_t(\omega) = \left(A_t^{\gamma\xi} \tilde{A}_t^{\gamma(1-\xi)} x_t(\omega)\right)^{\alpha} \ell_t(\omega)^{1-\alpha}, \tag{5}$$

where $\ell_t(\omega)$ is firm ω's use of domestic labor, $x_t(\omega)$ is its use of AI, $\alpha \in (0,1)$ is AI's cost share, and $\tilde{A}_t \equiv \max\{A_t, A_t^*\}$ denotes the global AI frontier. The parameter $\gamma > 0$ governs the overall strength of the AI productivity spillover, and $\xi \in [0,1]$ governs the split between two channels. The domestic technology level A_t generates a local ecosystem spillover with elasticity $\gamma\xi$: as domestic AI capability rises, firms that use AI in production become more productive. The global frontier $\tilde{A}_t$ generates a knowledge spillover with elasticity $\gamma(1-\xi)$, capturing the idea that all AI-using firms benefit from advances at the world technology frontier regardless of where that innovation

originates. When $\xi = 1$ the spillover is purely domestic; when $\xi = 0$ the spillover is purely global. Neither spillover is internalized by AI firms.

To operate, firm ω must pay a fixed cost f in units of labor. Per-period profit net of this operating cost is

$$\Pi_t(\omega) = p_t(\omega)\, y_t(\omega) - w_t\, \ell_t(\omega) - \tilde{q}_t\, x_t(\omega) - w_t f, \tag{6}$$

where $\tilde{q}_t$ is the delivered price of AI in the Home market. Each firm chooses its price $p_t(\omega)$ and inputs $\{\ell_t(\omega), x_t(\omega)\}$ to maximize profits, taking input prices as given.

Free entry. There is a large mass of potential entrepreneurs in each country who can enter and produce a new variety. Free entry implies that firms enter until profits net of the fixed operating cost are zero.[2]

2.3 Artificial Intelligence

Each country is home to a single large AI firm. These firms face static production and pricing decisions as well as a dynamic innovation decision. We first describe production and pricing, then introduce the innovation process.

Production. The Home AI firm produces AI using labor,

$$y_t^{\text{AI}} = A_t\, \ell_t^{\text{AI}}, \tag{7}$$

with marginal cost $1/A_t$ given $w_t = 1$. Technology levels are determined endogenously through the innovation race described below; when choosing its production scale at date t, the firm takes A_t as given.

Price competition. The Home and Foreign AI firms compete à la Bertrand in a global market for AI; AI inputs produced in the two countries are perfect substitutes. Under Bertrand competition, each firm quotes a price—q_t for Home, q_t^* for Foreign—and the firm quoting the lower price captures the entire market. The prevailing world price is $\tilde{q}_t = \min\{q_t, q_t^*\}$.

We show in Appendix A.2 that aggregate world demand for AI is unit elastic in price. The unconstrained monopoly price is therefore unbounded, and Bertrand competition yields limit pricing: the leader prices at the follower's marginal cost. When Home is the AI leader with $A_t > A_t^*$,

$$\tilde{q}_t = \frac{1}{A_t^*}. \tag{8}$$

[2] We show in Appendix A.1 that monopolistic pricing yields a constant markup $\mu = \sigma/(\sigma - 1)$ over marginal cost, and free entry pins per-firm revenue at $R_t = \sigma f$.

Symmetrically, when Foreign leads, $\tilde{q}_t = 1/A_t$. When both firms have the same technology level, they share the market at the common marginal cost and earn zero variable profit from AI production.

Innovation. Both countries start at date 0 with the same technology level, $A_0 = A_0^* = A$. The AI firms engage in a winner-takes-all innovation race modeled as a one-shot Poisson game. Each firm invests in R&D to increase its hazard rate of reaching a superior technology $\bar{A} > A$. The first firm to innovate wins: the winner's technology jumps to $\bar{A}$ and the loser remains at A. After the race concludes, R&D ceases. The winner-takes-all structure reflects a distinctive feature of AI as a technology: once one firm opens a sufficiently large capability lead, the gap tends to be self-reinforcing. A frontier AI system can be used to accelerate its own development through recursive self-improvement, generating rapid capability gains that a lagging rival cannot easily replicate from a lower technological base.[3] In our two-phase structure, the jump from A to $\bar{A}$ with no subsequent R&D is a reduced-form representation of this mechanism. Rather than modeling the self-improvement process explicitly, we capture its consequence, a persistent technology gap, as a permanent state. This keeps the model tractable while preserving the economically essential feature that the identity of the winner has lasting consequences for the international distribution of AI capability.

R&D requires labor. We parameterize the innovation technology so that an R&D intensity e_t generates a Poisson arrival rate e_t of reaching the frontier and costs

$$\ell_t^{\text{R\&D}} = \frac{1}{2} e_t^2 \tag{9}$$

units of labor. The government may subsidize R&D at rate $s_t \in [0, 1)$, so the firm's private cost of R&D labor is $(1 - s_t)\ell_t^{\text{R\&D}}$.

AI firm profits. The Home AI firm's flow profit is

$$\Pi_t^{\text{AI}} = q_t y_t^{\text{AI}} - \ell_t^{\text{AI}} - (1 - s_t)\ell_t^{\text{R\&D}}, \tag{10}$$

comprising revenue from AI sales, less production labor, less R&D expenditure net of the subsidy. The AI firm chooses price quotes q_t and R&D intensity e_t to maximize the expected net present value of profits, discounting at rate ρ and taking the Foreign firm's R&D intensity e_t^* as given.

[3] The possibility of recursive self-improvement has been widely discussed as a distinguishing feature of AI relative to conventional technologies; see, e.g., Aghion et al. (2018) and Nordhaus (2021). The empirical pattern of scaling laws (Kaplan et al., 2020) and the use of AI to automate AI research reinforce this channel.

2.4 Government

Social cost of AI. Advancing AI capability imposes a reduced-form social cost $r_t = r(A_t^{\text{acc}})$ in units of labor, where A_t^{acc} is the highest AI technology level accessible to the domestic economy and $r(\cdot)$ is increasing. The leading component is labor displacement: AI renders some tasks obsolete, requiring costly worker retraining. The cost also captures broader risks that scale with capability, such as AI-enabled cyberattacks, bioweapon development by malicious actors, and other misuse externalities whose expected damages grow with the frontier technology level. Under the AI trade maintained throughout Sections 2–3, both countries access the global frontier, so $A_t^{\text{acc}} = \max\{A_t, A_t^*\}$.

Fiscal policy. The government sets the R&D subsidy s_t and finances both the subsidy and social costs through lump-sum taxation. The balanced-budget constraint is

$$0 = T_t + s_t \ell_t^{\text{R\&D}} + r_t, \tag{11}$$

and symmetrically for Foreign. After the innovation race concludes, $\ell_t^{\text{R\&D}} = 0$.

2.5 Market Clearing and Equilibrium

Market clearing. The world market for the numeraire good clears when $a_t + a_t^* = \ell_t^a + \ell_t^{*,a}$. Differentiated varieties clear in the world market accounting for iceberg trade costs,

$$y_t(\omega) = c_{H,t}(\omega) + \tau\, c_{H,t}^*(\omega) \qquad \text{and} \qquad y_t^*(\omega) = \tau\, c_{F,t}(\omega) + c_{F,t}^*(\omega). \tag{12}$$

Local labor markets clear in each country,

$$L = \ell_t^a + \int_0^{N_t} \Big[\ell_t(\omega) + f\Big] d\omega + \ell_t^{\text{AI}} + \ell_t^{\text{R\&D}} + r_t, \tag{13}$$

where the right-hand side accounts for agriculture, differentiated-goods production (variable and fixed), AI production, R&D, and the social cost of AI. The world market for AI clears with the leading firm supplying all demand,

$$y_t^{\text{AI}} = \left(\int_0^{N_t} x_t(\omega)\, d\omega + \int_0^{N_t^*} x_t^*(\omega)\, d\omega \right) \mathbf{1}_{A_t > A_t^*}, \tag{14}$$

and symmetrically for Foreign. When both firms have the same technology, they split the market.

Definition 1 (Equilibrium)**.** *Given initial technology $A_0 = A_0^* = A$, R&D subsidies $\{s_t, s_t^*\}$, and structural parameters, a world equilibrium consists of prices $\{w_t, \tilde{p}_{j,t}(\omega), q_t\}$, consumption allocations $\{a_t, c_{j,t}(\omega)\}$, firm allocations $\{N_t, \ell_t(\omega), x_t(\omega)\}$, AI allocations $\{y_t^{AI}, \ell_t^{AI}, e_t\}$, transfers $\{T_t\}$, technology paths $\{A_t\}$, and Foreign counterparts, such that:*

(i) *Households choose consumption to maximize utility subject to budget constraints, taking prices as given.*

(ii) *Conditional on operating, differentiated-goods firms choose prices and inputs to maximize profits, taking input prices as given.*

(iii) *Free entry is satisfied.*

(iv) *AI firms set prices competitively (Bertrand) and choose R&D intensity to maximize the expected net present value of profits, taking the rival's strategy as given (Nash).*

(v) *Government budget constraints are balanced.*

(vi) *All markets clear.*

Maintained assumptions. We assume throughout that allocations are interior: numeraire consumption is positive in both countries, both countries host strictly positive masses of differentiated-goods firms ($N_t > 0$ and $N_t^* > 0$), and R&D intensities are interior. Intuitively, trade costs must be large enough relative to any technology gap between the two countries that neither country is fully driven out of differentiated-goods production. A sufficient condition on primitives is given in Appendix A.1.

2.6 Innovation Game

We now describe the dynamic innovation competition between the two AI firms. In the pre-innovation phase, both firms have technology A and Bertrand competition implies zero AI profits. The innovation race determines which firm reaches $\bar{A}$ first.

Private innovation incentives. Under the simplified structure—symmetric initial technology and the losing firm exiting AI production—the Home AI firm's problem admits the recursive representation

$$\rho \mathcal{V} = \max_{e \geq 0} \left\{ -(1-s)\frac{1}{2}e^2 + e\big(\mathcal{V}^w - \mathcal{V}\big) - e^* \mathcal{V} \right\}, \tag{15}$$

where $\mathcal{V}$ is the firm's continuation value in the pre-innovation state and $\mathcal{V}^w = \frac{1}{\rho}\pi^w$ is its continuation value from winning. Here π^w denotes the flow AI profit in the post-innovation state, which equals the margin between the world AI price $\tilde{q} = 1/A$ and the winner's marginal cost $1/\bar{A}$, times total AI demand.[4] The three terms in (15) represent the flow cost of R&D net of the subsidy, the expected capital gain from winning at rate e, and the expected loss from the rival innovating first at

[4] We show in Appendix A.3 that $\pi^w = 2\frac{\alpha}{\mu}\beta\big(1 - \frac{A}{\bar{A}}\big)$, where $\mu = \sigma/(\sigma-1)$ is the markup. AI profits are independent of the trade cost τ because world spending on differentiated goods is constant at 2β and AI's cost share is constant under Cobb–Douglas.

rate e^*—in which case the Home firm's continuation value drops to zero. The first-order condition is

$$e^{\text{priv}}(e^*) = \frac{1}{1-s}\big(\mathcal{V}^w - \mathcal{V}\big), \tag{16}$$

so the firm's R&D intensity is proportional to the value gap between winning and the current state, scaled by the effective R&D cost $1 - s$.

Social innovation incentives. The Home planner chooses R&D intensity to maximize household welfare, taking Foreign R&D intensity e^* as given (Nash). The planner's problem admits the recursive representation

$$\rho\mathcal{W} = \max_{e \geq 0} \left\{ \tilde{u} - \frac{1}{2}e^2 + e\big(\mathcal{W}^w - \mathcal{W}\big) + e^*\big(\mathcal{W}^\ell - \mathcal{W}\big) \right\}, \tag{17}$$

where $\tilde{u}$ is Home household flow utility gross of R&D expenditure, $\mathcal{W}^w$ is the social continuation value from Home winning the race, and $\mathcal{W}^\ell$ is the continuation value from Foreign winning. In the post-innovation phase, $\mathcal{W}^w = u^w/\rho$ and $\mathcal{W}^\ell = u^\ell/\rho$, where u^w and u^ℓ denote flow utility in the winning and losing states.[5] The first-order condition is

$$e^{\text{soc}}(e^*) = \mathcal{W}^w - \mathcal{W}. \tag{18}$$

The planner internalizes the full social cost of R&D and does not benefit from a subsidy, so the FOC equates intensity directly to the social value gap. Crucially, the social continuation values $\mathcal{W}^w$ and $\mathcal{W}^\ell$ reflect not only AI profits but also the domestic ecosystem effect on the cost of living (through the CES price index P_t) and the social cost r_t. This is the fundamental source of divergence between private and social innovation incentives. Since both countries access the frontier under AI trade, the social cost $r(\bar{A})$ is the same regardless of which firm wins; we define the incremental social cost $K \equiv r(\bar{A}) - r(A)$. We also define the scooping weight $\theta = e^*/\rho$, which captures the strategic pressure from Foreign innovation and will govern the planner's relative valuation of winning versus losing the race.

Discussion. Both the private and social best-response functions have the same functional form: R&D intensity equals the gap between winning and current continuation values, modulated by the cost of R&D. The two problems differ in what the continuation values reflect. The private firm values winning through AI market profits alone. The planner values winning through the combined effect on household welfare—profits, consumer prices, and social costs. Whether the social return exceeds the private return therefore depends on the balance between the ecosystem benefit of winning, which lowers the cost of living (and which the firm does not capture), and

[5] In the winning state, Home earns AI profits and benefits from a stronger domestic ecosystem that lowers the cost of living. In the losing state, Home earns no AI profits and faces a weaker ecosystem as firms shift toward Foreign, raising the cost of living. Both states incur the same social cost $r(\bar{A})$ since both countries access the frontier under AI trade.

the social cost (which the firm does not bear). Section 3 characterizes the conditions under which international competition tilts this balance toward subsidizing innovation, even when autarky would call for restricting it.

2.7 Closed-Economy Benchmark

To isolate the role of international competition, we define a single-country closed-economy benchmark. The closed economy is identical to one country of the model above with three modifications: (i) there is no trade in differentiated goods, (ii) the domestic AI market is disciplined by a competitive open-source fringe that supplies AI at a constant price $q^{\text{OS}} = 1/A$ both before and after innovation, and (iii) innovation is not a winner-takes-all race—the domestic firm innovates independently. The open-source fringe pins the AI price at $1/A$ across both regimes, ensuring that the comparison between the closed and open economy isolates the trade and strategic-competition channels rather than mechanically switching the AI market structure between competitive and monopoly pricing.

In this benchmark, $\tilde{A}_t = A_t$, so the production spillover collapses to $A_t^{\gamma\xi} A_t^{\gamma(1-\xi)} = A_t^{\gamma}$ regardless of ξ. The goods block therefore reduces to a closed-economy monopolistic competition model with free entry and a mass of varieties $N = \beta/(\sigma f)$.[6] The social planner's innovation problem retains the structure of (17) but with no rival hazard ($e^* = 0$) and continuation values that reflect only the direct marginal-cost gain, not the domestic ecosystem channel. The closed economy thus provides a clean baseline: if the social cost K exceeds the direct consumer-price benefit $\beta\alpha\gamma \log(\bar{A}/A)$, the closed-economy planner restricts AI innovation.

