# Artificial Intelligence, Strategic Homogenization, and the Misalignment of Adoption

Feng Chi* Nathan Yang†

May 1, 2026

**Abstract**

Enterprise artificial intelligence is increasingly delivered through shared infrastructure such as foundation models and common APIs. When firms rely on a common processing layer, their actions inherit a shared error component that is uninformative about fundamentals but induces cross-sectional correlation — a signal structure that standard coordination-information frameworks do not produce. This paper develops a Bayesian game in which artificial intelligence reduces idiosyncratic noise while introducing such a shared shock into adopters' signals. The model yields a closed-form equilibrium, a misalignment result in which adoption is excessive regardless of strategic-interaction sign, and a four-regime typology including a coordination trap.

*Keywords:* Artificial Intelligence; Beauty Contest; Coordination; Attention-Based View; Technology Adoption; Information Design

*University of Illinois Urbana-Champaign, Department of Economics. Email: fchi@illinois.edu.
†University of Illinois Urbana-Champaign, Gies College of Business. Email: ncyang@illinois.edu.

# 1 Introduction

In almost every industry, firms now face decisions about whether and how to integrate artificial intelligence into their decision-making processes. Hiring, lending, pricing, inventory management, and product recommendation systems increasingly rely on shared infrastructure: foundation models accessed through common APIs, cloud-hosted scoring systems, and third-party recommendation engines. A striking feature of contemporary enterprise artificial intelligence is its concentration. The same small set of frontier models now mediates the analytics of large numbers of firms in finance, retail, healthcare, and beyond. As a result, decisions that appear decentralized may in fact be routed through a common processing layer.

Recent empirical evidence highlights the economic importance of these technologies. Firms investing in artificial intelligence experience faster growth, driven primarily by increased product innovation (Babina et al. 2024a); deployment of generative AI tools at the worker level produces measurable gains in productivity and output quality (Brynjolfsson et al. 2025); AI adoption is now sufficiently widespread to be measurable across the firm size distribution (Acemoglu et al. 2022); and AI investment is associated with increases in firms' systematic risk that are not observed for other technology categories (Babina et al. 2024b). This makes understanding the equilibrium consequences of widespread reliance on shared AI systems a central economic question, and one that financial regulators have independently identified as a source of systemic risk (Bank of England Financial Policy Committee 2025).[1]

Recent work on algorithmic decision-making provides complementary evidence that such ho-

[1] The Bank of England's Financial Policy Committee notes that "[f]rom a systemic risk perspective, the potential for AI-based participants to take increasingly correlated positions is an important consideration. This could be driven by the widespread use of a small number of open-source or vendor-provided models or underlying data sets, or a more general convergence on very similar model designs across the market." See *Financial Stability in Focus: Artificial intelligence in the financial system*, April 2025.

mogenization can arise from shared computational systems. Algorithmic advice can function as a strategic signal that coordinates behavior across agents even in the absence of direct communication (Rebholz et al. 2025), and experimental and simulation evidence shows that AI-driven decision-makers can generate sustained supra-competitive outcomes through tacit coordination in standard market environments (Deshpande and Jacobson 2026). A related line of work studies algorithmic monoculture, in which multiple decision-makers rely on a common model or shared inputs, leading to correlated evaluations and potentially systematic distortions in outcomes (Kleinberg and Raghavan 2021, Peng and Garg 2024). More broadly, the use of shared components such as training data and foundation models can induce homogeneous outcomes across decision-makers, even when they operate independently (Bommasani et al. 2021, 2022). These findings suggest that when firms rely on common algorithmic systems, correlation in behavior need not reflect shared information or institutional pressures, but can instead emerge from shared processing itself.

This paper takes that claim to a setting in which the channel is not institutional or cognitive but technological. When firms process their data through a shared artificial intelligence system, they standardize not only what information they use but how that information is translated into decisions. The resulting correlation has a distinct structure: firms' actions inherit a common component that reflects the shared processing layer rather than the underlying fundamental. The economic question is then immediate. When homogenization arises from shared digital infrastructure rather than from social or organizational forces, what are its equilibrium and welfare consequences?

Existing economic models of information in coordination settings have, to our knowledge, not previously studied a signal structure in which the common component is orthogonal to the fundamental. In the canonical framework of Morris and Shin (2002) and its extensions (Angeletos and Pavan 2007, Hellwig and Veldkamp 2009, Myatt and Wallace 2012), firms receive signals

about a common fundamental and choose actions that depend both on that fundamental and on others' actions.[2] Correlation arises either from common fundamentals or from public signals, and these sources of correlation are informative: even when public signals induce coordination, they convey information about the underlying state, so that aggregating across firms refines beliefs. The resulting trade-off between coordination and information generates the non-monotone comparative statics that organize this literature.

Shared artificial intelligence infrastructure violates this structure. When firms feed their data through a common processing model, their signals contain a component that is correlated across adopters but uninformative about the fundamental. Setting any precision parameter to any value in the standard frameworks does not yield a signal of this kind: orthogonality to the fundamental is a feature of the processing technology rather than a limit of an information-quality parameter. The implication is that the offsetting force which generates the non-monotone comparative statics in prior work — that more correlation also brings more information — is structurally absent here. As a result, greater reliance on the shared signal monotonically inflates aggregate volatility, and the welfare wedge has a determinate sign that does not depend on whether actions are strategic complements or substitutes.

This structural difference connects to several adjacent literatures without being subsumed by them. In models of endogenous information acquisition, agents coordinate on similar signals and equilibrium behavior can exhibit substantial comovement (Veldkamp 2006); improvements in data-processing technology can shift agents toward extracting information from others' trades rather than producing new information (Farboodi and Veldkamp 2020). Correlated fluctuations need not reflect fundamentals: endogenous information acquisition and imperfect aggregation can generate

[2] For a survey of this literature, see Pavan et al. (2015).

non-fundamental volatility even in equilibrium (Hébert and La'O 2023, Lorenzoni 2009), and dispersed beliefs can drive aggregate fluctuations through sentiment-like components (Angeletos and La'o 2013). In rational-inattention environments, agents optimally choose what information to process, and the resulting attention allocation can produce correlated behavior or muted aggregate responses (Sims 2003, Matějka and McKay 2015). In each of these settings, correlation remains tied to information — about $\theta$, about prices, or about other agents' actions. The mechanism here is different: correlation arises from a common transformation of underlying data, with no informational link to the fundamental.

The decision problem facing a firm is therefore different from the standard information-acquisition problem. A firm contemplating artificial intelligence adoption is choosing a processing technology rather than a signal. The benefit is improved idiosyncratic precision, as shared infrastructure typically outperforms bespoke systems. The cost is exposure to a common processing shock that is shared with other adopters. Whether this trade is privately attractive depends on the firm's coordination motive. Whether it is socially desirable depends on whether firms internalize the effect of synchronized errors on aggregate outcomes. Modern artificial intelligence provides a natural empirical instance of this structure: foundation models and shared APIs implement a common mapping from data to decisions across firms (Veldkamp and Chung 2024), making the induced correlation a property of the processing function itself rather than of the underlying information.

This paper develops a tractable Bayesian game that isolates this mechanism. Artificial intelligence is modeled as a shared processing technology that reduces idiosyncratic noise while introducing an uninformative common shock into adopters' signals. The model admits a closed-form linear equilibrium under heterogeneous adoption and allows the private return to adoption to be characterized as a function of penetration. The central result is a misalignment theorem:

for any value of the strategic-interaction parameter, equilibrium adoption exceeds the level chosen by a planner who internalizes the systemic-risk externality generated by the shared processing layer. The invariance follows from a single structural feature — the orthogonality of the shared shock to the fundamental — and is not obtained by adding a comparable externality to the canonical Morris-Shin or Angeletos-Pavan settings, where the welfare direction depends on parameter regions.