3 AI Innovation Policy

We now characterize optimal innovation policy in the open and closed economies. Three forces shape the comparison between private and social innovation incentives: a productivity spillover from AI to differentiated production that the AI firm does not internalize, a domestic ecosystem channel through which trade openness amplifies the social return to innovation, and a social cost that rises with AI capability. Our main result is that international competition can reverse the direction of optimal policy: for an intermediate range of social costs, the closed-economy planner bans innovation while the open-economy planner subsidizes it. We first establish the key properties of the goods equilibrium that drive this result, then characterize the divergence between private and social innovation incentives, and finally state the main policy reversal result and its sufficient conditions.

[6] The closed-economy CPI satisfies $P = N^{1/(1-\sigma)} p$, where p is the common producer price. Innovation lowers p through the combined AI productivity spillover A^{γ} but does not trigger any ecosystem reallocation, since all varieties are domestic.

3.1 Equilibrium Properties

The goods equilibrium derived from the model of Section 2 admits closed-form expressions for firm-level allocations, the CES price index, and AI profits. Because the global frontier spillover $\tilde{A}_t^{\gamma(1-\xi)}$ benefits both countries equally, relative prices depend only on the domestic spillover component. We summarize this through the technology gap $z_t = (A_t/A_t^*)^\vartheta$ with $\vartheta = \alpha\gamma\xi(\sigma-1)$ and the trade openness index $\tau^{1-\sigma} \in (0,1)$.[7] We collect two properties that are essential for understanding the innovation results below.

AI profits are independent of trade costs. Under Bertrand competition with unit-elastic AI demand, the leader prices at the follower's marginal cost and captures the entire world AI market.[8] When Home wins the innovation race, flow AI profits are

$$\pi^w = 2\frac{\alpha}{\mu}\beta\left(1-\frac{A}{\bar{A}}\right),$$

which depends on world differentiated-good expenditure 2β, AI's revenue share α/μ, and the technology gap $\bar{A}/A$, but not on the trade cost τ.[9] This is a key feature of the model: trade openness affects the *social* return to innovation through the consumer price index but does not affect the *private* return through AI profits.

The domestic ecosystem and the cost of living. Innovation outcomes reshape each country's production ecosystem. When Home wins the race, its firms benefit from a larger domestic technology spillover, lowering their marginal costs. The resulting cost advantage widens the technology gap z, making Home varieties cheaper relative to Foreign varieties. Under free entry, revenue per firm is pinned at $R = \sigma f$, fixing the world mass of varieties at $N + N^* = 2\beta/(\sigma f)$. The distribution across countries, however, shifts with z:

$$N = \frac{\beta}{\sigma f}\frac{z\left(1+\tau^{2(1-\sigma)}\right) - 2\tau^{1-\sigma}}{\left(1-\tau^{1-\sigma}z\right)\left(z-\tau^{1-\sigma}\right)}.$$

As the gap widens, N increases at the expense of N^*: the domestic AI technology spillover lowers Home firms' marginal costs, drawing more entrants into the Home economy.[10]

[7] We derive the full equilibrium characterization in Appendix A.1. Under free entry, revenue per firm is $R_t = \sigma f$, world expenditure on differentiated goods is 2β, and the total mass of varieties satisfies $N_t + N_t^* = 2\beta/(\sigma f)$. All closed-form expressions used in this section are derived there.

[8] World AI demand is $N_t x_t + N_t^* x_t^* = (\alpha/\mu)(2\beta/\tilde{q}_t)$, where $\mu = \sigma/(\sigma-1)$. The log-elasticity of demand with respect to $\tilde{q}_t$ is -1, so the unconstrained monopoly price is unbounded and Bertrand competition yields limit pricing. See Appendix A.2.

[9] The τ-independence follows from Cobb-Douglas production in the differentiated sector: AI's cost share α is constant, and world spending on differentiated goods is pinned at 2β by quasi-linear preferences. See Appendix A.3.

[10] We derive this expression in Appendix A.1 (Lemma 11). In autarky ($\tau \to \infty$), each country has $N = \beta/(\sigma f)$ regardless of technology; trade openness is what allows the winner's ecosystem to expand.

Because domestic varieties enter the Home CPI without the iceberg penalty while imports enter with the factor $\tau^{1-\sigma}$, a larger Home ecosystem lowers the cost of living. Conversely, when Foreign wins, firms shift toward Foreign and the Home CPI rises. Specifically, we show in Appendix B.1 that, under the interiority condition maintained in Section 2,

$$P^w < P \qquad \text{and} \qquad P^w < P^\ell,$$

where P, P^w, and P^ℓ denote the Home CES price index in the pre-innovation state, the Home-wins state, and the Foreign-wins state, respectively. The winner's CPI falls unconditionally. Whether the loser's CPI rises above the pre-innovation level ($P^\ell > P$) depends on ξ: the full ranking $P^\ell > P > P^w$ holds when the domestic spillover share ξ is sufficiently close to 1, and holds unconditionally when $\xi = 1$.[11] This ranking captures the central mechanism of the paper: winning the innovation race is socially valuable not only because it generates AI profits, but also because it strengthens the domestic ecosystem, attracting firms to the Home economy and lowering the cost of living. Losing is costly relative to winning for the symmetric reason. The private AI firm captures the profit channel but not the ecosystem channel. The divergence between private and social innovation incentives therefore depends on whether the CPI gains outweigh the social costs—and how trade openness affects this balance.

3.2 Innovation Incentives

We now compare the private and social best-response functions for R&D intensity. The private AI firm chooses R&D intensity to maximize the expected net present value of profits, taking the Foreign firm's intensity e^* as given. The Home planner chooses R&D intensity to maximize household welfare, also taking Foreign R&D as given (Nash). Both problems are described in Section 2.6.

Welfare decomposition. The comparison between social and private innovation incentives turns on the difference in effective prizes—the gain from winning net of the gain from continuing in the pre-innovation state. The private effective prize reflects only AI profits. The social effective prize additionally reflects the ecosystem channel and social costs.

To see this, consider the welfare gap between the winning state and the pre-innovation state. Using the indirect utility derived in Appendix A.1, flow utility in each state decomposes into AI profits, social costs, and a log-CPI term. The social gain from winning is therefore

$$u^w - \tilde{u} = \pi^w - \pi - K - \beta \log \frac{P^w}{P}, \tag{19}$$

[11] The CPI ratios across states admit closed forms. Let $\bar{z} = (\bar{A}/A)^{\vartheta}$. Then $(P^w/P)^{1-\sigma} = (\bar{A}/A)^{\alpha\gamma(1-\xi)(\sigma-1)}\bar{z}(1-\tau^{1-\sigma})/(1-\tau^{1-\sigma}\bar{z})$ and $(P^\ell/P)^{1-\sigma} = (\bar{A}/A)^{\alpha\gamma(1-\xi)(\sigma-1)}(1-\tau^{1-\sigma})/(1-\tau^{1-\sigma}/\bar{z})$. The factor $(\bar{A}/A)^{\alpha\gamma(1-\xi)(\sigma-1)} \geq 1$ reflects the global frontier spillover, which lowers both the winning and losing CPIs relative to the pre-innovation state. In the win-vs-lose ratio this factor cancels: $(P^w/P^\ell)^{1-\sigma} = (\bar{z}-\tau^{1-\sigma})/(1-\tau^{1-\sigma}\bar{z})$.

where $\tilde{u}$ denotes flow utility gross of R&D expenditure in the pre-innovation state, $\pi^w - \pi = \pi^w$ is the profit gain (since pre-innovation profits are zero under symmetric technology), $K = r(\bar{A}) - r(A)$ is the incremental social cost, and $-\beta \log(P^w/P) > 0$ is the CPI gain from winning.[12]

The decomposition (19) makes the source of divergence between private and social incentives transparent. The private firm values winning through π^w alone. The planner values winning through $\pi^w - K - \beta \log(P^w/P)$. The social return exceeds the private return whenever the CPI gain exceeds the social cost. As we show next, trade openness amplifies the CPI gain while leaving the private profit unchanged, and this is what can reverse the optimal policy from restriction to promotion.

The following proposition formalizes this comparison. We state it in terms of the best-response functions: $e^{\text{soc}}(e^*)$ is the planner's optimal R&D intensity given Foreign intensity e^*, and $e^{\text{priv}}(e^*)$ is the private firm's optimal intensity absent any R&D subsidy.[13]

Proposition 2 (Social vs. Private Innovation Incentives)**.** *The socially optimal R&D best response exceeds the private best response,*

$$e^{soc}(e^*) > e^{priv}(e^*),$$

if and only if

$$\underbrace{-\beta \log \frac{P^w}{P}}_{\substack{\text{direct CPI} \\ \text{gain from winning}}} - \underbrace{\theta\beta \log \frac{P^w}{P^\ell}}_{\substack{\text{race premium:} \\ \text{win vs. lose}}} > \underbrace{K}_{\substack{\text{social} \\ \text{cost}}} \tag{20}$$

where $K = r(\bar{A}) - r(A)$ is the incremental social cost and $\theta = e^*/\rho$ is the scooping weight from Section 2.6.

Proposition 2 characterizes when the social planner wants more AI innovation than the private market delivers. The left-hand side of condition (20) has two components, both of which are positive: the first because $P^w < P$, the second because $P^w < P^\ell$.

The first term, $-\beta \log(P^w/P)$, is the direct CPI gain from winning the innovation race. This gain arises because winning strengthens the domestic ecosystem, drawing more firms to Home and lowering the cost of living. The private AI firm does not internalize this ecosystem externality. This term is present regardless of what Foreign does and captures the planner's incentive to innovate even in the absence of strategic pressure.

The second term, $-\theta\beta \log(P^w/P^\ell)$, is a race premium. It reflects the additional social value of winning relative to *losing*, rather than relative to the pre-innovation state. The ratio P^w/P^ℓ captures the full ecosystem gap of the winner-takes-all AI innovation race: winning expands the domestic

[12] Since both countries access the frontier under AI trade, the social cost $r(\bar{A})$ is the same whether Home or Foreign wins. Therefore $K = r^w - r = r^\ell - r$, and $r^w - r^\ell = 0$.

[13] Both best-response functions have the same functional form: R&D intensity equals the value gap between winning and the current state, modulated by the cost of R&D. They differ only in what the continuation values reflect. The closed-form expressions are derived in Appendix B.2. Setting $s = 0$ for the private comparison isolates the wedge between private and social incentives from any existing subsidy.

ecosystem and lowers the CPI, while losing shrinks it and raises the CPI. The weight $\theta = e^*/\rho$ measures the strategic pressure from Foreign: when Foreign innovates more aggressively, the planner puts more weight on avoiding this losing state. The race premium captures the fact that, in international competition, the planner not only wants to win but also wants to avoid losing.

The right-hand side is the incremental social cost of AI, $K = r(\bar{A}) - r(A)$, the increase in the reduced-form cost $r(\cdot)$ when AI technology moves from the current level A to the frontier $\bar{A}$. The private firm does not bear this cost, but the planner internalizes it. If K is sufficiently large, the social return to AI innovation falls below the private return, and the planner wants to restrict rather than promote AI innovation. The question is whether trade openness can flip this comparison.

The role of trade openness. Both CPI terms on the left-hand side of (20) are increasing in trade openness. Lower trade costs make the equilibrium ecosystem more responsive to technology differences. Under free entry, the world mass of varieties is fixed, but the distribution across countries shifts with the technology gap z. As τ falls, a given cost advantage from winning translates into a larger reallocation of firms toward Home, and each firm that relocates replaces an imported variety, which carries the iceberg cost τ, with a domestic one, which does not, lowering the Home CPI. When τ is high, this reallocation is muted: closer to autarky, the equilibrium ecosystem becomes inelastic, and in the limit each country's firm mass is pinned at $\beta/(\sigma f)$ regardless of technology. The ecosystem channel shuts down entirely.

Collecting both CPI terms on the left-hand side of (20), define the threshold

$$K^{\text{crit}}(\tau) = \frac{\beta}{\sigma-1}\log\left(\left(\frac{\bar{A}}{A}\right)^{\alpha\gamma(1-\xi)(\sigma-1)}\bar{z}\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}\right) + \theta\frac{\beta}{\sigma-1}\log\left(\frac{\bar{z}-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}\right),$$

where $\bar{z} = (\bar{A}/A)^{\vartheta}$. Condition (20) is equivalent to $K < K^{\text{crit}}(\tau)$: the planner wants to subsidize innovation beyond the private level whenever the social cost falls below this threshold. The first term is the direct CPI gain from winning; the second is the race premium, which depends only on $\bar{z}$ because the global frontier cancels in the win-vs-lose comparison.

The first term in $K^{\text{crit}}(\tau)$ decomposes further into a direct marginal-cost gain and a trade amplification term,

$$\frac{\beta}{\sigma-1}\log\left(\left(\frac{\bar{A}}{A}\right)^{\alpha\gamma(1-\xi)(\sigma-1)}\bar{z}\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}\right) = \underbrace{\beta\alpha\gamma\log\frac{\bar{A}}{A}}_{\text{direct marginal-cost gain}} + \underbrace{\frac{\beta}{\sigma-1}\log\left(\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}\right)}_{\text{trade-induced amplification}}.$$

The direct marginal-cost gain is the welfare benefit of lowering the cost of Home differentiated production through the full AI productivity spillover, combining both the domestic and global channels ($A_t^{\gamma\xi}\bar{A}_t^{\gamma(1-\xi)}$). This term is the full $\beta\alpha\gamma\log(\bar{A}/A)$ regardless of ξ and survives under autarky. The trade-induced amplification term captures the additional CPI gain from ecosystem

reallocation under trade: it depends only on the domestic spillover ratio $\bar{z}$ and vanishes as $\tau \to \infty$. The second term in $K^{\text{crit}}(\tau)$ is the race premium, which is strictly positive whenever $\theta > 0$ and the economy is open.

The threshold $K^{\text{crit}}(\tau)$ is strictly decreasing in trade cost τ: the planner is willing to tolerate strictly larger social costs when trade is more open. As $\tau \to \infty$,

$$\lim_{\tau \to \infty} K^{\text{crit}}(\tau) = (1 + \theta\xi)\beta\alpha\gamma \log \frac{\bar{A}}{A}.$$

As τ falls toward the interiority bound $\underline{\tau} = (\bar{z} + \sqrt{\bar{z}^2 - 1})^{1/(\sigma-1)}$, the lowest trade cost at which both countries still produce differentiated goods, the threshold rises to a finite maximum. The monotonicity of $K^{\text{crit}}(\tau)$ implies a cutoff trade cost: for any K in the feasible range, there exists a unique $\tau^*(K)$ such that the social return to innovation exceeds the private return if and only if $\tau < \tau^*(K)$.

3.3 Optimal Innovation Policy

We are now ready to state the main results. We begin with the closed-economy benchmark defined in Section 2.7, which isolates the trade and strategic-competition channels by eliminating both differentiated-goods trade and the winner-takes-all innovation race.

Closed-economy benchmark. In the closed economy, there is no domestic ecosystem channel: with no trade, the mass of domestic varieties is pinned at $N = \beta/(\sigma f)$ regardless of technology. The CPI falls with AI innovation only through the direct marginal-cost channel. Moreover, there is no strategic interaction with a foreign rival, so the race premium is absent. The following proposition classifies the three possible regimes for autarky innovation policy.