The model also delivers a four-regime typology of adoption dynamics. When strategic complementarities are strong and the common shock is sufficiently large, the economy exhibits a coordination trap: both low- and high-adoption equilibria are stable, even though artificial intelligence is intrinsically unattractive in isolation. The high-adoption equilibrium is sustained purely by coordination motives. A phase diagram organizes these regimes and provides a unified view of how technological parameters and strategic structure jointly determine adoption outcomes.

The framework yields three empirical signatures that distinguish the technological channel from adjacent mechanisms. First, among adopters, actions contain a common component that is absent among non-adopters, generating excess covariance proportional to $(k^A)^2\sigma_\xi^2$ that the institutional-isomorphism and cognitive-map literatures cannot generate at high frequency. Second, aggregate volatility rises with adoption through $K_\xi(q)^2\sigma_\xi^2$ without a corresponding improvement in fundamental tracking, distinguishing the mechanism from rational-inattention and information-acquisition channels in which any rise in correlation is accompanied by gains in informational efficiency at some other margin. Third, when the underlying foundation model is updated, the realization of $\xi$ shifts discretely, so the model predicts that adopter behavior should comove with model-update events even when fundamentals are stable. These signatures point to empirical strategies that isolate residual comovement after controlling for observable fundamentals and relate

it to measures of shared AI usage.

The mechanism connects naturally to recent evidence that algorithmic decision systems can induce coordinated behavior and supra-competitive outcomes in decentralized environments (Calvano et al. 2020, Assad et al. 2024). Misalignment in the model arises not from informational frictions in the usual sense, but from the structure of the decision technology itself: each additional adopter increases the contribution of the shared component to aggregate outcomes, while firms' payoffs depend on coordination with the aggregate rather than on its variance.

# 2 Model

This section formalizes the mechanism described in the introduction. The model isolates the consequences of adopting a shared processing technology in an otherwise standard coordination environment. Firms choose actions that balance accuracy with coordination, but differ in how they process information about the underlying fundamental. Artificial intelligence is modeled as a common processing layer that alters the structure of signals by introducing a shared component across adopters. Adoption is endogenous, determined by firms' trade-off between precision and exposure to this shared component, and beliefs are formed using Bayesian filtering given the information structure. The environment, information technologies, adoption decision, and filtering problem are laid out in turn.

## 2.1 Environment

A unit continuum of firms $f \in [0,1]$ chooses actions $x_f \in \mathbb{R}$ to maximize the quadratic beauty-contest payoff

$$\pi_f = -(x_f - \theta)^2 - \beta(x_f - \bar{x})^2,$$

where $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$ is an unobserved fundamental and $\bar{x} \equiv \int_0^1 x_g \, dg$ is the average action. The parameter $\beta \in (-1, \infty)$ governs the strategic interaction: $\beta > 0$ corresponds to strategic complementarity, while $\beta < 0$ corresponds to strategic substitutability. The restriction $\beta > -1$ ensures existence of a unique linear equilibrium. The payoff trades off accuracy and coordination. Firms care about matching the fundamental, but also about aligning with the behavior of others. This structure captures a wide range of environments in reduced form. Pricing decisions with demand spillovers, capacity choices under congestion, R&D direction in settings with knowledge spillovers, and portfolio allocations in markets where prices reflect aggregate behavior all map into this formulation with appropriately signed $\beta$. Throughout, $\beta$ is treated as a primitive feature of the environment rather than an object of choice.

## 2.2 Information and Processing Technologies

Firms differ in how they process information about $\theta$. A firm using a custom technology observes

$$\tilde{s}_f^C = \theta + u_f^C,$$

where $u_f^C \sim \mathcal{N}(0, \sigma_C^2)$ is idiosyncratic noise. A firm using artificial intelligence observes

$$\tilde{s}_f^A = \theta + \xi + u_f^A,$$

where $u_f^A \sim \mathcal{N}(0, \sigma_A^2)$ is idiosyncratic noise and $\xi \sim \mathcal{N}(0, \sigma_\xi^2)$ is a common processing shock shared by all AI adopters. All shocks are mutually independent and independent of $\theta$.

This specification provides a reduced-form representation of shared algorithmic decision-making. A growing literature studies environments in which multiple decision-makers rely on a common algorithm or shared model, generating correlated evaluations across agents. In screening and matching settings, the use of a common ranking or score induces dependence across firms' decisions even when each firm would prefer to rely on a more accurate assessment (Kleinberg and Raghavan 2021, Peng and Garg 2024). More broadly, modern machine learning systems are built around shared components, such as training data and foundation models, which can induce systematic similarity in outcomes across decision-makers (Bommasani et al. 2021, 2022).

The formulation $\tilde{s}_f^A = \theta + \xi + u_f^A$ captures this idea in the simplest possible way. The term $\xi$ represents the component of the signal that is common across all adopters, arising from the shared processing layer. It can be interpreted as the error induced by a common model, a shared training dataset, or any feature of the algorithm that affects all users simultaneously. This component is not under the control of individual firms and is not diversifiable across them. By contrast, $u_f^A$ captures residual idiosyncratic variation, reflecting firm-specific data, implementation, or noise. In this sense, the specification isolates a setting in which signals contain a common component that is orthogonal to fundamentals but nonetheless affects equilibrium behavior through expectations and coordination (Hellwig 2005).

The distinction between the two technologies is therefore not only a difference in precision but a difference in structure. Artificial intelligence may reduce idiosyncratic noise, often corresponding to $\sigma_A^2 < \sigma_C^2$, but it does so by introducing a shared component into the signal. This feature mirrors the logic of algorithmic monoculture: when multiple agents rely on the same system, their errors become correlated because they are driven by a common source. Custom firms, by contrast, retain purely idiosyncratic noise and never observe $\xi$, so their signals remain independent across the cross-section.

This structure departs in a fundamental way from the public-signal benchmark of Morris and Shin (2002). A public signal introduces correlation because it is informative about $\theta$, so that aggregation improves inference even as actions become more aligned. More generally, even when signals confound different components of uncertainty, correlation typically remains tied to economically relevant shocks and can amplify aggregate volatility (Bergemann et al. 2015). Here the correlation arises from a source that is orthogonal to the fundamental. The shared component therefore survives aggregation without contributing any informational content. The cross-section of firms ceases to refine beliefs about $\theta$ and instead carries a common error inherited from the processing technology itself.

**Remark 1** (Behavioral interpretation)**.** The shared component $\xi$ admits a behavioral interpretation that may be useful for readers familiar with the sentiment literature in finance. When many firms adopt the same artificial-intelligence system for a given decision, it is as if each firm has hired the same employee, whose idiosyncratic biases, training, and blind spots generate correlated outputs across employers. The variance $\sigma_\xi^2$ measures the magnitude of this shared employee's biases, while the orthogonality of $\xi$ to $\theta$ captures that these biases are unrelated to the underlying state. This

structure parallels the treatment of investor sentiment as a common factor orthogonal to fundamentals (Baker and Wurgler 2006, 2007), and the role of correlated belief errors among noise traders in generating non-fundamental comovement (De Long et al. 1990). The mechanism here differs in source rather than form: the shared component arises from a common decision technology rather than from correlated psychology across agents.

## 2.3 Adoption Stage

Each firm draws a switching cost $c_f \sim F$ independently, where $F$ is a continuous cumulative distribution function on $\mathbb{R}$ with positive density on its support. The firm chooses its processing technology to maximize expected payoff net of the switching cost, taking the equilibrium adoption rate $q$ as given. An adoption equilibrium is a rational-expectations fixed point in which $q$ is consistent with individual choices.

For expositional convenience, it is useful to allow $F$ to have bounded support $[0, \bar{c}]$ when discussing corner outcomes (Section 5). In that case, if $\Delta\Pi(0) \leq 0$ then $F(\Delta\Pi(0)) = 0$, and zero adoption can arise as an equilibrium. When $F$ has unbounded support, as with the logistic distribution used in the numerical calibration of Figure 3, the corresponding corner outcomes are replaced by interior equilibria with $q^*$ close to but bounded away from 0 or 1. This distinction does not affect any of the qualitative results that follow.

## 2.4 Bayesian Filtering

Firms form beliefs using Bayesian-optimal filtering given their information structure and common knowledge of the data-generating process, including the presence and variance of the common shock $\xi$. Let $V_A \equiv \sigma_\theta^2 + \sigma_\xi^2 + \sigma_A^2$ and $V_C \equiv \sigma_\theta^2 + \sigma_C^2$ denote unconditional signal variances, and

define the projection coefficients

$$\alpha_\theta^A \equiv \frac{\sigma_\theta^2}{V_A}, \qquad \alpha_\xi^A \equiv \frac{\sigma_\xi^2}{V_A}, \qquad \alpha_\theta^C \equiv \frac{\sigma_\theta^2}{V_C}.$$

Then $\mathbb{E}_f^A[\theta] = \alpha_\theta^A \tilde{s}_f^A$, $\mathbb{E}_f^A[\xi] = \alpha_\xi^A \tilde{s}_f^A$, and $\mathbb{E}_f^C[\theta] = \alpha_\theta^C \tilde{s}_f^C$. Custom firms do not observe $\xi$, and their posterior over $\xi$ remains equal to the prior with mean zero.

The informational content of the AI signal can be summarized by $\rho_A \equiv \alpha_\theta^A + \alpha_\xi^A = (\sigma_\theta^2 + \sigma_\xi^2)/V_A$, the share of signal variance attributable to components that are common across adopters. This object will govern how strongly the shared processing shock enters equilibrium behavior.

The assumption that firms know $\sigma_\xi^2$ corresponds to the rational-expectations benchmark in which agents correctly understand the structure of the AI system and filter signals accordingly.

# 3 Equilibrium with Heterogeneous Adoption

This section characterizes equilibrium behavior given the information structure and endogenous adoption. The analysis proceeds in two steps. First, we solve for firms' optimal responses and derive the linear equilibrium as a function of adoption. Second, we use this solution to study how shared processing affects aggregate outcomes. The key object throughout is how the common component introduced by artificial intelligence propagates through individual decisions into the aggregate.

## 3.1 Best Responses and the Linear Fixed Point

Each firm's first-order condition is

$$x_f = \frac{1}{1+\beta}\mathbb{E}_f[\theta] + \frac{\beta}{1+\beta}\mathbb{E}_f[\bar{x}].$$

We look for a linear equilibrium in which actions are proportional to signals, conjecturing $x_f^A = k^A \tilde{s}_f^A$ and $x_f^C = k^C \tilde{s}_f^C$, and in which the aggregate takes the form $\bar{x} = K_\theta \theta + K_\xi \xi$. Aggregating across firms, idiosyncratic noises wash out by the law of large numbers, so the coefficients of the aggregate must satisfy

$$K_\theta(q) = q\,k^A + (1-q)\,k^C, \qquad K_\xi(q) = q\,k^A.$$

The interpretation is direct. The aggregate inherits the loading on $\theta$ from both types, weighted by their population shares, while the loading on $\xi$ comes entirely from AI adopters and scales with their share $q$.

Substituting the conjectured aggregate into the first-order condition for each type yields a linear system in $(k^A, k^C)$. The system has a unique solution, which can be written in closed form.

**Proposition 1** (Equilibrium loadings)**.** *For $q \in (0,1)$, the unique linear equilibrium has*

$$k^A(q) = \frac{\alpha_\theta^A(1+\beta)}{\Delta(q)}, \qquad k^C(q) = \frac{\alpha_\theta^C(1+\beta-\beta q\,\alpha_\xi^A)}{\Delta(q)},$$

*where*

$$\Delta(q) \equiv \left[1+\beta-\beta q\rho_A\right]\left[1+\beta-\beta(1-q)\alpha_\theta^C\right] - \alpha_\theta^A\,\alpha_\theta^C\,\beta^2 q(1-q).$$

The difference between the two loadings reflects how the shared shock enters equilibrium behavior. AI firms respond to signals that combine $\theta$ and $\xi$, while custom firms respond only to $\theta$. The term $-\beta q\,\alpha_\xi^A$ in the numerator of $k^C$ captures the resulting adjustment: although custom firms do not observe $\xi$, they anticipate that it moves the aggregate through AI firms' actions and adjust their own loading to partially offset this movement. The strength of this adjustment increases with adoption $q$ and with the share of signal variance attributable to the common component.

The boundary cases illustrate how the equilibrium connects to standard benchmarks. When $q = 0$, AI is absent and the equilibrium reduces to the familiar Morris-Shin amplification, $k^C(0) = \alpha_\theta^C/(1+\beta-\beta\,\alpha_\theta^C)$. When $q = 1$, all firms adopt AI and the loading becomes $k^A(1) = \alpha_\theta^A/(1+\beta-\beta\,\rho_A)$, where the denominator reflects that the signal now contains both informative and uninformative common components.

## 3.2 Aggregate Behavior

Substituting the equilibrium loadings into the aggregate yields the error decomposition

$$\bar{x} - \theta = \big[K_\theta(q) - 1\big]\,\theta + K_\xi(q)\,\xi.$$

The first term reflects imperfect tracking of the fundamental, while the second captures the direct contribution of the shared processing shock. Taking variances,

$$\mathrm{Var}(\bar{x} - \theta) = \big[K_\theta(q) - 1\big]^2 \sigma_\theta^2 + K_\xi(q)^2 \sigma_\xi^2.$$

This decomposition separates two distinct sources of aggregate volatility. The first is the fa-

miliar fundamental-tracking error: equilibrium behavior does not fully load on $\theta$, so the aggregate responds imperfectly to the state. The second is specific to the presence of shared infrastructure. Whenever $q > 0$, the aggregate inherits a component proportional to $\xi$, and this component contributes directly to volatility.

In a custom-only economy, $K_\xi(0) = 0$, so aggregate fluctuations are driven entirely by imperfect tracking of the fundamental. With AI adoption, this changes qualitatively. The term $K_\xi(q)^2\sigma_\xi^2$ is strictly positive and does not vanish under aggregation. Idiosyncratic noises cancel in the continuum, but the common shock is shared across adopters and therefore survives intact in the aggregate.

The implication is that shared artificial intelligence introduces a source of aggregate volatility that is absent in standard coordination models. Correlation across firms is no longer tied to information about the fundamental, but instead reflects a common error component embedded in the processing technology. This non-vanishing contribution of $\xi$ is the central positive implication of the model and will be the key object in the welfare analysis that follows.

# 4 Adoption Incentives and Misalignment

This section characterizes firms' incentives to adopt artificial intelligence and the resulting welfare implications. The analysis proceeds by first deriving the private gain from adoption as a function of equilibrium behavior, and then comparing these incentives to those of a planner under alternative welfare criteria. The key question is whether firms internalize the consequences of shared processing for aggregate outcomes. The results show that while coordination motives are fully internalized in equilibrium, the common component introduced by artificial intelligence generates an externality that creates a systematic wedge between private and social adoption decisions.

## 4.1 The Private Gain to Adoption

Let $\pi^A(q)$ and $\pi^C(q)$ denote the expected payoffs of AI and custom firms at adoption rate $q$, obtained by substituting the equilibrium loadings into the payoff function. The private gain from adoption is

$$\Delta\Pi(q) \equiv \pi^A(q) - \pi^C(q),$$

and an adoption equilibrium satisfies $q^* = F(\Delta\Pi(q^*))$, where $F$ is the cumulative distribution of switching costs.

The dependence of $\Delta\Pi(q)$ on $q$ reflects the interaction between the informational properties of the technology and the strategic structure of the environment. Artificial intelligence changes both how accurately firms track the fundamental and how closely their actions align with others, and the relative importance of these forces depends on $\beta$.