Proposition 3 (Innovation Policy under Autarky)**.** *In the closed-economy benchmark of Section 2.7, define*

$$\bar{K} = \frac{\beta\alpha}{\mu}\left(1 - \frac{A}{\bar{A}}\right) + \beta\alpha\gamma \log \frac{\bar{A}}{A} \qquad \textit{and} \qquad \underline{K} = \beta\alpha\gamma \log \frac{\bar{A}}{A}.$$

Let e^{soc} denote the socially optimal AI R&D intensity. Then:

(i) Ban: If the social cost of AI is sufficiently high ($K \geq \bar{K}$), the planner sets $e^{soc} = 0$.

(ii) Tax: If the social cost is moderate ($\underline{K} \leq K < \bar{K}$), the planner sets $e^{soc} > 0$ but $e^{soc} < e^{priv}$.

(iii) Subsidize: If the social cost is low ($K < \underline{K}$), the planner sets $e^{soc} > e^{priv}$.

The two thresholds in Proposition 3 have transparent economic content. The lower threshold $\underline{K} = \beta\alpha\gamma \log(\bar{A}/A)$ is the CPI gain from AI innovation under autarky, which operates entirely

through the direct marginal-cost reduction in the differentiated sector via the productivity spillover A^{γ}. When the social cost falls below this threshold, the social CPI gain exceeds the social cost, and the planner wants to subsidize innovation beyond the private level.

The upper threshold $\bar{K}$ adds the AI profit gain $(\beta\alpha/\mu)(1 - A/\bar{A})$ to the CPI gain. When social costs exceed even the combined profit and CPI gains, any positive level of AI innovation is socially wasteful, and the planner bans AI innovation entirely. Between the two thresholds, the planner allows innovation but at a level below the private optimum—the social return is positive but smaller than the private return.

Three features of the autarky benchmark are worth noting. First, the AI profit term carries the factor $\beta\alpha/\mu$ (not $2\beta\alpha/\mu$) because the closed economy has domestic market size β, not the world market 2β. Second, $\bar{K} > \underline{K}$ for all $\bar{A} > A$, so the tax regime always exists. Third, the thresholds depend on the technology step $\bar{A}/A$, the productivity spillover γ, and the market-structure parameters α, σ, but not on trade costs, since trade is absent.[14]

The policy reversal. We now state the paper's central result: for an intermediate range of social costs, autarky calls for a complete ban on AI innovation while trade calls for subsidizing AI innovation above the private level.

Proposition 4 (Innovation Policy Reversal)**.** *Maintain the interiority conditions of Section 2.5. If the social cost satisfies*

$$\underbrace{\frac{\beta\alpha}{\mu}\left(1-\frac{A}{\bar{A}}\right)+\beta\alpha\gamma\log\frac{\bar{A}}{A}}_{\bar{K}\text{: autarky ban threshold}} < K < \underbrace{\frac{\beta}{\sigma-1}\log\left(\left(\frac{\bar{A}}{A}\right)^{\alpha\gamma(1-\xi)(\sigma-1)}\bar{z}\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}\right)+\theta\frac{\beta}{\sigma-1}\log\left(\frac{\bar{z}-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}\right)}_{K^{crit}(\tau)\text{: open-economy social-innovation threshold}}, \tag{21}$$

then:

(i) Under the closed-economy benchmark of Section 2.7, the socially optimal level of AI innovation is $e^{soc} = 0$.

(ii) Under trade, the socially optimal R&D best response exceeds the private best response: $e^{soc}(e^*) > e^{priv}(e^*)$.

Proposition 4 establishes that international competition can qualitatively reverse optimal innovation policy. A closed-economy planner facing the same social cost would ban AI innovation entirely. But the presence of trade and a foreign rival creates social returns to innovation—through the domestic ecosystem channel and the race premium, that are large enough to make subsidizing innovation socially optimal.

[14] We defer the formal proof to Appendix B.3. The ban threshold follows from verifying that the planner's continuation value from winning exceeds the no-innovation value if and only if $K < \bar{K}$.

The mechanism works as follows. In the closed economy, the only social benefit of innovation is the direct marginal-cost reduction in the differentiated sector. When social costs exceed this benefit plus the private profit gain, any positive AI innovation is welfare-reducing. In the open economy, two additional forces appear. First, innovation shifts the composition of world varieties toward Home, lowering the Home CPI by more than the direct marginal-cost effect alone. This ecosystem channel operates through free entry and is stronger when trade is more open because imports carry a larger CPI penalty. Second, because the innovation race is winner-takes-all, international competition creates a race premium: the planner values winning not only for its direct benefits but to avoid losing, which contracts the domestic ecosystem and cedes AI profits to Foreign. This premium is absent in the closed economy, where there is no rival.

The condition (21) places social costs in an interval whose bounds have natural interpretations. The lower bound $\bar{K}$ is the autarky ban threshold from Proposition 3: costs above this level make AI innovation socially wasteful in autarky. The upper bound $K^{\text{crit}}(\tau)$ is the open-economy threshold from Proposition 2: costs below this level make the social return exceed the private return under trade. The interval is non-empty whenever $K^{\text{crit}}(\tau) > \bar{K}$; we characterize when this inequality holds below.

Comparative statics. The interval in (21) widens as trade becomes more open (lower τ), as the technology step $\bar{A}/A$ increases, and as the scooping weight θ rises. The first effect operates through the CPI channel: lower trade costs make the ecosystem effect more powerful. The second effect operates through both channels: a larger technology step raises the competitiveness gain from winning and thereby strengthens both the CPI gain and the race premium. The third effect operates through the race premium: a higher θ means Foreign innovates more aggressively relative to discounting, so the Home faces a greater risk of being scooped in the AI innovation race and the planner places more weight on the CPI gap between winning and losing.

The interval shrinks as $\tau \to \infty$: in the autarky limit, $K^{\text{crit}}(\tau) \to (1+\theta\xi)\beta\alpha\gamma \log(\bar{A}/A)$, and whether the interval is non-empty depends on the balance between the race premium and the appropriable private prize. The ξ enters because the race premium at autarky depends only on the domestic spillover: the win-vs-lose CPI gap scales with $\bar{z}$ rather than the full technology gap. Without scooping ($\theta = 0$), the interval shrinks to zero at $\tau \to \infty$—the autarky ban threshold exactly equals the autarky critical threshold—and the reversal can only occur if trade is sufficiently open.

Sufficient conditions. The following proposition identifies primitive conditions under which the interval in Proposition 4 is non-empty for every trade cost satisfying interiority.

Proposition 5 (Sufficient Condition for Policy Reversal)**.** *If*

$$\theta\gamma\xi \log \frac{\bar{A}}{A} > \frac{\sigma-1}{\sigma}\left(1 - \frac{A}{\bar{A}}\right), \tag{22}$$

then for every $\tau > \underline{\tau}(\bar{z})$ satisfying the interiority condition, there exists a non-empty interval of social costs K such that the conclusion of Proposition 4 holds.

Proposition 5 provides a clean sufficient condition that is independent of τ.

The left-hand side, $\theta\gamma\xi \log(\bar{A}/A)$, is the effective social gain from winning the innovation race, combining the domestic spillover strength $\gamma\xi$ with the scooping weight θ. Only the domestic spillover component enters because the global frontier cancels in the win-vs-lose comparison that drives the race premium. The right-hand side, $(\sigma-1)/\sigma \cdot (1 - A/\bar{A})$, is the appropriable private gain from innovating, scaled by the monopolistic competition wedge $1/\mu = (\sigma-1)/\sigma$ that maps spending into variable-cost demand for AI. A notable feature is that α and β cancel from the condition: under Cobb-Douglas production, both the private prize and the social CPI gain scale proportionally with the importance of AI in production and with the scale of the differentiated sector.

For small innovation steps ($\bar{A}/A \approx 1$), both sides are approximately linear in $\bar{A}/A - 1$, and the condition reduces to $\theta\gamma\xi > (\sigma-1)/\sigma$. For large innovation steps ($\bar{A}/A \to \infty$), $\log(\bar{A}/A) \to \infty$ while $1 - A/\bar{A} \to 1$, so the condition holds automatically whenever $\theta\gamma\xi > 0$. Economically, large technology steps always generate a region of social costs where the policy reversal occurs, provided there is any strategic pressure from Foreign innovation and the domestic spillover share is positive.

When condition (22) fails, the reversal may still occur but only for sufficiently open trade. In this case, there exists a unique trade-cost threshold τ^* such that the interval in Proposition 4 is non-empty if and only if $\tau < \tau^*$. The threshold τ^* is decreasing in θ (a larger scooping weight relaxes the openness requirement) and decreasing in $\bar{z}$ (a larger technology gap relaxes it), but increasing in the appropriable private prize (a larger private gain requires more openness to outweigh).

Why openness changes the policy wedge. The core of the mechanism is that trade openness raises the social return to innovation without raising the private return. AI profits are independent of τ because world spending on differentiated goods is pinned at 2β and AI's cost share is constant under Cobb-Douglas. Instead, openness raises the consumer-surplus component of the prize: under trade, domestic varieties enter the CPI without the iceberg penalty, so a stronger domestic ecosystem translates directly into a lower cost of living. Innovation makes the winner more competitive, and free entry strengthens the winner's ecosystem, delivering a CPI gain that the private firm does not capture. This is why an open economy can exhibit $e^{\text{soc}}(e^*) > e^{\text{priv}}(e^*)$ even when the closed economy exhibits the opposite.

Three modeling assumptions isolate the domestic ecosystem mechanism cleanly: free entry pins world variety creation and makes its composition responsive to competitiveness, iceberg trade costs create a CPI wedge between domestic and imported varieties, and Bertrand AI pricing with unit-elastic demand makes AI profits τ-independent.[15] These assumptions are conservative. Relax-

[15] The ecosystem channel can be understood as a variant of the home market effect in Krugman (1980) and the

ing the last feature, for instance by introducing less elastic AI demand, imperfect substitutability across AI suppliers, or AI trade costs, would add a direct user-price channel that further strengthens the social return to winning and can enlarge the reversal region.[16]

3.4 Cooperative Innovation Policy

We now ask what innovation policy a global planner would choose if she could jointly set R&D intensities in both countries. This cooperative benchmark strips out the two forces that drive national planners toward subsidizing AI innovation: the beggar-thy-neighbor motive embedded in the domestic ecosystem channel and the arms-race premium arising from fear of being scooped. The comparison between cooperative and non-cooperative policy therefore isolates the role of international competition in shaping innovation incentives. The main result is that national competition always leads to over-innovation relative to the cooperative benchmark, and that the magnitude of the excess is driven by precisely the forces that generate the policy reversal of Proposition 4.

The global planner's problem. A global planner maximizes world welfare $\mathcal{W}^G = \mathcal{W} + \mathcal{W}^*$, defined as the sum of Home and Foreign household utility. Because the model is symmetric—both countries start at $A_0 = A_0^* = A$ and share identical structural parameters—the post-innovation world state delivers the same aggregate welfare regardless of which country innovates first. If Home wins, world flow utility is $u^w + u^\ell$; if Foreign wins, it is $u^\ell + u^w$. The identity of the winner is a pure distributional question from the global perspective.

This is the fundamental asymmetry between the cooperative and non-cooperative problems: the national planner faces a race with differentiated outcomes (winning vs. losing), while the global planner faces a simple adoption problem (pre-innovation vs. post-innovation). Because both countries' innovations lead to the same world state, the combined arrival rate is $e + e^*$ and the global planner's recursive problem is

$$\rho \mathcal{W}^G = \max_{e, e^* \geq 0} \left\{ 2\tilde{u} - \frac{1}{2}e^2 - \frac{1}{2}(e^*)^2 + (e + e^*)(\mathcal{W}^{G,\text{post}} - \mathcal{W}^G) \right\},$$

where $\mathcal{W}^{G,\text{post}} = \frac{1}{\rho}(u^w + u^\ell)$ is the post-innovation world continuation value and $\tilde{u}$ is per-country flow utility gross of R&D expenditure. Three features distinguish this problem from the national planner's problem (17). First, there is a single arrival rate $e + e^*$ rather than separate winning and losing hazards—the $e^*(\mathcal{W}^\ell - \mathcal{W})$ term is absent because losing and winning deliver the same world

agglomeration forces in Krugman (1991). In those frameworks, increasing returns and trade costs jointly determine where production locates. Here, the AI productivity spillover acts as an external increasing return at the industry level: the winner's firms all become more competitive, attracting entrants and shifting production home. The key difference is that the total mass of varieties is fixed by free entry, so the reallocation is purely redistributive across countries rather than variety-creating.

[16] Section 4 discusses several such extensions. The domestic ecosystem channel studied here is the cleanest analytical mechanism, but it likely understates the full social return to innovation in richer environments.

welfare. Second, the effective discount rate is ρ, not $\tilde{\rho}(e^*) = \rho + e^*$, because the global planner faces no urgency from a rival. Third, the R&D cost is the world total $\frac{1}{2}e^2 + \frac{1}{2}(e^*)^2$.

The first-order conditions yield $e = e^* \equiv e^G$: with convex R&D costs, the cheapest way to achieve any target combined arrival rate $\lambda = e + e^*$ is to split effort equally. Substituting back and solving the implied quadratic, the cooperative R&D intensity satisfies

$$(e^G)^2 + \rho e^G = \Lambda^G,$$

where Λ^G is the global effective prize defined below. The unique positive solution is $e^G = \frac{1}{2}(-\rho + \sqrt{\rho^2 + 4\Lambda^G})$, which is positive if and only if $\Lambda^G > 0$.[17]

The global effective prize. The global effective prize admits a transparent decomposition. Using the indirect utility expression from Appendix A.1 and noting that pre-innovation profits vanish under symmetric technology ($\pi = 0$), we have

$$\Lambda^G = \pi^w - 2K \underbrace{-\beta \log \frac{P^w}{P}}_{\substack{\text{winner's} \\ \text{CPI gain}}} \underbrace{-\beta \log \frac{P^\ell}{P}}_{\substack{\text{loser's} \\ \text{CPI change}}},$$

where the factor of 2 on the social cost $K = r(\bar{A}) - r(A)$ reflects the fact that since both countries access the frontier under AI trade, innovation raises social costs in *both* countries. The national planner internalizes only K; the global planner internalizes $2K$.

The net world CPI effect decomposes into a world marginal-cost gain and an ecosystem-reallocation residual:

$$-\beta \log \frac{P^w}{P} - \beta \log \frac{P^\ell}{P} = \underbrace{\beta\alpha\gamma(2-\xi) \log \frac{\bar{A}}{A}}_{\text{world marginal-cost gain}} + \underbrace{\frac{\beta}{\sigma - 1} \log \frac{(1-\tau^{1-\sigma})^2}{(1-\tau^{1-\sigma}\bar{z})(1-\tau^{1-\sigma}/\bar{z})}}_{\text{net ecosystem residual}}. \tag{23}$$

The first term is the combined consumer-price benefit from lower production costs across both countries. The winner gains $\beta\alpha\gamma \log(\bar{A}/A)$ from the full domestic and global spillover, and the loser gains $\beta\alpha\gamma(1-\xi) \log(\bar{A}/A)$ from the global frontier spillover $\tilde{A}_t^{\gamma(1-\xi)}$ alone. The sum is a genuine world efficiency gain that scales with $(2-\xi)$: when the spillover is mostly global (ξ small), both countries benefit nearly equally from innovation and the world marginal-cost gain approaches $2\beta\alpha\gamma \log(\bar{A}/A)$; when the spillover is purely domestic ($\xi = 1$), only the winner benefits and the world gain is just $\beta\alpha\gamma \log(\bar{A}/A)$.