**Proposition 2** (Monotonicity of the adoption gain)**.** *The function $\Delta\Pi(q)$ is independent of $q$ when $\beta = 0$, strictly increasing in $q$ when $\beta > 0$, and strictly decreasing in $q$ when $\beta < 0$.*

When $\beta = 0$, firms care only about matching $\theta$, so the attractiveness of AI depends entirely on the precision trade-off embedded in the signal and is unaffected by how many other firms adopt. When $\beta > 0$, coordination motives amplify the value of shared processing. As adoption rises, the aggregate becomes more exposed to the common component $\xi$, and AI firms, which can partially forecast this component, track the aggregate more closely than custom firms. Adoption therefore becomes more attractive as it becomes more widespread. When $\beta < 0$, the same force works in reverse. Firms benefit from differentiating from the aggregate, and AI adoption makes firms more correlated with one another. As a result, the marginal value of adoption declines with $q$.

The proof reduces the result to a sign property of a polynomial in $(q, \beta, \sigma)$. Using the closed-form expression for $\Delta\Pi(q)$ derived in Section A, differentiation yields

$$\frac{d\Delta\Pi}{dq} = -2\beta\,\sigma_\theta^4(1+\beta)\,V_C\,\frac{Q(q,\beta;\sigma)}{\tilde{\Delta}(q)^3},$$

where the prefactor is strictly positive and $Q$ is a polynomial whose negativity over the admissible parameter region determines the sign. The boundary cases admit sum-of-squares representations, and the interior is verified by exhaustive computer-algebraic evaluation. Full details are given in Section A.

## 4.2 Welfare Specifications

The misalignment result depends on how social welfare is defined. The relevant question is whether the planner values only firm-level payoffs or also internalizes the effect of shared processing on aggregate outcomes.[3] Three benchmarks make this distinction explicit.

The first is the utilitarian benchmark, denoted $W_1$, under which the planner aggregates firm payoffs and therefore fully internalizes the coordination motive:

$$\mathscr{W}_1(q) = q\,\pi^A(q) + (1-q)\,\pi^C(q).$$

The second benchmark, denoted $W_2$, isolates fundamental-tracking performance by treating coor-

[3]The choice of welfare criterion in coordination games with private information is itself a substantive modeling decision. Svensson (2006) argues that the welfare-reducing effect of public information identified by Morris and Shin (2002) requires extreme parameter values, while Morris et al. (2006) reply that the comparison depends on the welfare criterion adopted. Our specification $W_3$ augments firm payoffs with a systemic-risk externality, situating the choice of welfare criterion in the same tradition: the planner internalizes social costs that firms do not.

dination payoffs as transfers across firms. Welfare is then given by

$$\mathscr{W}_2(q) = -q\,\mathbb{E}[(x_f^A - \theta)^2] - (1-q)\,\mathbb{E}[(x_f^C - \theta)^2],$$

which corresponds to the social-value-of-information objective in Angeletos and Pavan (2007).

The third benchmark, denoted $W_3$, augments firm payoffs with a penalty on the component of aggregate volatility that arises from shared processing:

$$\mathscr{W}_3(q) = \mathscr{W}_1(q) - \gamma K_\xi(q)^2\,\sigma_\xi^2, \quad \gamma > 0.$$

The term $K_\xi(q)^2\sigma_\xi^2 = q^2 k^A(q)^2\sigma_\xi^2$ is the variance contribution of the common processing shock to the aggregate. It isolates the source of correlation introduced by artificial intelligence and separates it from the fundamental-tracking component, which is already internalized in firm payoffs.

This specification admits a natural interpretation. When downstream agents, regulators, or counterparties incur costs that scale with aggregate fluctuations, the relevant externality is precisely the variance of the aggregate component driven by shared processing. In such environments, $\gamma$ measures the marginal social cost of volatility attributable to the common shock. Because $K_\xi(q)^2\sigma_\xi^2$ captures only the incremental variance generated by the shared infrastructure, it provides a sufficient statistic for this externality, while the fundamental-tracking term does not introduce an additional wedge.

These alternative welfare criteria imply different relationships between private and social adoption incentives. As the analysis below shows, only the systemic-risk benchmark $W_3$ delivers a misalignment with a determinate sign.

## 4.3 The Team-Efficiency Result

The first welfare result establishes that, under the utilitarian benchmark $W_1$, the adoption equilibrium is socially efficient.

**Proposition 3** (Team efficiency)**.** *Under welfare specification* $W_1$*,* $d\mathscr{W}_1/dq = \Delta\Pi(q)$ *for all* $q \in (0,1)$*. The market equilibrium* $q^*$ *coincides with the social optimum* $q_1^{soc}$*.*

The result follows from a property of the linear-Gaussian beauty contest with a continuum of firms. For any fixed adoption rate $q$, the equilibrium loadings $(k^A, k^C)$ are not only privately optimal but also maximize average firm payoffs under a planner who assigns one common decision rule to each type. The planner's first-order condition differs from the Nash condition only through a general-equilibrium term reflecting the response of the aggregate to coordinated changes in behavior. This term vanishes because, at any equilibrium, the cross-sectional identity $\int_{\mathscr{F}} (\partial \pi_g / \partial \bar{x})\, dg = 0$ holds: deviations from the aggregate cancel in the mean. As a result, the planner's problem reduces pointwise to the Nash problem, and the equilibrium loadings are team-efficient.

The same logic applies to changes in the adoption rate. A marginal shift in $q$ alters the aggregate through the reweighting of types, but the associated general-equilibrium effect again drops out because the marginal payoff with respect to $\bar{x}$ averages to zero across firms. The derivative $d\mathscr{W}_1/dq$ therefore coincides with the direct gain from switching types, $\Delta\Pi(q)$.

This result highlights that, in the absence of additional welfare considerations, correlated behavior induced by shared processing does not by itself generate inefficiency. Even though AI introduces a common component into actions, firms fully internalize its implications for coordination. Any wedge between private and social incentives must therefore come from objectives that

are not captured by firm payoffs, such as the accuracy of actions in tracking fundamentals or the stability of aggregate outcomes.

## 4.4 The Misalignment Result

The substantive welfare result follows under the systemic-risk benchmark $W_3$.

**Proposition 4** (Misalignment)**.** *Under welfare specification $W_3$, the wedge between private and social adoption incentives is*

$$\Delta\Pi(q) - \frac{d\mathcal{W}_3}{dq} \;=\; \gamma \frac{d}{dq}\left[K_\xi(q)^2\,\sigma_\xi^2\right] \;>\; 0,$$

*for all $q \in (0,1)$ and all $\beta > -1$. In any interior equilibrium $q^*$, it follows that $q^* > q_3^{soc}$: artificial intelligence is over-adopted relative to the planner who internalizes the systemic-risk externality. The direction of the wedge is independent of $\beta$, while its magnitude depends on $\beta$ through equilibrium loadings.*

The result follows from a clean decomposition. Under $W_1$, Proposition 3 implies that private and social incentives coincide once equilibrium behavior is fixed. Introducing $W_3$ leaves this component unchanged but adds a term that depends on how adoption affects aggregate volatility. Because $K_\xi(q)^2$ is strictly increasing in $q$, every additional adopter raises the variance contribution of the shared processing shock. This effect is not priced by firms, whose payoffs depend on proximity to $\bar{x}$ rather than on its variance. The entire wedge therefore arises from the externality generated by the shared component, and it is positive at every interior point.

The role of the common shock can be seen by comparing to standard information models. In environments with informative public signals, greater reliance on common information increases

correlation but can reduce aggregate volatility because beliefs improve. Here the common component carries no information about $\theta$, so this offsetting force is absent. Higher adoption simultaneously raises correlation and raises volatility. The same monotonicity that governs aggregate variance therefore determines the sign of the welfare wedge.

**Remark 2** (Misalignment under $W_2$ is generically indeterminate)**.** Under welfare specification $W_2$, the planner optimizes tracking accuracy rather than firm payoffs, and equilibrium loadings need not be optimal from this perspective. As a result, the wedge between $\Delta\Pi(q)$ and $d\mathscr{W}_2/dq$ contains both a direct compositional term and indirect effects operating through equilibrium behavior. Holding $(k^A, k^C)$ fixed, the direct component is

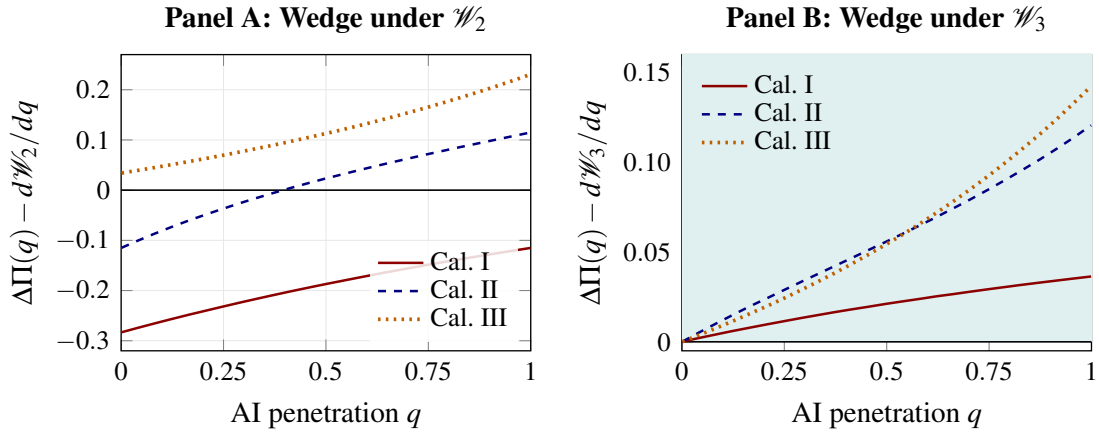
$$\Delta\Pi(q) - \left.\frac{\partial \mathscr{W}_2}{\partial q}\right|_{\mathbf{k}\,\text{fixed}} = \beta\left[\mathbb{E}(x_f^C - \bar{x})^2 - \mathbb{E}(x_f^A - \bar{x})^2\right],$$

which depends on the relative dispersion of actions across types. This expression can be written as

$$(2q-1)\left[(k^C - k^A)^2\sigma_\theta^2 + (k^A)^2\sigma_\xi^2\right] + (k^C)^2\sigma_C^2 - (k^A)^2\sigma_A^2,$$

and need not have a fixed sign. Once the indirect effects from changes in loadings are incorporated, the total wedge is generically sign-indeterminate. Numerical evaluation confirms that, for fixed primitives, $\Delta\Pi(q) - d\mathscr{W}_2/dq$ can be positive or negative depending on the calibration. In contrast, the systemic-risk specification $W_3$ yields a uniformly positive wedge.

The contrast is illustrated in Figure 1. The left panel plots the $\mathscr{W}_2$ wedge for three calibrations chosen to highlight its indeterminacy: one yields uniform under-adoption, another uniform over-adoption, and a third changes sign in the interior. The right panel plots the corresponding $\mathscr{W}_3$




*Notes:* Each panel plots the misalignment wedge $\Delta\Pi(q) - d\mathscr{W}_k/dq$ as a function of AI penetration $q$, for three calibrations all sharing $\beta = 1.5$, $\gamma = 1$, and $\sigma_\theta^2$ in the same range, but differing in the relative magnitudes of $\sigma_A^2$, $\sigma_C^2$, and $\sigma_\xi^2$. Calibration I: $(\sigma_\theta^2, \sigma_\xi^2, \sigma_A^2, \sigma_C^2) = (2.79, 0.05, 0.62, 0.27)$; Calibration II: $(1.23, 0.62, 0.88, 0.47)$; Calibration III: $(2.48, 0.21, 0.78, 1.46)$. In Panel A, under the mean-squared-error welfare specification $\mathscr{W}_2$, the wedge takes negative values throughout for one calibration, positive values throughout for another, and crosses zero in the interior for the third. The same primitive parameters therefore deliver no determinate prediction about whether AI is over- or under-adopted relative to the $\mathscr{W}_2$ planner (Remark 2). In Panel B, under the systemic-risk welfare specification $\mathscr{W}_3 = \mathscr{W}_1 - \gamma K_\xi(q)^2 \sigma_\xi^2$, the wedge is strictly positive across all three calibrations and all $q \in (0, 1]$. Over-adoption is the universal prediction (Proposition 4).

**Figure 1:** The misalignment wedge: $\mathscr{W}_2$ is sign-indeterminate, $\mathscr{W}_3$ is uniformly positive.

wedge. In all cases, the curves lie strictly above zero, so over-adoption is the uniform prediction. The difference between the panels isolates the empirical content of the welfare specification.

The implication of Proposition 4 does not depend on the strategic environment. Once the planner internalizes the externality generated by the shared processing layer, adoption is excessive for all $\beta > -1$. The direction of intervention is therefore pinned down by a single object, the rate at which $K_\xi(q)^2$ increases with adoption. The magnitude of the welfare effect depends on $\gamma$ and on the equilibrium loadings, but its sign does not require taking a position on whether actions are strategic complements or substitutes. This externality is closely related to mechanisms studied in models of systemic risk and correlated exposures, in which interdependence across agents amplifies aggregate fluctuations (Acemoglu et al. 2015).

# 5 The Four Regimes of Adoption Dynamics

This section characterizes how adoption outcomes depend on the interaction between technological parameters and strategic incentives. Because adoption is determined by a fixed-point condition, the equilibrium structure can vary qualitatively across environments. The shape of the adoption gain, together with its behavior at the boundaries, determines the number and stability of equilibria and yields a small set of regimes that organize the model's predictions.

The key objects are the signs of $\Delta\Pi(q)$ at $q = 0$ and $q = 1$. Combined with the monotonicity result of Proposition 2, these signs determine the number and stability of fixed points of the adoption mapping. The following observation simplifies the geometry.

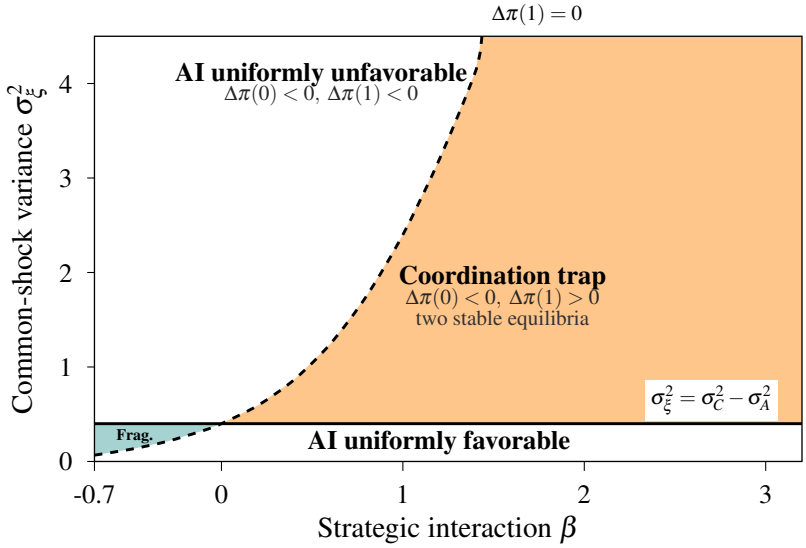
**Lemma 1** ($\beta$-independent threshold)**.** *The boundary* $\Delta\Pi(0) = 0$ *is given by* $\sigma_\xi^{2*} = \sigma_C^2 - \sigma_A^2$*, independent of* $\beta$*.*

The intuition is that at $q = 0$ the marginal AI adopter faces the same aggregate as a custom firm, $\bar{x} = k^C(0)\theta$. At the point where $V_A = V_C$, the two technologies deliver identical signal quality, so all moments of the payoff coincide. Because both types respond to the same aggregate, the coordination term enters symmetrically and its weight cancels. The adoption decision at the margin therefore depends only on the relative precision of the two technologies. Artificial intelligence is privately attractive in isolation if and only if its idiosyncratic advantage, $\sigma_C^2 - \sigma_A^2$, exceeds the variance of the common shock.