The second term, the net ecosystem residual, is the key object for understanding the divergence between national and global incentives.

[17] We derive the full solution in Appendix B.6. Compare with the national planner's quadratic $(e^N)^2 + \tilde{\rho}(e^*)e^N = 2\Lambda^{\text{soc}}$: the cooperative problem has a lower effective discount rate (ρ vs. $\rho + e^*$) and a different effective prize.

The domestic ecosystem channel as a near-zero-sum transfer. The domestic ecosystem channel is the central mechanism of the paper: innovation shifts the equilibrium composition of world varieties toward the winner, strengthening the winner's ecosystem and weakening the loser's. From the national planner's perspective, the full winner's ecosystem gain $\frac{\beta}{\sigma-1}\log\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}$ is a benefit of winning the race. From the global planner's perspective, this gain is largely offset by the loser's ecosystem loss $\frac{\beta}{\sigma-1}\log\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}/\bar{z}} < 0$.

The net ecosystem residual in (23) is strictly positive but small relative to the gross winner's ecosystem gain. The economic intuition is straightforward. Total world expenditure on differentiated goods is pinned at 2β. Free entry pins total world varieties at $N + N^* = 2\beta/(\sigma f)$. Innovation does not change how much the world spends or how many firms operate. It only changes *where* those firms locate, concentrating production in the lower-cost country and reducing aggregate iceberg waste. The national planner counts the full CPI benefit of this relocation. The global planner largely nets it out.

Proposition 6 (Cooperative Innovation Policy)**.** *Maintain the interiority conditions of Section 2.5. Define the global ban threshold*

$$\bar{K}^G = \frac{1}{2}\left[\pi^w - \beta\log\frac{P^w}{P} - \beta\log\frac{P^\ell}{P}\right]$$

and the global subsidy threshold

$$\underline{K}^G = \frac{1}{2}\left[-\beta\log\frac{P^w}{P} - \beta\log\frac{P^\ell}{P}\right].$$

Then the cooperative planner's optimal innovation policy satisfies:

(i) Ban: If $K \geq \bar{K}^G$, the global planner sets $e^G = 0$.

(ii) Tax: If $\underline{K}^G \leq K < \bar{K}^G$, the global planner sets $e^G > 0$ but $e^G < e^{priv}(e^G)$.

(iii) Subsidize: If $K < \underline{K}^G$, the global planner sets $e^G > e^{priv}(e^G)$.

The structure of Proposition 6 mirrors that of Proposition 3, with global thresholds replacing the autarky thresholds. The subsidy threshold $\underline{K}^G$ is half the net world CPI gain from innovation: it represents the per-country social cost at which the net consumer-surplus externality exactly offsets the social cost.[18]

[18] We defer the proof to Appendix B.6. The ban threshold follows from $\Lambda^G = 0$; the subsidy threshold follows from comparing e^G with the private best response at $e^* = e^G$.

National versus cooperative incentives. Comparing the global effective prize Λ^G with the national planner's effective prize from Proposition 2, several forces systematically make the cooperative planner more cautious about AI innovation. The national planner values winning through the full CPI gain $-\beta \log(P^w/P)$, which includes the large gross ecosystem component, and the race premium $\theta\beta \log(P^w/P^\ell)$, whose win-vs-lose CPI gap depends only on relative prices and hence only on the domestic spillover. The global planner values innovation through the net world CPI effect, in which the ecosystem gains and losses nearly cancel, and has no race-premium term at all. At the same time, the national planner's effective discount rate $\tilde{\rho}(e^*) = \rho + e^*$ exceeds ρ because every instant without innovation risks being scooped, raising national innovation intensity for any given prize level; the global planner faces no such urgency. Finally, the global planner internalizes $2K$ rather than K, since both countries incur the social cost under AI trade. Moreover, the ecosystem and scooping channels compound each other: the race premium exists only because the win-vs-lose CPI gap is driven by the domestic spillover, and the higher discount rate $\rho + e^*$ magnifies the national prize that this gap inflates. Without the ecosystem channel, neither force would create a wedge between national and cooperative incentives.

The cooperative planner bans innovation at a strictly lower social cost than the autarky planner: $\bar{K}^G < \bar{K}$. At goods autarky ($\tau \to \infty$), the ecosystem residual vanishes and the gap reduces to

$$\bar{K} - \bar{K}^G\Big|_{\tau\to\infty} = \frac{\xi}{2}\beta\alpha\gamma \log\frac{\bar{A}}{A} > 0,$$

which is positive because the cooperative planner internalizes social costs in both countries while the autarky planner internalizes only one. The ranking extends by continuity to trade costs near autarky.

Proposition 7 (Over-Innovation under National Competition)**.** *Maintain the interiority conditions of Section 2.5. Then at any symmetric interior Nash equilibrium of the national planners' game, the innovation intensity exceeds the cooperative optimum:*

$$e^N > e^G.$$

The proof of Proposition 7 follows from two observations.[19] First, the national planner's effective prize exceeds the per-country global prize at any common rival intensity, because the national planner captures the full gross reallocation gain rather than half the net world effect, and internalizes only K rather than $2K$. Second, the national planner faces a higher effective discount rate $\rho + e^N > \rho$, which further raises innovation intensity.

[19] The formal proof is in Appendix B.7. The best-response comparison requires a fixed-point argument to translate into an equilibrium comparison. Under the strategic complementarity derived in the innovation game, the comparison of best responses implies the comparison of equilibrium intensities.

The three-way policy comparison. Combining Propositions 3, 4, and 6, we can rank innovation across three regimes: autarky, cooperative open-economy, and non-cooperative open-economy. The ranking reveals the extent to which the policy reversal of Proposition 4 is driven by an arms-race externality.

When social costs are low ($K < \underline{K}^G$), all three planners want more AI innovation than the private market delivers. The ranking is $e^N > e^G > e^{\text{priv}}$: both planners subsidize, but national competition leads to more aggressive innovation due to the ecosystem externality and race premium.

The starkest case arises in the intermediate range identified by Proposition 4. Since $\bar{K}^G < \bar{K} < K^{\text{crit}}(\tau)$ (at least near goods autarky), there exists a sub-interval $[\bar{K}^G, \bar{K}]$ in which the autarky planner bans innovation, the cooperative planner also bans innovation, but the non-cooperative national planner subsidizes it. In this region, international competition drives countries to innovate aggressively even when both the closed-economy and the cooperative benchmarks call for no innovation at all. The entire innovation incentive is driven by beggar-thy-neighbor ecosystem competition and the fear of being scooped, forces that are redistributive from a global perspective.

For even larger social costs ($K > K^{\text{crit}}(\tau)$), all planners ban or restrict AI innovation. But the national planner is the last to ban, because the ecosystem externality and race premium sustain her innovation incentive even at high social costs.

In summary, the cooperative planner occupies a middle ground: she internalizes the genuine world efficiency gains from innovation—the world marginal-cost reduction and the small net ecosystem residual, but strips out the beggar-thy-neighbor and arms-race motives. The national planner over-innovates relative to this benchmark, and the magnitude of the excess is increasing in trade openness and the scooping weight, precisely the forces that make the domestic ecosystem externality quantitatively important.

4 AI Export Controls and the Geopolitics of the Innovation Race

Once a country wins the AI race, frontier technology becomes a strategic asset. The winner can deny rivals access through export controls; the loser can refuse imports to preserve domestic autonomy. We endogenize these choices as a second-stage policy game following the innovation race of Section 3. The equilibrium balances a static gain from accessing the frontier spillover through trade, a social-cost saving from remaining on lagging AI, and the option value of domestic catch-up R&D. Three qualitatively distinct regimes can emerge, each with different implications for first-stage innovation incentives. The main finding is that when the winner optimally bans AI exports to preserve its competitive advantage, the welfare-versus-profit wedge of Section 3 strictly widens: the private return to AI innovation falls while the social return rises, strengthening the policy reversal by expanding the range of social costs for which international competition flips a ban into a subsidy. We first describe the policy game and its key objects, then characterize the equilibrium, and finally map the second-stage outcomes back into first-stage innovation incentives.

4.1 The Post-Innovation Policy Game

Suppose Home has won the innovation race, so $A_t = \bar{A}$ and $A_t^* = A$. Goods trade continues at iceberg cost τ. The new question is whether the loser also accesses the winner's frontier AI technology.

The accessible frontier. We maintain throughout this section that an open-source AI vintage at level A is freely available in both countries.[20] The AI price is $\tilde{q} = 1/A$ in every country and every regime. Export controls affect not the price of AI as an intermediate input, but whether the loser's production sector benefits from the frontier spillover. Recall from Section 2.2 that differentiated-goods production features both a domestic spillover $A_t^{\gamma\xi}$ and a global frontier spillover $\tilde{A}_t^{\gamma(1-\xi)}$, where $\tilde{A}_t = \max\{A_t, A_t^*\}$ and $\xi \in (0,1)$ governs the split between the two channels. Under AI trade, both countries access the frontier, so $A_t^{\text{acc}} = A_t^{*,\text{acc}} = \bar{A}$. Under AI autarky, the loser loses access to the frontier spillover:

$$A_t^{*,\text{acc}} = A \qquad \text{and} \qquad A_t^{\text{acc}} = \bar{A}.$$

This distinction has a direct consequence for the goods equilibrium. Under AI trade, only the domestic technology differential governs relative competitiveness, so the effective relative-productivity gap is $z^{\text{trade}} = \bar{z}$. Under AI autarky, the loser also forfeits the global frontier spillover, and the effective relative-productivity gap rises to

$$z^{\text{aut}} = \left(\frac{\bar{A}}{A}\right)^{\alpha\gamma(\sigma-1)} > z^{\text{trade}},$$

which reflects the full technology gap rather than just the domestic spillover component $\bar{z} = (\bar{A}/A)^{\alpha\gamma\xi(\sigma-1)}$.[21] Since the effective technology gap facing the loser's producers is larger under autarky than under trade ($z^{\text{aut}} > z^{\text{trade}}$), autarky shifts more varieties toward the winner and worsens the loser's cost of living relative to the trade regime.

Social costs. Under AI trade, both countries access the frontier and incur social cost $r(\bar{A})$. Under AI autarky, the loser's accessible technology remains at A, so its social cost is only $r(A)$—the loser avoids the incremental cost K by forfeiting frontier access.

The geopolitical AI trade game. The second-stage interaction is an offer-and-response game:

[20] This assumption extends the open-source fringe from the closed-economy benchmark of Section 2.7 to the post-innovation phase of the open economy. Under AI trade, Bertrand competition between the winner and the open-source fringe drives the AI price to $1/A$. Under AI autarky, the domestic fringe alone delivers the same price. The policy choice therefore does not move the AI price—it moves access to the frontier production spillover.

[21] Under AI trade, the global frontier $\tilde{A}_t = \bar{A}$ benefits both countries' production equally, so only the domestic spillover differential $\bar{z}$ governs relative competitiveness. Under AI autarky, the loser loses the frontier spillover and the full technology gap $z^{\text{aut}} = (\bar{A}/A)^{\alpha\gamma(\sigma-1)}$ determines relative competitiveness.

(i) The winning country chooses whether to ban AI exports or to allow exports at a per-period fee $T \geq 0$.

(ii) The losing country chooses whether to accept trade at the offered fee or to reject it and remain in AI autarky.

If the winner bans exports or the loser rejects, AI autarky prevails. If both sides agree, AI trade occurs at fee T and the resulting state is terminal: the technology gap persists but the loser gains frontier access. The fee T is a sovereign transfer from the losing country's government to the winning country's government. It enters social continuation values but does not accrue to the private AI firms.[22]

The catch-up option. Under AI autarky, the loser can invest in domestic catch-up R&D at intensity $e^{*,\text{soc}}$, at labor cost $\frac{1}{2}(e^{*,\text{soc}})^2$, generating a Poisson arrival rate $e^{*,\text{soc}}$ of reaching technology $\bar{A}$. If catch-up succeeds, the two countries return to symmetric technology levels, but AI trade remains banned, so each AI firm serves only its domestic market.[23] The loser's planner chooses catch-up intensity to maximize household welfare. Let

$$\Delta^{*,\text{catch}} \equiv \frac{\alpha\beta}{\mu}\left(1 - \frac{A}{\bar{A}}\right) - K + \frac{\beta}{\sigma - 1}\log\frac{z^{\text{aut}} - \tau^{1-\sigma}}{1 - \tau^{1-\sigma}}$$

denote the flow-utility gain from catching up: a domestic AI-profit stream, less the incremental social cost, plus the CPI improvement from eliminating the ecosystem distortion. The socially optimal catch-up intensity is then

$$\widehat{e}^{*,\text{soc}} = -\rho + \sqrt{\rho^2 + 2(\Delta^{*,\text{catch}})_+}. \tag{24}$$

When $\Delta^{*,\text{catch}} \leq 0$, the loser does not invest in catch-up and autarky is effectively terminal. When the social cost K is small relative to the profit and CPI gains, the catch-up option is valuable and the loser invests at a strictly positive rate.

4.2 Equilibrium of the AI Trade Game

The equilibrium of the geopolitical AI trade game is determined by two thresholds: the loser's maximum acceptable fee $T^{\max}$ and the winner's reservation fee T^{res}. We characterize each in turn.

[22] We adopt two tie-breaking conventions: the loser accepts at indifference, and the winner allows trade at the maximum acceptable fee when indifferent between exporting and banning.

[23] After catch-up, both firms have technology $\bar{A}$ but each limit-prices against the domestic open-source fringe at $1/A$, earning domestic-monopoly profits $\frac{\alpha\beta}{\mu}(1 - A/\bar{A})$. The ecosystem distortion disappears because relative productivity returns to 1.

The loser's willingness to pay. The loser accepts AI trade at fee T if and only if trade delivers at least as much continuation welfare as AI autarky. The maximum acceptable fee equates the two continuation values.

Proposition 8 (Loser's Willingness to Pay for Frontier AI)**.** *In the second stage after Home wins ($\bar{A} > A$), the loser accepts AI trade if and only if $T \leq T^{max}$, where*

$$T^{max} = \underbrace{\frac{\beta}{\sigma - 1} \log \frac{z^{aut} - \tau^{1-\sigma}}{z^{trade} - \tau^{1-\sigma}}}_{\text{static access gain}} - \underbrace{K}_{\substack{\text{social-} \\ \text{cost saving}}} - \underbrace{\frac{1}{2}(\widehat{e}^{*,soc})^2}_{\substack{\text{catch-up} \\ \text{option value}}},$$

with $\widehat{e}^{*,soc}$ given by (24). This holds under the interiority conditions of Section 2.5 applied to both z^{trade} and z^{aut}.

Proposition 8 decomposes the loser's willingness to pay into three terms with transparent economic content.

The first term is the loser's static gain from frontier access. Under AI autarky, the loser faces a productivity disadvantage of z^{aut} rather than z^{trade} because it forfeits the global frontier spillover. This amplifies the ecosystem distortion: more firms shift toward the winner, raising the loser's cost of living. Trade restores the frontier spillover and softens this ecosystem imbalance. The term is strictly positive because the loser faces a wider technology gap under autarky ($z^{\text{aut}} > z^{\text{trade}}$) and is increasing in the global spillover component $(1 - \xi)$.

The second term is the social-cost saving from autarky. By remaining on the lagging AI vintage A, the loser avoids the incremental social cost K associated with using frontier AI. This saving is weakly positive as a motive for autarky and grows with the sensitivity of social costs to AI capability.