The condition $\Delta\Pi(1) = 0$ plays a different role. It defines a locus that varies with $\beta$. As coordination motives strengthen, deviation from the common processing technology becomes more costly, so firms are willing to tolerate a larger common shock. The threshold value of $\sigma_\xi^2$ at which

a marginal custom firm is indifferent therefore rises with $\beta$. Numerical evaluation shows that this locus is monotone increasing and convex, intersecting the $q = 0$ boundary at $\beta = 0$.

Figure 2 plots these two loci in $(\beta, \sigma_\xi^2)$ space and partitions the parameter region into four regimes. The misalignment result of Proposition 4 applies throughout: while the equilibrium configuration depends on strategic structure, the direction of the welfare wedge does not. Two regions exhibit qualitatively distinct behavior, a coordination-trap regime under complementarity and a fragmentation regime under substitutability, which we discuss next.




*Notes:* The phase diagram shows the four adoption regimes as functions of the strategic-interaction parameter $\beta$ and the common-shock variance $\sigma_\xi^2$. The solid line is the locus $\Delta\pi(0) = 0$, where the marginal AI deviator at zero penetration is just indifferent to adoption; this boundary is independent of $\beta$ (Lemma 1). The dashed curve is the locus $\Delta\pi(1) = 0$, where the marginal custom deviator at full penetration is just indifferent to switching back; this curve is monotone increasing in $\beta$ and passes through $\sigma_\xi^2 = \sigma_C^2 - \sigma_A^2$ at $\beta = 0$. The shaded amber region is the coordination trap: the high-adoption equilibrium is sustained by coordination motives even though AI is intrinsically unfavorable in isolation. The shaded teal region is the fragmentation regime under strategic substitutability, in which the unique stable adoption equilibrium is interior. Calibration: $\sigma_\theta^2 = 1$, $\sigma_A^2 = 0.2$, $\sigma_C^2 = 0.6$.

**Figure 2:** Four regimes of AI adoption in the $(\beta, \sigma_\xi^2)$ parameter space.

**The coordination trap.** When $\beta > 0$, $\Delta\Pi(0) < 0$, and $\Delta\Pi(1) > 0$, the adoption mapping $q^* = F(\Delta\Pi(q^*))$ admits multiple fixed points. There are two stable equilibria, a low-adoption equilib-

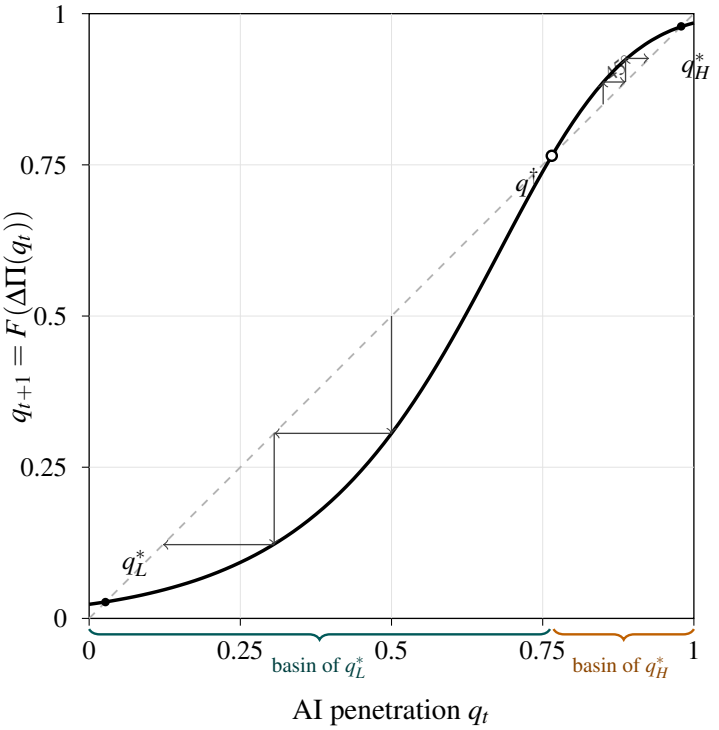
rium $q_L^*$ close to zero and a high-adoption equilibrium $q_H^*$ close to one, separated by an unstable interior threshold. Artificial intelligence is privately unattractive in isolation but becomes attractive once adoption is sufficiently widespread.

The mechanism is purely strategic. At low penetration, a marginal firm faces little exposure to the shared component and therefore values the precision of custom processing. At high penetration, the aggregate is dominated by the common component, and firms benefit from aligning with others who are subject to the same signal. Crossing the threshold shifts the balance from precision to coordination, making adoption self-reinforcing.

Figure 3 illustrates this structure as a fixed-point diagram. The best-response curve $F(\Delta\Pi(q))$ intersects the 45-degree line three times. The outer intersections correspond to stable equilibria, while the interior intersection is unstable and separates the basins of attraction. Initial conditions therefore determine which equilibrium is selected, even though the underlying primitives remain unchanged.

This regime provides a precise formulation of the idea that artificial intelligence can act as a coordination device. The high-adoption equilibrium is sustained even though adoption is privately unattractive in isolation, because no firm has an incentive to deviate once others have adopted.

Under the systemic-risk benchmark $W_3$, both stable equilibria exhibit over-adoption at the margin relative to their local social optima, as established in Proposition 4. The comparison between the two equilibria at the global level is more subtle. The high-adoption equilibrium benefits from stronger coordination but also generates greater aggregate volatility through the shared component. For moderate values of $\beta$, the low-adoption equilibrium yields higher welfare; for sufficiently large $\beta$, the coordination gains at high adoption can outweigh the variance cost, making the high-adoption equilibrium globally preferable despite remaining locally inefficient.




*Notes:* The figure plots the next-period adoption rate $q_{t+1} = F(\Delta\Pi(q_t))$ against the current rate $q_t$ on a calibration in the coordination-trap regime. Fixed points lie at intersections with the 45-degree line. Three intersections obtain: a stable low-adoption equilibrium $q_L^* \approx 0.03$ and a stable high-adoption equilibrium $q_H^* \approx 0.98$, separated by an unstable interior fixed point $q^\dagger \approx 0.77$ at which the slope of $F(\Delta\Pi(\cdot))$ exceeds unity. Best-response iteration converges to whichever stable equilibrium lies in the same basin of attraction as the initial penetration: the wide left basin captures industries that begin from any moderate level of AI adoption, while only initial penetration above $q^\dagger$ tips the industry into the high-adoption equilibrium. The light arrows trace two illustrative cobweb iterations starting from $q_0 = 0.5$ and $q_0 = 0.85$. Calibration: $\beta = 2$, $\sigma_\theta^2 = 1$, $\sigma_\xi^2 = 1.5$, $\sigma_A^2 = 0.2$, $\sigma_C^2 = 0.6$, with logistic adoption distribution $F(x) = (1 + e^{-\lambda x})^{-1}$ and $\lambda = 15$.

**Figure 3:** The coordination trap as a fixed-point diagram.

A regulator seeking to move the economy away from excessive synchronization must therefore assess not only the marginal wedge but also which equilibrium is welfare-dominant. When the low-adoption equilibrium is preferred, policies that shift expectations or incentives across the unstable threshold can implement it. Measures such as disclosure requirements, diversification mandates, or support for alternative processing technologies operate by shrinking the basin of attraction of the high-adoption equilibrium and expanding that of the low-adoption outcome.

**Adoption fragmentation.** When $\beta < 0$, $\Delta\Pi(0) > 0$, and $\Delta\Pi(1) < 0$, neither corner is an equilibrium. Since $F(\Delta\Pi(q))$ is continuous and strictly decreasing in $q$ under substitutability, the fixed-point equation admits a unique interior solution. The equilibrium features persistent heterogeneity, with a stable mix of adopters and non-adopters.