The third term is the dynamic option value of catch-up innovation. Under autarky, the loser can invest in R&D to reach $\bar{A}$ independently. The value of this option, $\frac{1}{2}(\widehat{e}^{*,\text{soc}})^2$, is strictly positive whenever the catch-up gain $\Delta^{*,\text{catch}}$ is positive and vanishes otherwise.[24]

When the loser prefers autarky. The loser rejects all non-negative fees ($T^{\text{max}} < 0$) when the combined social-cost and option-value motives outweigh the static access gain. This is more likely when K is large (frontier AI is socially costly), when ρ is low (the catch-up option is more valuable), and when the global spillover component $(1 - \xi)$ is small (losing frontier access matters less for production). Conversely, when the global spillover is quantitatively important, the positive access term dominates and the loser is willing to pay a substantial fee for frontier AI.

[24] The option-value term takes the quadratic form $\frac{1}{2}(\widehat{e}^{*,\text{soc}})^2$ because the loser's planning problem has the same linear-quadratic structure as the first-stage problem of Section 2.6.

The winner's reservation fee. The winner faces an analogous trade-off. Allowing AI exports generates revenue from serving both markets and eliminates catch-up risk, but banning exports preserves a stronger competitive advantage in the goods market by keeping the effective technology gap facing the loser's producers at z^{aut} rather than z^{trade}. The winner's reservation fee T^{res}—the minimum fee that makes trade at least as attractive as an export ban—summarizes this comparison. We show in Appendix C.1 that

$$T^{\text{res}} = \underbrace{-\frac{\alpha\beta}{\mu}\left(1-\frac{A}{\bar{A}}\right)}_{\text{forgone export profit}} + \underbrace{\frac{\beta\rho}{(\rho+\widehat{e}^{*,\text{soc}})(\sigma-1)}\log\frac{1-\tau^{1-\sigma}z^{\text{trade}}}{1-\tau^{1-\sigma}z^{\text{aut}}}}_{\text{static strategic gain from autarky}} + \underbrace{\left(-\frac{\beta\widehat{e}^{*,\text{soc}}}{(\rho+\widehat{e}^{*,\text{soc}})(\sigma-1)}\log\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}z^{\text{trade}}}\right)}_{\text{dynamic benefit of eliminating catch-up risk}}. \quad (25)$$

The three terms in (25) have opposing signs. The first term is strictly negative: by banning exports, the winner forgoes the profit stream from serving the loser's market, which lowers the fee needed to make trade attractive. The second term is strictly positive: because the technology gap is wider under autarky ($z^{\text{aut}} > z^{\text{trade}}$), the winner enjoys a more favorable variety allocation and hence a lower cost of living. The third term is weakly negative and is strictly negative when $\widehat{e}^{*,\text{soc}} > 0$: trade removes the risk that the loser catches up and erodes the winner's competitive advantage.[25]

The sign of T^{res} is therefore ambiguous. When the static strategic gain from keeping the loser behind dominates, T^{res} is positive and the winner requires a substantial fee to allow trade. When the export-profit and risk-elimination channels dominate, T^{res} is negative and the winner prefers trade even at zero fee. The following proposition characterizes the equilibrium.

Proposition 9 (Equilibrium of the AI Trade Game)**.** *In the second stage after Home wins ($\bar{A} > A$), under the interiority conditions of Section 2.5:*

(i) Loser-initiated autarky. If $T^{max} < 0$, the loser rejects every non-negative fee. AI autarky prevails and the loser retains the catch-up option.

(ii) AI trade with fee extraction. If $T^{max} \geq 0$ and $T^{max} \geq T^{res}$, the winner charges $T = T^{max}$ and the loser accepts. AI trade occurs.

(iii) Winner-initiated export ban. If $T^{max} \geq 0$ and $T^{max} < T^{res}$, the winner bans AI exports. AI autarky prevails despite the loser's willingness to pay.

The first two cases are standard: either the loser's outside option is too attractive for any deal to

[25] After catch-up, relative productivity returns to 1 and the winner loses the ecosystem benefit. The winner's CPI rises after catch-up, making autarky less attractive when catch-up risk is high.

occur, or gains from trade exist and the winner captures them. The third case is the economically distinctive regime. Here the loser would benefit from frontier access and is willing to pay for it, but the maximum acceptable fee does not compensate the winner for the loss of its competitive advantage. The winner sacrifices export profits and accepts catch-up risk in order to keep the rival further behind in the goods market. This is a geoeconomic motive: the winner uses AI export controls as a tool of trade policy, exploiting the link between frontier AI access and competitive advantage in differentiated goods.

Which case prevails. Case (iii) is more likely when trade costs are moderate (so ecosystem composition is consequential) and the technology step $\bar{A}/A$ is large (so the stakes of the competitive-advantage game are high). The domestic spillover share ξ plays a non-monotone role. When $\xi = 1$ (purely domestic spillover), AI trade confers no frontier-access benefit on the loser, so $T^{\max}$ is more likely negative and the loser itself rejects trade (Case (i)). As ξ falls toward zero, the global spillover component grows and the loser's willingness to pay rises, but so does the winner's strategic motive for an export ban, because denying access to the frontier spillover widens the productivity gap from $\bar{z}$ all the way to z^{aut}. Case (iii) therefore requires ξ low enough that the loser values frontier access ($T^{\max} \geq 0$) yet high enough that the winner's competitive advantage is sensitive to export controls ($T^{\text{res}} > T^{\max}$).

4.3 Implications for Innovation Incentives

We now map the three equilibrium cases back into first-stage innovation incentives. As in Section 3.2, the comparison between social and private best responses turns on the effective prizes Λ^{soc} and Λ^{priv}—the respective gaps between winning and current continuation values. The second-stage geopolitical AI trade game modifies these prizes by changing the continuation values that the winner and loser anticipate.

Continuation values. Under AI trade, the private AI firm's continuation values are unchanged from Section 3: the winning firm serves both markets and earns $\pi^{w,\text{trade}} = 2\frac{\alpha\beta}{\mu}(1 - A/\bar{A})$, while the losing firm earns zero.[26] Under AI autarky, two changes occur. First, the winning firm serves only the domestic market and earns $\pi^{w,\text{aut}} = \frac{\alpha\beta}{\mu}(1 - A/\bar{A})$—exactly half the trade profit. Second, the losing firm retains a positive catch-up option and earns a strictly positive continuation value. Both changes reduce the private effective prize relative to the baseline.

For the planner, the continuation values additionally reflect CPI effects, social costs, and the fee transfer T. The social effective prizes under each case are derived in Appendix C.4. Their comparison with the baseline prizes from Section 3 yields the following result.

[26] The fee T is a sovereign transfer and does not accrue to the AI firm. Hence $\mathcal{V}^{w,\text{trade}} = \pi^{w,\text{trade}}/\rho$ and $\mathcal{V}^{\ell,\text{trade}} = 0$ exactly as in Section 3.

Proposition 10 (Feedback to First-Stage Innovation Incentives)**.** *Fix the rival's first-stage R&D intensity e^* and set $s = 0$ for the private comparison. Relative to the baseline of Section 3:*

(i) Loser-initiated autarky (Case (i) of Proposition 9). The private best-response R&D intensity is strictly lower than in the baseline. The socially optimal R&D intensity is higher than in the baseline if and only if

$$T^{res} > \frac{e^*}{\rho + e^*}(-T^{max}).$$

(ii) AI trade with fee extraction (Case (ii)). The private best-response R&D intensity is unchanged. The socially optimal R&D intensity is weakly higher, with the increase proportional to T^{max}:

$$\Lambda^{soc}_{trade} - \Lambda^{soc}_{baseline} = \frac{\rho + 2e^*}{\rho} T^{max} \geq 0.$$

(iii) Winner-initiated export ban (Case (iii)). The private best-response R&D intensity is strictly lower than in the baseline, while the socially optimal R&D intensity is strictly higher. The welfare-versus-profit wedge is strictly wider:

$$\Lambda^{soc}_{ban} - \Lambda^{priv}_{ban} > \Lambda^{soc}_{baseline} - \Lambda^{priv}_{baseline}.$$

Proposition 10 delivers the central message of this section: endogenous AI trade policy does not uniformly intensify or dampen the innovation race. It creates three distinct first-stage environments that affect private and social incentives asymmetrically.

Fee extraction and the social return (Case (ii)). When AI trade occurs with fee extraction, the private innovation game is exactly as in the baseline because the fee is a sovereign transfer, not private revenue. But the planner's prize increases by $\frac{\rho+2e^*}{\rho} T^{\max}$, reflecting the fiscal rent that the winning country extracts from the loser. If Home wins, it receives $T^{\max}$ per period; if Home loses, it pays $T^{\max}$ per period. Both effects raise the social value of winning relative to losing, so fee extraction amplifies the baseline divergence between social and private innovation incentives identified in Section 3.

Export bans and the geoeconomic arms race (Case (iii)). Case (iii) is the most striking regime. The private best-response R&D intensity falls for two reasons: the winning AI firm now serves only the domestic market, halving post-innovation profits, and the losing firm retains a positive catch-up option, making defeat less costly. Both forces reduce the private effective prize. The socially optimal R&D intensity, by contrast, rises. The winner gains strategically from keeping the loser further behind: the amplified productivity gap $z^{\text{aut}} > z^{\text{trade}}$ shifts more varieties toward the winner, delivering a CPI improvement that the AI firm does not internalize. At the same

time, losing becomes more costly for the planner because the loser faces the amplified ecosystem distortion without frontier access.

The result is that private and social incentives move in opposite directions: the firm wants to innovate less when export bans are anticipated, while the planner wants to innovate more. The welfare-versus-profit wedge strictly widens relative to the baseline, strengthening the case for innovation subsidies. This is a geoeconomic arms race in which the anticipation of post-innovation strategic competition raises the social stakes of the first-stage innovation race even as it reduces the private stakes.

Loser-initiated autarky and the ambiguity (Case (i)). In Case (i), the private best-response R&D intensity falls for the same reasons as in Case (iii). The socially optimal R&D intensity, however, moves ambiguously. Autarky benefits the winner through the strategic goods-market channel, but it hurts the loser by denying frontier access. The condition $T^{\text{res}} > \frac{e^*}{\rho+e^*}(-T^{\max})$ captures the relevant balance: the left-hand side is the winner-side strategic value of autarky, and the right-hand side is the probability-weighted loser-side welfare loss from foregone trade. If the winner's strategic gain dominates, the socially optimal R&D intensity rises despite the loser-side cost. The key difference from Case (iii) is that in Case (i), the loser-side welfare loss ($-T^{\max} > 0$) must be weighed against the winner-side gain, whereas in Case (iii) both margins push the social prize upward because $T^{\max} \geq 0$.

Connection to the baseline policy reversal. The baseline policy reversal of Proposition 4 established that for an intermediate range of social costs, autarky calls for banning AI innovation while the open economy calls for subsidizing it. Proposition 10 shows that endogenous AI trade policy can amplify this reversal. Under Case (ii), the private R&D intensity is unchanged while the socially optimal level weakly rises, so the welfare-versus-profit wedge weakly widens. Under Case (iii), the wedge is strictly wider because private and social incentives move in opposite directions. In both cases, the range of social costs for which the planner wants to subsidize innovation beyond the private level is weakly larger than in the baseline. The geoeconomic game reinforces the pro-innovation force of international competition.[27]

Discussion. The three-case structure reflects a fundamental geopolitical tension in AI trade policy. From the loser's perspective, importing frontier AI provides immediate productivity benefits but surrenders the catch-up option and incurs social costs. From the winner's perspective, exporting AI generates revenue and eliminates catch-up risk but erodes the competitive advantage that drives the ecosystem benefit of winning. The equilibrium outcome depends on which side's strategic motives dominate. The feedback to the innovation race then depends on what both countries anticipate at the time of the first-stage R&D decision.

[27] Formally, the effective critical social-cost threshold $K^{\text{crit}}(\tau)$ from Proposition 2 is replaced by a weakly higher threshold under Case (ii) and a strictly higher threshold under Case (iii), expanding the interval in the main condition (21).

The key insight is that the most economically interesting regime—the winner-initiated export ban—creates the largest divergence between private and social innovation incentives, precisely because it combines the strongest geoeconomic motive (keeping the rival behind) with the weakest private motive (serving only the domestic market). In this regime, the social planner's case for subsidizing innovation is strictly stronger than in the baseline, while the private firm's incentive to innovate is strictly weaker. The gap between optimal and laissez-faire innovation policy therefore widens when export controls are endogenous.

5 Conclusion

This paper develops a two-country model of AI innovation policy in which a winner-takes-all technology race interacts with trade in differentiated goods. The central finding is a policy reversal: a technology whose social costs justify a ban under autarky can become worth subsidizing in an open economy. Two channels, absent in the closed economy, drive the reversal. Winning the innovation race strengthens the domestic production ecosystem by attracting more firms home, lowering the cost of living. The private AI firm does not internalize this externality. In addition, because the race is winner-takes-all, the planner internalizes a loss from falling behind, creating a race premium that has no counterpart in the closed economy. A global cooperative planner would still ban the technology, confirming that the national subsidy is purely beggar-thy-neighbor. Endogenizing post-innovation export controls sharpens the result: an export ban lowers the winning firm's profits while raising ecosystem gains, so private and social returns move in opposite directions and the national case for subsidies grows stronger.

Several extensions merit future investigation. Our model features two symmetric countries competing head-to-head. Introducing a third country that does not field a frontier AI firm but consumes AI-intensive goods, as the European Union arguably does relative to the United States and China, raises distinct policy questions. Such a country cannot win the race, yet its welfare depends on who wins and on what terms frontier technology is shared. Whether a non-contender should align with one rival, invest in its own capacity, or focus regulation on deployment rather than development is an open question with direct relevance to the design of frameworks like the EU AI Act. A second direction concerns domestic market structure. We assign a single AI firm to each country. In practice, several firms within the same jurisdiction compete for the frontier, as OpenAI and Anthropic do in the United States. Intra-country competition may dissipate rents and reduce the private return to innovation, altering the welfare-profit wedge that drives our results. It also forces the planner to decide whether to concentrate subsidies on a national champion or spread support across competitors, a choice that interacts with the international rivalry in ways our current framework does not capture.

References

Acemoglu, Daron. 2024. The simple macroeconomics of AI. *NBER Working Paper No. 32487*.

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

—. 2019. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

Aghion, Philippe, Benjamin F Jones, and Charles I Jones. 2018. Artificial intelligence and economic growth. In *The Economics of Artificial Intelligence: An Agenda*, pp. 237–282. University of Chicago Press.

Akcigit, Ufuk, Sina T Ates, and Giammario Impullitti. 2018. Innovation and trade policy in a globalized world. *NBER Working Paper No. 24543*.

Athey, Susan and Fiona Scott Morton. 2025. Artificial intelligence, competition, and welfare. *Working Paper*.

Bai, Yan, Keyu Jin, and Dan Lu. 2025. Technological rivalry and optimal dynamic policy in an open economy. *NBER Working Paper No. 31703*.

Becko, John and Daniel O'Connor. 2025. Strategic (dis)integration. Working Paper.

Brander, James A and Barbara J Spencer. 1985. Export subsidies and international market share rivalry. *Journal of International Economics*, 18(1–2):83–100.

Chatterji, Aaron K and Fiona Murray. 2025. How geopolitics is changing the economics of innovation. *Working Paper*.