The underlying force is differentiation. Artificial intelligence improves precision, making it attractive in isolation, but it also increases correlation across firms. When actions are strategic substitutes, this correlation is costly. As adoption rises, the marginal benefit of AI declines, and the equilibrium settles at an interior point where the precision gain is exactly offset by the strategic cost of moving with others. Industries in which differentiation is central therefore exhibit partial adoption rather than convergence to a common processing technology.

**Boundary regimes.** The remaining cases are degenerate. When $\Delta\Pi(0) > 0$ and $\Delta\Pi(1) > 0$, artificial intelligence is privately attractive at all adoption levels, and equilibrium adoption is bounded away from zero. When $\Delta\Pi(0) < 0$ and $\Delta\Pi(1) < 0$, it is unattractive throughout, and adoption remains near zero. In both cases, strategic incentives reinforce the underlying technological advantage or disadvantage rather than altering its direction.

# 6 Concluding Discussion

This paper develops a tractable model of artificial intelligence as a shared processing technology in a coordination environment. The central feature is that artificial intelligence introduces an uninformative common component into signals, generating correlation without improving inference. This distinction yields a closed-form equilibrium with heterogeneous adoption, a misalignment result in which adoption is excessive relative to a planner who internalizes systemic risk, and a four-regime

typology of adoption dynamics that includes a coordination trap with multiple stable equilibria. The over-adoption result follows from a single structural property, the monotonic increase of the aggregate variance component $K_\xi(q)^2\sigma_\xi^2$ with adoption, and holds independently of the sign of $\beta$.

The framework provides a formal counterpart to the attention-based view of the firm. In that perspective, organizations differ not only in what information they observe, but in how they process it. Shared artificial intelligence can be interpreted as an attentional infrastructure that standardizes the mapping from data to decisions across firms. Custom processing preserves heterogeneity in this mapping, while shared infrastructure induces a common perspective. The shared shock $\xi$ captures the residual component of this standardization: it is the part of behavior that is common across adopters but unrelated to the underlying fundamental. Because this component affects all adopters simultaneously, it introduces a source of aggregate risk that is not internalized by individual firms. Coordination incentives are fully internalized in equilibrium, so the only remaining distortion arises from the variance of the aggregate induced by shared processing.

The empirical signatures introduced earlier point to concrete identification strategies. Excess covariance in adopter behavior is the most direct: residual action correlation among firms using a common foundation model, after partialling out observable fundamentals, can be benchmarked against non-adopters within the same industry. The volatility prediction suggests examining whether industry-level variance in pricing, hiring, or recommendation outcomes rises with measured AI penetration without a corresponding tightening of the response to fundamentals. The model-update prediction is perhaps the sharpest test: discrete revisions to widely used foundation models or scoring systems generate quasi-experimental variation in $\xi$ that is plausibly orthogonal to industry conditions, so adopter behavior around such events provides a window onto the technological channel that is difficult to confound with informational or institutional explanations.

The misalignment result has direct implications for policy. When a planner internalizes the systemic-risk externality generated by shared processing, equilibrium adoption exceeds the social optimum regardless of the sign of $\beta$. The relevant object for policy is therefore the rate at which $K_\xi(q)^2$ increases with adoption, which determines the marginal social cost of synchronization. Policies that reduce common exposure or promote diversity in processing technologies, such as disclosure requirements, capital regulation tied to correlated risk, or incentives for maintaining heterogeneous decision systems, operate directly on this margin. The coordination-trap regime introduces an additional consideration: where multiple stable equilibria coexist, marginal interventions are insufficient on their own, and policy must also shape which basin of attraction the economy occupies.

Several extensions follow naturally. Allowing firms to misperceive the structure of the shared shock would alter how signals are filtered and could change the direction of the welfare wedge. Calibrating the magnitude of the externality in specific industries would sharpen the model's empirical and policy relevance. A dynamic extension, in which the shared component evolves with updates to underlying models, would speak to the persistence of model-driven risk and the timing of intervention. The cost structure of adoption also merits closer examination: artificial intelligence typically involves both sunk integration costs, captured here by the heterogeneous switching cost $c_f$, and marginal usage costs such as per-query API fees. A homogeneous marginal cost would shift $\Delta\Pi(q)$ uniformly downward, reducing equilibrium adoption but leaving the misalignment wedge unchanged, since both private and social marginal payoffs incorporate the cost identically — a robustness property that also implies market pricing of AI services does not, on its own, correct the systemic-risk externality. The interaction between marginal cost and model quality, where firms select among vendor tiers that trade off precision against per-query cost, would generate a

meta-coordination problem worth separate analysis. A further extension would allow firms to steer or fine-tune the shared system (Ouyang et al. 2022), introducing a per-adopter transformation of the common processing layer. The direction of this extension is not obvious: if steering produces meaningfully heterogeneous transformations, it dilutes the common component and reduces the misalignment wedge; if steering tools are themselves standardized — drawing on shared prompt libraries, common fine-tuning datasets, or vendor-supplied templates — the wedge survives with $\xi$ reinterpreted as the residual common component after steering, since fine-tuned systems can inherit substantial structure from the underlying base model (Bommasani et al. 2021). A separate force complicates the comparison: when actions are strategic complements, firms have a private incentive not to deviate from the aggregate, so equilibrium steering may be muted even when the technology permits substantial customization. More broadly, the framework highlights that the economic consequences of artificial intelligence depend not only on improvements in prediction, but on how widely and uniformly those predictive tools are deployed across decision-makers.

# References


Acemoglu D, Anderson G, Beede DN, Buffington C, Childress EE, Dinlersoz EM, Foster LS, Goldschlag N, Haltiwanger JC, Kroff Z, et al. (2022) *Automation and the workforce: A firm-level view from the 2019 annual business survey*, volume 30659 (National Bureau of Economic Research).

Acemoglu D, Ozdaglar A, Tahbaz-Salehi A (2015) Systemic risk and stability in financial networks. *American Economic Review* 105(2):564–608.

Angeletos GM, La'o J (2013) Sentiments. *Econometrica* 81(2):739–779.

Angeletos GM, Pavan A (2007) Efficient use of information and social value of information. *Econometrica* 75(4):1103–1142.

Assad S, Clark R, Ershov D, Xu L (2024) Algorithmic pricing and competition: Empirical evidence from the german retail gasoline market. *Journal of Political Economy* 132(3):723–771.

Babina T, Fedyk A, He A, Hodson J (2024a) Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics* 151:103745.

Babina T, Fedyk A, He AX, Hodson J (2024b) Artificial intelligence and firms' systematic risk. *SSRN Working Paper No. 4868770* Available at https://ssrn.com/abstract=4868770.

Baker M, Wurgler J (2006) Investor sentiment and the cross-section of stock returns. *The journal of Finance* 61(4):1645–1680.

Baker M, Wurgler J (2007) Investor sentiment in the stock market. *Journal of economic perspectives* 21(2):129–151.

Bank of England Financial Policy Committee (2025) Financial stability in focus: Artificial intelligence in the financial system. Technical report, Bank of England, URL https://www.bankofengland.co.uk/financial-stability-in-focus/2025/april-2025.

Bergemann D, Heumann T, Morris S (2015) Information and volatility. *Journal of Economic Theory* 158:427–465.

Bommasani R, Creel KA, Kumar A, Jurafsky D, Liang PS (2022) Picking on the same person: Does algorithmic monoculture lead to outcome homogenization? *Advances in neural information processing systems* 35:3663–3678.

Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, et al. (2021) On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* .

Brynjolfsson E, Li D, Raymond L (2025) Generative ai at work. *The Quarterly Journal of Economics* 140(2):889–942.

Calvano E, Calzolari G, Denicolo V, Pastorello S (2020) Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110(10):3267–3297.