Clayton, Christopher, Matteo Maggiori, and Jesse Schreger. 2025. A theory of economic coercion and fragmentation. Working Paper.

—. 2026. A framework for geoeconomics. *Econometrica*, 94(1):105–136.

Costinot, Arnaud and Iván Werning. 2023. Robots, trade, and luddism: A sufficient statistic approach to optimal technology regulation. *Review of Economic Studies*, 90(5):2261–2291.

Eaton, Jonathan and Gene M Grossman. 1986. Optimal trade and industrial policy under oligopoly. *Quarterly Journal of Economics*, 101(2):383–406.

Gans, Joshua S. 2025. Regulating the direction of innovation. *Journal of Public Economics*.

Goldfarb, Avi and Daniel Trefler. 2018. AI and international trade. In *The Economics of Artificial Intelligence: An Agenda*, pp. 463–492. University of Chicago Press.

Guerreiro, Joao, Sergio Rebelo, and Pedro Teles. 2022. Should robots be taxed? *Review of Economic Studies*, 89(1):279–311.

—. 2023. Regulating artificial intelligence. *NBER Working Paper No. 31921*.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Krugman, Paul. 1980. Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70(5):950–959.

—. 1991. Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.

Liu, Jin, Martin Rotemberg, and Sharon Traiberman. 2024. Sabotage as industrial policy. *NBER Working Paper No. 32798*.

Liu, Xueyue, Yu Liu, Alexey Makarin, and Jaya Wen. 2025. Export controls and innovation in sanctioned countries. *Working Paper 25-004*.

Mattoo, Aaditya, Michele Ruta, and Robert W Staiger. 2024. Geopolitics and the world trading system. *Policy Research Working Paper No. 11009, World Bank*.

Nordhaus, William D. 2021. Are we approaching an economic singularity? Information technology and the future of economic growth. *American Economic Journal: Macroeconomics*, 13(1):299–332.

The White House. 2025. Winning the race: America's AI action plan. Technical report, Executive Office of the President, Washington, DC.

Thuemmel, Uwe. 2023. Optimal taxation of robots. *Journal of the European Economic Association*, 21(3):1154–1190.

A Additional Model Details

A.1 Firm Problem and Equilibrium Characterization

We collect here the derivations underlying the equilibrium characterization used in Sections 3 and 4. Throughout, we suppress the time subscript t when no confusion arises.

Demand system. The Home household maximizes $u = a + \beta \log c$ subject to the budget constraint (3), where c is the CES composite (2). Define the CES price index

$$P = \left(\int_0^N \tilde{p}_H(\omega)^{1-\sigma} d\omega + \int_0^{N^*} \tilde{p}_F(\omega)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}.$$

Standard two-stage budgeting yields optimal expenditure $Pc = \beta$ on differentiated goods, with individual demands

$$c_j(\omega) = \left(\frac{\tilde{p}_j(\omega)}{P} \right)^{-\sigma} c.$$

Under symmetry, with $p(\omega) = p$ and $p^*(\omega) = p^*$, delivered prices satisfy $\tilde{p}_H(\omega) = p$ and $\tilde{p}_F(\omega) = \tau p^*$, so

$$P = \left(N p^{1-\sigma} + N^* (\tau p^*)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

Firm optimization. Taking the AI price $\tilde{q}$ and $w = 1$ as given, a Home differentiated firm with production function (5) minimizes cost $\ell(\omega) + \tilde{q} x(\omega)$ subject to the output constraint. The cost-minimizing input ratio satisfies $\tilde{q} x(\omega) / \ell(\omega) = \alpha / (1 - \alpha)$, and marginal cost is

$$mc = \kappa \frac{\tilde{q}^{\alpha}}{A^{\gamma \xi \alpha} \tilde{A}^{\gamma (1-\xi) \alpha}},$$

where $\kappa = (\alpha^{\alpha} (1-\alpha)^{1-\alpha})^{-1}$ and $\tilde{A} = \max\{A, A^*\}$ is the global AI frontier. Under CES demand with elasticity $\sigma > 1$, profit maximization yields the constant markup

$$p(\omega) = \mu \, mc \qquad \text{where} \qquad \mu = \frac{\sigma}{\sigma - 1}.$$

Input demands are proportional to revenue $R(\omega) = p(\omega) y(\omega)$:

$$\tilde{q} \, x(\omega) = \alpha \, mc \, y(\omega) = \frac{\alpha}{\mu} R(\omega),$$

$$\ell(\omega) = (1 - \alpha) \, mc \, y(\omega) = \frac{1 - \alpha}{\mu} R(\omega).$$

Per-firm profit is $\Pi(\omega) = R(\omega) / \sigma - f$.

Free entry. Free entry requires $\Pi(\omega) = 0$, hence $R = \sigma f$ for every active firm in both countries. Each country spends β on differentiated goods. Total world spending is therefore 2β, and world revenue from differentiated production satisfies

$$NR + N^* R^* = 2\beta.$$

Since $R = R^* = \sigma f$, the total mass of varieties is

$$N + N^* = \frac{2\beta}{\sigma f}. \tag{26}$$

Under free entry, input demands simplify to $x = (\alpha/\mu)(\sigma f/\tilde{q})$ and $\ell = ((1-\alpha)/\mu)\sigma f$, both independent of the firm's identity.

Relative prices. Both Home and Foreign firms face the same AI price $\tilde{q}$ and the same global frontier $\tilde{A}$. The ratio of marginal costs therefore depends only on the domestic spillover component:

$$\frac{p}{p^*} = \frac{mc}{mc^*} = \left(\frac{A^*}{A}\right)^{\alpha\gamma\xi}.$$

Define the relative-technology index

$$z = \left(\frac{A}{A^*}\right)^{\vartheta} \qquad \text{where} \qquad \vartheta = \alpha\gamma\xi(\sigma - 1). \tag{27}$$

Then $p^{1-\sigma} = (p^*)^{1-\sigma} z$.

Equilibrium CPI. Define bilateral expenditure shares under symmetry:

$$\pi_{HH} = \frac{Np^{1-\sigma}}{Np^{1-\sigma} + N^*(\tau p^*)^{1-\sigma}} = \frac{Nz}{Nz + N^*\tau^{1-\sigma}},$$
$$\pi_{FH} = \frac{N(\tau p)^{1-\sigma}}{N^*(p^*)^{1-\sigma} + N(\tau p)^{1-\sigma}} = \frac{Nz\tau^{1-\sigma}}{N^* + Nz\tau^{1-\sigma}}.$$

Aggregate Home differentiated revenue is $NR = \beta(\pi_{HH} + \pi_{FH})$, and aggregate Foreign revenue is $N^*R^* = \beta(2 - \pi_{HH} - \pi_{FH})$. Using $R = R^* = \sigma f$ and solving the system $N + N^* = 2\beta/(\sigma f)$ together with $NR = \beta(\pi_{HH} + \pi_{FH})$ yields

$$P^{1-\sigma} = (p^*)^{1-\sigma} \frac{\beta}{\sigma f} z \frac{1 - \tau^{2(1-\sigma)}}{1 - \tau^{1-\sigma} z}. \tag{28}$$

Lemma 11 (Closed-Form Mass of Varieties). *Suppose* $N > 0$, $N^* > 0$, *and* $1 + \tau^{2(1-\sigma)} - 2\tau^{1-\sigma} z > 0$. *Then*

$$N = \frac{\beta}{\sigma f} \frac{z(1+\tau^{2(1-\sigma)}) - 2\tau^{1-\sigma}}{(1-\tau^{1-\sigma}z)(z-\tau^{1-\sigma})},$$

$$N^* = \frac{\beta}{\sigma f} \frac{z(1+\tau^{2(1-\sigma)} - 2z\tau^{1-\sigma})}{(1-\tau^{1-\sigma}z)(z-\tau^{1-\sigma})}.$$

Proof. Free entry implies $N + N^* = 2\beta/(\sigma f)$. The CES price index and relative price identity give $Nz + N^*\tau^{1-\sigma} = (\beta/\sigma f) z(1-\tau^{2(1-\sigma)})/(1-\tau^{1-\sigma}z)$. Solving this two-equation linear system in (N, N^*) delivers the stated expressions. ■

Indirect utility. Recall Home flow utility is $u = a + \beta \log c$. Optimal differentiated expenditure gives $Pc = \beta$ and hence $a = Y - \beta$. With $w = 1$ and free entry ($\Pi = 0$), income is $Y = L + \Pi^{\text{AI}} + T$. Using the government budget constraint (11), $T = -s\ell^{\text{R\&D}} - r$. Defining AI profits gross of R&D as $\pi^{\text{AI}} = \Pi^{\text{AI}} + (1-s)\ell^{\text{R\&D}}$, we have $Y = L + \pi^{\text{AI}} - \ell^{\text{R\&D}} - r$. Therefore

$$u = L + \pi^{\text{AI}} - \ell^{\text{R\&D}} - r - \beta + \beta \log \beta - \beta \log P. \tag{29}$$

The key channel is $-\beta \log P$: a lower CES price index raises welfare. Defining flow utility gross of R&D expenditure as $\tilde{u} = u + \ell^{\text{R\&D}}$, we have $\tilde{u} = L + \pi^{\text{AI}} - r - \beta + \beta \log \beta - \beta \log P$.

CPI expressions in three states. Under the innovation game of Section 2.6, the three relevant states are: baseline (A, A), Home wins $(\bar{A}, A)$, and Foreign wins $(A, \bar{A})$. In all three states, the AI price is $\tilde{q} = 1/A$ (Bertrand competition or symmetric marginal cost). The global frontier is $\tilde{A} = A$ in the baseline and $\tilde{A} = \bar{A}$ in both the winning and losing states. Define

$$\bar{z} = \left(\frac{\bar{A}}{A}\right)^{\vartheta} \quad \text{and} \quad \bar{z}_g = \left(\frac{\bar{A}}{A}\right)^{\alpha\gamma(1-\xi)(\sigma-1)}.$$

The relative-technology index takes values $z = 1$ (baseline), $z = \bar{z}$ (Home wins), $z = \bar{z}^{-1}$ (Foreign wins). Using (28) and the Foreign producer prices in each state, the CPI ratios are

$$\left(\frac{P^w}{P}\right)^{1-\sigma} = \bar{z}_g \bar{z} \frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}},$$

$$\left(\frac{P^\ell}{P}\right)^{1-\sigma} = \bar{z}_g \frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}/\bar{z}},$$

$$\left(\frac{P^w}{P^\ell}\right)^{1-\sigma} = \frac{\bar{z}-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}.$$

The log-CPI gain from winning decomposes as

$$-\beta \log \frac{P^w}{P} = \underbrace{\beta\alpha\gamma \log \frac{\bar{A}}{A}}_{\text{marginal-cost gain}} + \underbrace{\frac{\beta}{\sigma-1} \log \frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}}}_{\text{variety reallocation}}, \tag{30}$$

where the first term combines the domestic and global spillover channels ($\gamma\xi + \gamma(1-\xi) = \gamma$) and the second captures the trade-induced CPI benefit from shifting varieties toward Home.

A.2 AI Demand

Under free entry, each differentiated firm spends the fraction α/μ of its revenue $R = \sigma f$ on AI. World AI demand is

$$Nx + N^*x^* = \frac{\alpha}{\mu}\frac{NR + N^*R^*}{\tilde{q}} = \frac{\alpha}{\mu}\frac{2\beta}{\tilde{q}}.$$

The log-elasticity of world AI demand with respect to $\tilde{q}$ is therefore

$$\frac{\partial \log(Nx + N^*x^*)}{\partial \log \tilde{q}} = -1.$$

With unit-elastic demand, the unconstrained monopoly price tends to infinity, so Bertrand competition between the Home and Foreign AI firms yields limit pricing: the leader prices at the follower's marginal cost. When $A > A^*$,

$$\tilde{q} = \frac{1}{A^*}.$$

A.3 AI Profits

When Home leads with $A > A^*$, the AI price is $\tilde{q} = 1/A^*$ and the Home AI firm's marginal cost is $1/A$. Using the world AI demand from Appendix A.2, flow AI profits gross of R&D are

$$\begin{aligned}\pi^{\text{AI}} &= \left(\tilde{q} - \frac{1}{A}\right)(Nx + N^*x^*) = \left(\frac{1}{A^*} - \frac{1}{A}\right)\frac{\alpha}{\mu}\frac{2\beta}{\tilde{q}} \\ &= \left(\frac{1}{A^*} - \frac{1}{A}\right)\frac{\alpha}{\mu}2\beta A^* = 2\frac{\alpha}{\mu}\beta\left(1 - \frac{A^*}{A}\right).\end{aligned}$$

When Home wins the innovation race, $A = \bar{A}$ and $A^* = A$, so

$$\pi^w = 2\frac{\alpha}{\mu}\beta\left(1 - \frac{A}{\bar{A}}\right).$$

Crucially, π^w is independent of the trade cost τ: world AI spending equals $(\alpha/\mu)\cdot 2\beta$ regardless of τ because world differentiated expenditure is pinned at 2β and AI's cost share α is constant under Cobb–Douglas.

B Proofs and Derivations for Section 3

B.1 CPI Ranking

We prove the ranking $P^w < P$ and $P^w < P^\ell$ under the interiority condition $1 + \tau^{2(1-\sigma)} - 2\bar{z}\tau^{1-\sigma} > 0$ from Section 2.5.

Proof. From Appendix A.1, $(P^w/P)^{1-\sigma} = \bar{z}_g\bar{z}(1-\tau^{1-\sigma})/(1-\tau^{1-\sigma}\bar{z})$. Since $\sigma > 1$, we have $P^w < P$ if and only if $(P^w/P)^{1-\sigma} > 1$. Since $\bar{z}_g \geq 1$,

$$\bar{z}_g\bar{z}\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}} \geq \bar{z}\frac{1-\tau^{1-\sigma}}{1-\tau^{1-\sigma}\bar{z}} > 1,$$

where the last inequality holds because $\bar{z}(1-\tau^{1-\sigma}) > 1-\tau^{1-\sigma}\bar{z}$ simplifies to $(\bar{z}-1)(1-\tau^{1-\sigma}) > 0$, which is immediate from $\bar{z} > 1$ and $\tau^{1-\sigma} \in (0,1)$.

For $P^w < P^\ell$, we use $(P^w/P^\ell)^{1-\sigma} = (\bar{z}-\tau^{1-\sigma})/(1-\tau^{1-\sigma}\bar{z})$. We need $(\bar{z}-\tau^{1-\sigma})/(1-\tau^{1-\sigma}\bar{z}) > 1$, which reduces to $(\bar{z}-1)(1+\tau^{1-\sigma}) > 0$. This holds since $\bar{z} > 1$ and $\tau^{1-\sigma} > 0$. ■

B.2 Proof of Proposition 2

Private best-response function. Under the simplified structure of Section 2.6 (symmetric initial technology, $\pi = 0$, $\mathcal{V}^\ell = 0$), the Home AI firm's HJB (15) with $s = 0$ yields the FOC $e = \mathcal{V}^w - \mathcal{V}$. Substituting back and defining $\tilde{\rho} = \rho + e^*$:

$$\rho\mathcal{V} = \frac{1}{2}(\mathcal{V}^w - \mathcal{V})^2 - e^*\mathcal{V}$$
$$\tilde{\rho}\mathcal{V} = \frac{1}{2}e^2 + e^* \cdot 0 = \frac{1}{2}e^2.$$

Using $\mathcal{V} = \mathcal{V}^w - e$ and $\mathcal{V}^w = \pi^w/\rho$, the quadratic in e is

$$\frac{1}{2}e^2 + \tilde{\rho}\, e = \tilde{\rho}\, \mathcal{V}^w = \frac{\tilde{\rho}}{\rho}\pi^w \equiv \Lambda^{\text{priv}}.$$

The positive root is

$$e^{\text{priv}}(e^*) = -\tilde{\rho} + \sqrt{\tilde{\rho}^2 + 2\Lambda^{\text{priv}}}.$$

Social best-response function. The planner's HJB (17) yields the FOC $e = \mathcal{W}^w - \mathcal{W}$. Substituting and using $\tilde{u} = u + \frac{1}{2}e^2$:

$$\rho\mathcal{W} = \tilde{u} + \frac{1}{2}(\mathcal{W}^w - \mathcal{W})^2 + e^*(\mathcal{W}^\ell - \mathcal{W})$$
$$\tilde{\rho}\,\mathcal{W} = \tilde{u} + \frac{1}{2}e^2 + e^*\mathcal{W}^\ell.$$

Using $\mathcal{W} = \mathcal{W}^w - e$, $\mathcal{W}^w = u^w/\rho$, and $\mathcal{W}^\ell = u^\ell/\rho$, the quadratic in e is

$$\frac{1}{2}e^2 + \tilde{\rho}\, e = \tilde{\rho}\, \mathcal{W}^w - \tilde{u} - e^* \mathcal{W}^\ell \equiv \Lambda^{\text{soc}},$$

with positive root

$$e^{\text{soc}}(e^*) = -\tilde{\rho} + \sqrt{\tilde{\rho}^2 + 2\Lambda^{\text{soc}}}.$$

Comparing effective prizes. Since the best-response functions have the same functional form, $e^{\text{soc}}(e^*) > e^{\text{priv}}(e^*)$ if and only if $\Lambda^{\text{soc}} > \Lambda^{\text{priv}}$. We expand each prize. Using $\mathcal{W}^w - \mathcal{W}^\ell = (u^w - u^\ell)/\rho$ and $\mathcal{V}^w - \mathcal{V}^\ell = \pi^w/\rho$:

$$\Lambda^{\text{soc}} = (u^w - \tilde{u}) + e^*(\mathcal{W}^w - \mathcal{W}^\ell) = (u^w - \tilde{u}) + \theta(u^w - u^\ell),$$
$$\Lambda^{\text{priv}} = \pi^w + e^*(\mathcal{V}^w - \mathcal{V}^\ell) = \pi^w + \theta\pi^w = (1+\theta)\pi^w,$$

where $\theta = e^*/\rho$. Using the indirect utility (29):

$$u^w - \tilde{u} = \pi^w - K - \beta \log \frac{P^w}{P},$$
$$u^w - u^\ell = \pi^w - (r^w - r^\ell) - \beta \log \frac{P^w}{P^\ell}.$$

Since both countries access the frontier under AI trade, $r^w = r^\ell$, so $(r^w - r^\ell) = 0$. Therefore

$$\begin{aligned}\Lambda^{\text{soc}} - \Lambda^{\text{priv}} &= (u^w - \tilde{u} - \pi^w) + \theta(u^w - u^\ell - \pi^w) \\ &= \left(- K - \beta \log \frac{P^w}{P} \right) + \theta \left(- \beta \log \frac{P^w}{P^\ell} \right).\end{aligned}$$

Hence $\Lambda^{\text{soc}} > \Lambda^{\text{priv}}$ if and only if

$$-\beta \log \frac{P^w}{P} - \theta\beta \log \frac{P^w}{P^\ell} > K,$$

which is condition (20). Both terms on the left are positive: the first because $P^w < P$ (Appendix B.1), the second because $P^w < P^\ell$ and $\theta > 0$. ■

Closed-form expression. Substituting the CPI ratios from Appendix A.1:

$$-\beta \log \frac{P^w}{P} = \frac{\beta}{\sigma - 1} \log \left(\bar{z}_g \bar{z} \frac{1 - \tau^{1-\sigma}}{1 - \tau^{1-\sigma}\bar{z}} \right),$$
$$-\beta \log \frac{P^w}{P^\ell} = \frac{\beta}{\sigma - 1} \log \frac{\bar{z} - \tau^{1-\sigma}}{1 - \tau^{1-\sigma}\bar{z}}.$$

The condition $K < K^{\text{crit}}(\tau)$ is therefore equivalent to (20).

B.3 Proof of Proposition 3

Autarky planner's problem. In the closed-economy benchmark of Section 2.7, there is no foreign rival ($e^* = 0$), no winner-takes-all race, and the AI market is disciplined by an open-source fringe at price $1/A$. The domestic spillover collapses to A^γ regardless of ξ, and there is no variety reallocation (all varieties are domestic, $N = \beta/(\sigma f)$). The planner's HJB is

$$\rho \mathcal{W} = \max_{e \geq 0} \left\{ \tilde{u} - \frac{1}{2} e^2 + e(\mathcal{W}^w - \mathcal{W}) \right\}.$$

The FOC is $e = \mathcal{W}^w - \mathcal{W}$, and the quadratic in e is

$$\frac{1}{2} e^2 + \rho e = u^w - \tilde{u} \equiv \Lambda^{\text{aut}}.$$

The planner sets $e > 0$ if and only if $\Lambda^{\text{aut}} > 0$.

Autarky effective prize. Under autarky, AI profits are $\pi^{w,\text{aut}} = (\alpha\beta/\mu)(1 - A/\bar{A})$, since the domestic market size is β (not 2β). The autarky CPI ratio satisfies

$$\left(\frac{P^w}{P} \right)^{1-\sigma} \Big|_{\tau \to \infty} = \left(\frac{\bar{A}}{A} \right)^{\alpha\gamma(\sigma-1)}$$

because the global and domestic spillovers both operate through A^γ under autarky, and there is no variety-reallocation channel. Hence

$$-\beta \log \frac{P^w}{P} \Big|_{\text{aut}} = \beta\alpha\gamma \log \frac{\bar{A}}{A}.$$

Therefore

$$\Lambda^{\text{aut}} = \pi^{w,\text{aut}} - K - \beta \log \frac{P^w}{P} \Big|_{\text{aut}} = \frac{\beta\alpha}{\mu} \left(1 - \frac{A}{\bar{A}} \right) + \beta\alpha\gamma \log \frac{\bar{A}}{A} - K.$$

Threshold classification. Define $\bar{K} = (\beta\alpha/\mu)(1 - A/\bar{A}) + \beta\alpha\gamma \log(\bar{A}/A)$. Then:

(i) If $K \geq \bar{K}$, then $\Lambda^{\text{aut}} \leq 0$ and the planner sets $e = 0$.

(ii)+(iii) If $K < \bar{K}$, then $\Lambda^{\text{aut}} > 0$ and the planner sets $e > 0$.

To separate tax from subsidy, compare with the private optimum. In the closed economy, the private firm's effective prize is $\Lambda^{\text{priv,aut}} = \rho \mathcal{V}^{w,\text{aut}} = \pi^{w,\text{aut}}$. The social prize exceeds the private prize when $\Lambda^{\text{aut}} > \Lambda^{\text{priv,aut}}$, i.e.,

$$-K + \beta\alpha\gamma \log \frac{\bar{A}}{A} > 0 \quad \Longleftrightarrow \quad K < \underline{K} \equiv \beta\alpha\gamma \log \frac{\bar{A}}{A}.$$

Since $\bar{K} = \underline{K} + \pi^{w,\text{aut}} > \underline{K}$ for all $\bar{A} > A$, the three regimes are as stated. ■

B.4 Proof of Proposition 4

Proposition 4 combines the results of Propositions 2 and 3.

Proof. The lower bound $K \geq \bar{K}$ ensures that the autarky planner sets $e = 0$ (Proposition 3, case (i)). The upper bound $K < K^{\text{crit}}(\tau)$ ensures that the social best response exceeds the private best response in the open economy (Proposition 2). These two conditions are compatible if and only if $\bar{K} < K^{\text{crit}}(\tau)$, which holds whenever the trade cost τ is sufficiently low. ■

B.5 Proof of Proposition 5

Strategy. We show that condition (22) implies $K^{\text{crit}}(\tau) > \bar{K}$ for all $\tau > \underline{\tau}(\bar{z})$ in the interiority region.

Autarky limit. As $\tau \to \infty$ (equivalently $\tau^{1-\sigma} \to 0$):

$$K^{\text{crit}}(\tau) \to \frac{\beta}{\sigma - 1} \log \bar{z}_g \bar{z} + \theta \frac{\beta}{\sigma - 1} \log \bar{z} = \beta\alpha\gamma \log \frac{\bar{A}}{A} + \theta\beta\alpha\gamma\xi \log \frac{\bar{A}}{A} = (1 + \theta\xi)\beta\alpha\gamma \log \frac{\bar{A}}{A}.$$

Therefore

$$\lim_{\tau \to \infty} K^{\text{crit}}(\tau) - \bar{K} = \theta\xi\beta\alpha\gamma \log \frac{\bar{A}}{A} - \frac{\beta\alpha}{\mu}\left(1 - \frac{A}{\bar{A}}\right).$$

This is positive if and only if

$$\theta\gamma\xi \log \frac{\bar{A}}{A} > \frac{1}{\mu}\left(1 - \frac{A}{\bar{A}}\right) = \frac{\sigma - 1}{\sigma}\left(1 - \frac{A}{\bar{A}}\right),$$

which is condition (22).

From autarky limit to all τ. When (22) holds, $K^{\text{crit}}(\infty) > \bar{K}$. Since $K^{\text{crit}}(\tau)$ is continuous and strictly decreasing in τ on $(\underline{\tau}, \infty)$, and $\bar{K}$ is independent of τ, we have $K^{\text{crit}}(\tau) > K^{\text{crit}}(\infty) > \bar{K}$ for all $\tau \in (\underline{\tau}, \infty)$. Hence the interval $(\bar{K}, K^{\text{crit}}(\tau))$ is non-empty for every such τ. ■

Monotonicity of $K^{\text{crit}}(\tau)$. For completeness, we verify $\partial K^{\text{crit}}/\partial\tau < 0$. Both log terms in K^{crit} are increasing in $\tau^{1-\sigma}$. Since $d\tau^{1-\sigma}/d\tau = (1-\sigma)\tau^{-\sigma} < 0$, the result follows.

B.6 Proof of Proposition 6

Global planner's problem. The global planner maximizes $\mathcal{W}^G = \mathcal{W} + \mathcal{W}^*$. Because the post-innovation world state is symmetric in aggregate welfare (whether Home or Foreign wins, world

flow utility is $u^w + u^\ell$), the combined arrival rate is $e + e^*$ and the HJB is stated in Section 3.4. The FOCs $e = e^* = \mathcal{W}^{G,\text{post}} - \mathcal{W}^G$ yield equal intensities e^G by the convexity of R&D costs.

Solving the quadratic. Substituting $e = e^* = e^G$ and using $\mathcal{W}^{G,\text{post}} = (u^w + u^\ell)/\rho$:

$$\rho \mathcal{W}^G = 2\tilde{u} - (e^G)^2 + 2e^G(\mathcal{W}^{G,\text{post}} - \mathcal{W}^G) = 2\tilde{u} + (e^G)^2.$$

From the FOC, $\rho e^G = \rho \mathcal{W}^{G,\text{post}} - \rho \mathcal{W}^G = (u^w + u^\ell) - 2\tilde{u} - (e^G)^2$, giving

$$(e^G)^2 + \rho e^G = (u^w + u^\ell) - 2\tilde{u} \equiv \Lambda^G.$$

The unique positive solution is $e^G = (-\rho + \sqrt{\rho^2 + 4\Lambda^G})/2$, which is positive iff $\Lambda^G > 0$.

Decomposition of Λ^G. Using the indirect utility (29) with $\ell^{\text{R\&D}} = 0$ in the post-innovation states:

$$\begin{aligned}\Lambda^G &= (u^w + u^\ell) - 2\tilde{u} \\ &= (\pi^w + \pi^\ell) - (r^w + r^\ell - 2r) - \beta \log \frac{P^w}{P} - \beta \log \frac{P^\ell}{P}.\end{aligned}$$

Since both countries access the frontier, $r^w + r^\ell - 2r = 2K$, and with $\pi^\ell = 0$: $\Lambda^G = \pi^w - 2K - \beta \log(P^w/P) - \beta \log(P^\ell/P)$.

Threshold classification. Define $\bar{K}^G = \Lambda^G|_{K=\bar{K}^G} = 0$, giving $\bar{K}^G = \frac{1}{2}[\pi^w - \beta \log(P^w/P) - \beta \log(P^\ell/P)]$.

To separate tax from subsidy, compare with the private best response evaluated at $e^* = e^G$. The private effective prize at this point is $\Lambda^{\text{priv}}|_{e^*=e^G} = (1 + e^G/\rho)\pi^w$. The cooperative planner wants more innovation than the private market when $\Lambda^G/2 > \Lambda^{\text{priv}}|_{e^*=e^G}$. Note that the cooperative and private problems have different quadratic structures—$(e^G)^2 + \rho e^G = \Lambda^G$ versus $\frac{1}{2}(e^{\text{priv}})^2 + (\rho + e^G)e^{\text{priv}} = \Lambda^{\text{priv}}$—but since the positive root is strictly increasing in the effective prize for each quadratic, the comparison reduces to a prize comparison.

The global subsidy threshold $\underline{K}^G$ is the displacement cost at which Λ^G equals the critical value separating the subsidy and tax regimes. Setting the net CPI externality equal to zero: $\underline{K}^G = \frac{1}{2}[-\beta \log(P^w/P) - \beta \log(P^\ell/P)]$. Since $\bar{K}^G - \underline{K}^G = \pi^w/2 > 0$, the tax regime always exists for $\bar{A} > A$.

The three regimes follow: (i) ban when $K \geq \bar{K}^G$ ($\Lambda^G \leq 0$), (ii) tax when $\underline{K}^G \leq K < \bar{K}^G$, (iii) subsidize when $K < \underline{K}^G$. ■

B.7 Proof of Proposition 7

Strategy. We show that the national planner's best response exceeds the cooperative intensity at any common rival intensity, and then invoke the fixed-point structure to obtain the equilibrium comparison.

Best-response comparison. At any symmetric profile $e = e^* = \hat{e}$, the national planner's effective social prize is

$$\Lambda^{\text{soc}}(\hat{e}) = (u^w - \tilde{u}) + \frac{\hat{e}}{\rho}(u^w - u^\ell).$$

The national planner's quadratic is $\frac{1}{2}e^2 + (\rho + \hat{e})e = \Lambda^{\text{soc}}(\hat{e})$, while the cooperative quadratic (per country) is $(e^G)^2 + \rho e^G = \Lambda^G$. The national planner's effective prize exceeds the cooperative per-country prize for two reasons.

First, $\Lambda^{\text{soc}}(\hat{e})$ values the full winner's CPI gain $-\beta \log(P^w/P)$, including the variety-reallocation component, while $\Lambda^G/2$ values half the net world CPI effect, which is much smaller (Appendix B.6). Second, $\Lambda^{\text{soc}}(\hat{e})$ internalizes only K in displacement costs, while Λ^G internalizes $2K$.

Additionally, the national planner faces a higher effective discount rate $\rho + \hat{e} > \rho$, which for any given prize level raises the best-response intensity: the urgency of innovating before the rival increases the equilibrium effort.

To translate the best-response comparison into an equilibrium comparison, evaluate at the cooperative optimum: $e^{\text{soc}}(e^G) > e^G$ because the national prize exceeds the cooperative prize and the higher discount rate further raises the response. Since both best-response functions are continuous and have slopes in $[0, 1)$ under strategic complementarity, the national Nash equilibrium satisfies $e^N > e^G$. ■

Net world CPI decomposition. For reference, the net world CPI effect from (23) decomposes as

$$-\beta \log \frac{P^w}{P} - \beta \log \frac{P^\ell}{P} = \beta\alpha\gamma(2 - \xi) \log \frac{\bar{A}}{A} + \frac{\beta}{\sigma - 1} \log \frac{(1 - \tau^{1-\sigma})^2}{(1 - \tau^{1-\sigma}\bar{z})(1 - \tau^{1-\sigma}/\bar{z})}.$$

The reallocation residual is strictly positive since $\bar{z} > 1$ implies $(1 - \tau^{1-\sigma}\bar{z})(1 - \tau^{1-\sigma}/\bar{z}) < (1 - \tau^{1-\sigma})^2$, but it is second-order in $\tau^{1-\sigma}$.

C Proofs and Derivations for Section 4

C.1 Derivation of the Winner's Reservation Fee

The winner's reservation fee T^{res} is defined as the minimum per-period fee that makes allowing AI exports weakly preferred to an export ban:

$$T^{\text{res}} = \rho(\mathcal{W}^{w,\text{aut}} - \mathcal{W}^{w,\text{trade}}(0)).$$

Winner's trade continuation value. Under AI trade at fee T (terminal state), the winner's continuation value is

$$\mathcal{W}^{w,\text{trade}}(T) = \frac{u^{w,\text{trade}} + T}{\rho}.$$

At $T = 0$: $\mathcal{W}^{w,\text{trade}}(0) = u^{w,\text{trade}}/\rho$.

Winner's autarky continuation value. Under AI autarky, the winner does not invest in R&D but faces catch-up risk at rate $\widehat{e}^{*,\text{soc}}$. The winner's HJB is

$$\rho\mathcal{W}^{w,\text{aut}} = u^{w,\text{aut}} + \widehat{e}^{*,\text{soc}}(\mathcal{W}^{w,\text{catch}} - \mathcal{W}^{w,\text{aut}}).$$

Solving with $\mathcal{W}^{w,\text{catch}} = u^{w,\text{catch}}/\rho$:

$$\mathcal{W}^{w,\text{aut}} = \frac{u^{w,\text{aut}} + (\widehat{e}^{*,\text{soc}}/\rho)u^{w,\text{catch}}}{\rho + \widehat{e}^{*,\text{soc}}}.$$

Computing T^{res}.

$$\begin{aligned} T^{\text{res}} &= \rho\mathcal{W}^{w,\text{aut}} - u^{w,\text{trade}} \\ &= \frac{\rho u^{w,\text{aut}} + \widehat{e}^{*,\text{soc}}u^{w,\text{catch}}}{\rho + \widehat{e}^{*,\text{soc}}} - u^{w,\text{trade}} \\ &= \frac{\rho(u^{w,\text{aut}} - u^{w,\text{trade}}) + \widehat{e}^{*,\text{soc}}(u^{w,\text{catch}} - u^{w,\text{trade}})}{\rho + \widehat{e}^{*,\text{soc}}}. \end{aligned}$$

Now we compute each difference using the expressions from Appendix A.1.

Profit difference. $\pi^{w,\text{trade}} - \pi^{w,\text{aut}} = (\alpha\beta/\mu)(1 - A/\bar{A})$, since $\pi^{w,\text{trade}} = 2\pi^{w,\text{aut}}$.

CPI difference (autarky vs. trade for the winner). Under autarky, the effective productivity gap is z^{aut}, while under trade it is $z^{\text{trade}} = \bar{z}$. Since $z^{\text{aut}} > z^{\text{trade}}$, the winner's CPI is lower under autarky.

Both states incur the same displacement cost $r(\bar{A})$. Therefore

$$u^{w,\text{aut}} - u^{w,\text{trade}} = -\frac{\alpha\beta}{\mu}\left(1 - \frac{A}{\bar{A}}\right) + \frac{\beta}{\sigma - 1}\log\frac{1 - \tau^{1-\sigma}z^{\text{trade}}}{1 - \tau^{1-\sigma}z^{\text{aut}}}.$$

CPI difference (catch-up vs. trade for the winner). After catch-up, technologies are symmetric and the relative-productivity term returns to 1, so

$$u^{w,\text{catch}} - u^{w,\text{trade}} = -\frac{\alpha\beta}{\mu}\left(1 - \frac{A}{\bar{A}}\right) + \frac{\beta}{\sigma - 1}\log\frac{1 - \tau^{1-\sigma}}{1 - \tau^{1-\sigma}z^{\text{trade}}}.$$

Substituting into the T^{res} formula:

$$\begin{aligned} T^{\text{res}} = &-\frac{\alpha\beta}{\mu}\left(1 - \frac{A}{\bar{A}}\right) + \frac{\beta\rho}{(\rho + \widehat{e}^{*,\text{soc}})(\sigma - 1)}\log\frac{1 - \tau^{1-\sigma}z^{\text{trade}}}{1 - \tau^{1-\sigma}z^{\text{aut}}} \\ &- \frac{\beta\widehat{e}^{*,\text{soc}}}{(\rho + \widehat{e}^{*,\text{soc}})(\sigma - 1)}\log\frac{1 - \tau^{1-\sigma}}{1 - \tau^{1-\sigma}z^{\text{trade}}}, \end{aligned}$$

as stated in Section 4.2. The three terms have the signs described in the main text: the first is strictly negative (forgone export profit), the second is strictly positive (static strategic gain from keeping the rival behind), and the third is weakly negative (dynamic benefit of eliminating catch-up risk). ■

C.2 Proof of Proposition 8

Loser's comparison. The loser accepts AI trade at fee T if and only if $\mathcal{W}^{*,\text{trade}}(T) \geq \mathcal{W}^{*,\text{aut}}$. Since $\mathcal{W}^{*,\text{trade}}(T) = (u^{*,\text{trade}} - T)/\rho$, the maximum acceptable fee is

$$T^{\max} = u^{*,\text{trade}} - \rho\mathcal{W}^{*,\text{aut}}.$$

Loser's autarky value. Under AI autarky, the loser solves a catch-up innovation problem. The loser's HJB is

$$\rho\mathcal{W}^{*,\text{aut}} = \max_{e^{*,\text{soc}} \geq 0}\left\{\tilde{u}^{*,\text{aut}} - \frac{1}{2}(e^{*,\text{soc}})^2 + e^{*,\text{soc}}(\mathcal{W}^{*,\text{catch}} - \mathcal{W}^{*,\text{aut}})\right\},$$

where $\mathcal{W}^{*,\text{catch}} = u^{*,\text{catch}}/\rho$. The FOC gives $\widehat{e}^{*,\text{soc}} = \mathcal{W}^{*,\text{catch}} - \mathcal{W}^{*,\text{aut}}$, and substituting back:

$$\rho\mathcal{W}^{*,\text{aut}} = \tilde{u}^{*,\text{aut}} + \frac{1}{2}(\widehat{e}^{*,\text{soc}})^2.$$

The optimal catch-up intensity solves $\frac{1}{2}(\widehat{e}^{*,\text{soc}})^2 + \rho\widehat{e}^{*,\text{soc}} = (\Delta^{*,\text{catch}})_+$, yielding $\widehat{e}^{*,\text{soc}} = -\rho + \sqrt{\rho^2 + 2(\Delta^{*,\text{catch}})_+}$ with $\Delta^{*,\text{catch}}$ as defined in Section 4.1.

Static gap. The flow-utility comparison between trade and pre-catch-up autarky is

$$u^{*,\text{trade}} - \tilde{u}^{*,\text{aut}} = -K + \frac{\beta}{\sigma - 1} \log \frac{z^{\text{aut}} - \tau^{1-\sigma}}{z^{\text{trade}} - \tau^{1-\sigma}}.$$

The log term is strictly positive because $z^{\text{aut}} > z^{\text{trade}}$: it is the loser's static CPI gain from frontier access. The term $-K$ is weakly negative: the displacement-cost saving from staying on lagging AI.

Assembling $T^{\max}$.

$$T^{\max} = u^{*,\text{trade}} - \tilde{u}^{*,\text{aut}} - \frac{1}{2}(\hat{e}^{*,\text{soc}})^2 = \frac{\beta}{\sigma - 1} \log \frac{z^{\text{aut}} - \tau^{1-\sigma}}{z^{\text{trade}} - \tau^{1-\sigma}} - K - \frac{1}{2}(\hat{e}^{*,\text{soc}})^2,$$

as stated in Proposition 8. ■

C.3 Proof of Proposition 9

Proof. The equilibrium follows from the sequential structure of the geopolitical AI trade game. The loser accepts fee T if and only if $T \leq T^{\max}$. The winner weakly prefers trade at fee T to an export ban if and only if $T \geq T^{\text{res}}$.

Case (i). If $T^{\max} < 0$, no $T \geq 0$ satisfies the loser's acceptance condition. AI autarky prevails.

Case (ii). If $T^{\max} \geq 0$ and $T^{\max} \geq T^{\text{res}}$, the winner optimally charges $T = T^{\max}$ (extracting the full surplus), the loser accepts at indifference, and AI trade occurs.

Case (iii). If $T^{\max} \geq 0$ but $T^{\max} < T^{\text{res}}$, every fee acceptable to the loser ($T \leq T^{\max}$) fails to meet the winner's reservation ($T \geq T^{\text{res}}$). The winner bans exports. ■

C.4 Proof of Proposition 10

We derive the feedback of each second-stage case to first-stage innovation incentives. Throughout, e^* denotes the Foreign first-stage intensity and $s = 0$ for the private comparison. We denote the baseline (Section 3) effective prizes by $\Lambda^{\text{priv}}_{\text{baseline}}$ and $\Lambda^{\text{soc}}_{\text{baseline}}$.

Private continuation values. Under AI trade (baseline): $\mathcal{V}^{w,\text{trade}} = \pi^{w,\text{trade}}/\rho = 2(\alpha\beta/\mu)(1 - A/\bar{A})/\rho$ and $\mathcal{V}^{\ell,\text{trade}} = 0$.

Under AI autarky: $\mathcal{V}^{w,\text{aut}} = \pi^{w,\text{aut}}/\rho = \mathcal{V}^{w,\text{trade}}/2$ and $\mathcal{V}^{\ell,\text{aut}} = (\hat{e}^p)^2/(2\rho) > 0$, where $\hat{e}^p$ is the private catch-up intensity. The loser's value follows from the linear-quadratic catch-up problem: the loser earns zero AI flow profit but can invest to reach $\bar{A}$, earning $\pi^{w,\text{aut}}/\rho$ upon success.

Social continuation values. Under trade at fee T: $\mathcal{W}^{w,\text{trade}}(T) = (u^{w,\text{trade}} + T)/\rho$ and $\mathcal{W}^{\ell,\text{trade}}(T) = (u^{\ell,\text{trade}} - T)/\rho$.

Under autarky: $\mathcal{W}^{w,\text{aut}}$ and $\mathcal{W}^{\ell,\text{aut}}$ are as derived in Appendices C.1 and C.2.

Case (i): Loser-initiated autarky ($T^{\max} < 0$). The private effective prize changes to

$$\Lambda^{\text{priv}}_{\text{aut}} = (\rho + e^*)\mathcal{V}^{w,\text{aut}} - e^*\mathcal{V}^{\ell,\text{aut}}.$$

Since $\mathcal{V}^{w,\text{aut}} = \mathcal{V}^{w,\text{trade}}/2$ and $\mathcal{V}^{\ell,\text{aut}} > 0 = \mathcal{V}^{\ell,\text{trade}}$:

$$\Lambda^{\text{priv}}_{\text{aut}} - \Lambda^{\text{priv}}_{\text{baseline}} = -\frac{\rho + e^*}{\rho}\frac{\alpha\beta}{\mu}\left(1 - \frac{A}{\bar{A}}\right) - \frac{e^*}{2\rho}(\hat{e}^p)^2 < 0.$$

The private best response falls.

For the planner, the change in the social effective prize is

$$\Lambda^{\text{soc}}_{\text{aut}} - \Lambda^{\text{soc}}_{\text{baseline}} = \frac{\rho + e^*}{\rho}T^{\text{res}} + \frac{e^*}{\rho}T^{\max}.$$

Since $T^{\max} < 0$ in Case (i), the sign depends on whether the winner-side gain T^{res} outweighs the loser-side loss. The planner's best response rises iff $T^{\text{res}} > \frac{e^*}{\rho+e^*}(-T^{\max})$.

Case (ii): AI trade with fee extraction ($T^{\max} \geq 0$, $T^{\max} \geq T^{\text{res}}$). The fee $T^{\max}$ is a sovereign transfer, so private continuation values are unchanged: $\Lambda^{\text{priv}}_{\text{trade}} = \Lambda^{\text{priv}}_{\text{baseline}}$. For the planner, winning now brings $T^{\max}$ and losing costs $T^{\max}$:

$$\Lambda^{\text{soc}}_{\text{trade}} - \Lambda^{\text{soc}}_{\text{baseline}} = (\rho + e^*)\frac{T^{\max}}{\rho} + e^*\frac{T^{\max}}{\rho} = \frac{\rho + 2e^*}{\rho}T^{\max} \geq 0,$$

as stated in Proposition 10(ii). The social best response weakly rises.

Case (iii): Winner-initiated export ban ($T^{\max} \geq 0$, $T^{\max} < T^{\text{res}}$). The private effective prize is the same as in Case (i): $\Lambda^{\text{priv}}_{\text{ban}} = \Lambda^{\text{priv}}_{\text{aut}} < \Lambda^{\text{priv}}_{\text{baseline}}$. For the planner:

$$\Lambda^{\text{soc}}_{\text{ban}} - \Lambda^{\text{soc}}_{\text{baseline}} = \frac{\rho + e^*}{\rho}T^{\text{res}} + \frac{e^*}{\rho}T^{\max}.$$

Since $T^{\text{res}} > T^{\max} \geq 0$ in Case (iii), both terms on the right are positive, so $\Lambda^{\text{soc}}_{\text{ban}} > \Lambda^{\text{soc}}_{\text{baseline}}$. Moreover:

$$\frac{\rho + e^*}{\rho}T^{\text{res}} + \frac{e^*}{\rho}T^{\max} > \frac{\rho + e^*}{\rho}T^{\max} + \frac{e^*}{\rho}T^{\max} = \frac{\rho + 2e^*}{\rho}T^{\max} \geq 0.$$

The welfare-versus-profit wedge strictly widens:

$$(\Lambda^{\text{soc}}_{\text{ban}} - \Lambda^{\text{priv}}_{\text{ban}}) - (\Lambda^{\text{soc}}_{\text{baseline}} - \Lambda^{\text{priv}}_{\text{baseline}}) = (\Lambda^{\text{soc}}_{\text{ban}} - \Lambda^{\text{soc}}_{\text{baseline}}) - (\Lambda^{\text{priv}}_{\text{ban}} - \Lambda^{\text{priv}}_{\text{baseline}}) > 0,$$

since the first difference is strictly positive and the second is strictly negative. This establishes Proposition 10(iii). ■