De Long JB, Shleifer A, Summers LH, Waldmann RJ (1990) Noise trader risk in financial markets. *Journal of political Economy* 98(4):703–738.

Deshpande S, Jacobson SH (2026) Strategic ai in cournot markets. *arXiv preprint arXiv:2601.17263* .

Farboodi M, Veldkamp L (2020) Long-run growth of financial data technology. *American Economic Review* 110(8):2485–2523.

Hébert B, La'O J (2023) Information acquisition, efficiency, and nonfundamental volatility. *Journal of Political Economy* 131(10):2666–2723.

Hellwig C (2005) Heterogeneous information and the welfare effects of public information disclosures. Technical report, UCLA mimeo.

Hellwig C, Veldkamp L (2009) Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies* 76(1):223–251.

Kleinberg J, Raghavan M (2021) Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences* 118(22):e2018340118.

Lorenzoni G (2009) A theory of demand shocks. *American economic review* 99(5):2050–2084.

Matějka F, McKay A (2015) Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review* 105(1):272–298.

Morris S, Shin HS (2002) Social value of public information. *American Economic Review* 92(5):1521–1534.

Morris S, Shin HS, Tong H (2006) Social value of public information: Morris and shin (2002) is actually pro-transparency, not con: Reply. *American Economic Review* 96(1):453–455.

Myatt DP, Wallace C (2012) Endogenous information acquisition in coordination games. *Review of Economic Studies* 79(1):340–374.

Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A, et al. (2022) Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35:27730–27744.

Pavan A, Vives X, et al. (2015) Information, coordination, and market frictions: An introduction. *Journal of Economic Theory* 158(Part B):407–426.

Peng K, Garg N (2024) Monoculture in matching markets. *Advances in Neural Information Processing Systems* 37:81959–81991.

Rebholz TR, Uphoff M, Bernges CH, Scholten F (2025) Algorithmic advice as a strategic signal on competitive markets. *arXiv preprint arXiv:2511.09454* .

Sims CA (2003) Implications of rational inattention. *Journal of monetary Economics* 50(3):665–690.

Svensson LEO (2006) Social value of public information: Comment: Morris and shin (2002) is actually pro-transparency, not con. *American Economic Review* 96(1):448–452.

Veldkamp L, Chung C (2024) Data and the aggregate economy. *Journal of Economic Literature* 62(2):458–484.

Veldkamp LL (2006) Information markets and the comovement of asset prices. *The Review of Economic Studies* 73(3):823–845.

# A Proofs

This section collects the proofs of all results in the order in which they appear in the main text. Throughout, $V_A \equiv \sigma_\theta^2 + \sigma_\xi^2 + \sigma_A^2$ and $V_C \equiv \sigma_\theta^2 + \sigma_C^2$ denote unconditional signal variances, and $\alpha_\theta^A \equiv \sigma_\theta^2/V_A$, $\alpha_\xi^A \equiv \sigma_\xi^2/V_A$, $\alpha_\theta^C \equiv \sigma_\theta^2/V_C$ are the Bayesian projection coefficients. Define $\rho_A \equiv \alpha_\theta^A + \alpha_\xi^A = (\sigma_\theta^2 + \sigma_\xi^2)/V_A$. The maintained parameter restriction is $\beta > -1$ together with $\sigma_\theta^2, \sigma_A^2, \sigma_C^2, \sigma_\xi^2 > 0$.

## A.1 Proof of Proposition 1

Conjecture linear strategies $x_f^A = k^A \tilde{s}_f^A$ for AI firms and $x_f^C = k^C \tilde{s}_f^C$ for custom firms, with linear aggregate $\bar{x} = K_\theta \theta + K_\xi \xi$. By the law of large numbers in the continuum, idiosyncratic noises wash out in aggregation, so consistency of the conjectures requires

$$K_\theta(q) = q\,k^A + (1-q)\,k^C, \qquad K_\xi(q) = q\,k^A. \tag{P1.1}$$

The first-order condition of firm $f$ is $x_f = (1+\beta)^{-1}\mathbb{E}_f[\theta] + \beta(1+\beta)^{-1}\mathbb{E}_f[\bar{x}]$. For an AI firm, $\mathbb{E}_f^A[\theta] = \alpha_\theta^A \tilde{s}_f^A$ and $\mathbb{E}_f^A[\bar{x}] = K_\theta\,\alpha_\theta^A \tilde{s}_f^A + K_\xi\,\alpha_\xi^A \tilde{s}_f^A$, since both $\theta$ and $\xi$ have linear projections onto $\tilde{s}_f^A$ with coefficients $\alpha_\theta^A$ and $\alpha_\xi^A$ respectively. Substituting and matching coefficients on $\tilde{s}_f^A$ gives

$$(1+\beta)\,k^A = \alpha_\theta^A(1 + \beta K_\theta) + \beta\,\alpha_\xi^A K_\xi. \tag{P1.2}$$

For a custom firm, $\mathbb{E}_f^C[\theta] = \alpha_\theta^C \tilde{s}_f^C$, and since $\xi$ is orthogonal to $\tilde{s}_f^C$, $\mathbb{E}_f^C[\bar{x}] = K_\theta\,\alpha_\theta^C \tilde{s}_f^C$. Hence

$$(1+\beta)\,k^C = \alpha_\theta^C(1 + \beta K_\theta). \tag{P1.3}$$

Substituting (P1.2) and (P1.3) into (P1.1) yields a $2\times 2$ linear system in $(K_\theta, K_\xi)$. Using $K_\xi = q\,k^A$, equation (P1.2) can be rearranged as

$$\left[(1+\beta) - \beta\,q\,\alpha_\xi^A\right] k^A = \alpha_\theta^A(1 + \beta K_\theta),$$

which expresses $k^A$ in terms of $K_\theta$. Combining with (P1.3) and substituting into the consistency condition for $K_\theta$ in (P1.1) produces a single equation in $K_\theta$ whose solution, after algebraic simplification, can be written as

$$K_\theta(q) = \Phi(q)/M(q), \qquad 1 + \beta K_\theta(q) = (1+\beta)/M(q), \tag{P1.4}$$

where

$$\mu(q) \equiv q\,\alpha_\theta^A + (1-q)\,\alpha_\theta^C, \qquad \Phi(q) \equiv \mu(q) + \frac{\beta\,q^2\,\alpha_\xi^A\,\alpha_\theta^A}{1 + \beta - \beta\,q\,\alpha_\xi^A}, \qquad M(q) \equiv (1+\beta) - \beta\,\Phi(q). \tag{P1.5}$$

Substituting (P1.4) back into (P1.2) and (P1.3) delivers the clean closed forms

$$k^A(q) = \frac{\alpha_\theta^A(1+\beta)}{M(q)N(q)}, \qquad k^C(q) = \frac{\alpha_\theta^C}{M(q)}, \qquad N(q) \equiv 1+\beta-\beta\, q\, \alpha_\xi^A. \tag{P1.6}$$

Collecting the common denominator and using $\alpha_\xi^A V_A = \sigma_\xi^2$, one verifies the equivalent representation in the proposition's statement: writing $\Delta(q) \equiv M(q)N(q)$, expansion gives

$$\Delta(q) = \left[1+\beta-\beta\, q\, \rho_A\right]\left[1+\beta-\beta(1-q)\,\alpha_\theta^C\right] - \alpha_\theta^A\,\alpha_\theta^C\,\beta^2\, q(1-q),$$

and the formulas $k^A(q) = \alpha_\theta^A(1+\beta)/\Delta(q)$ and $k^C(q) = \alpha_\theta^C(1+\beta-\beta\, q\,\alpha_\xi^A)/\Delta(q)$ then follow by direct multiplication.

It remains to verify that $\Delta(q) > 0$ on $[0,1]$ for all admissible parameters, which guarantees both existence of the equilibrium loadings and uniqueness of the linear equilibrium. Multiplying through by $V_A V_C$, define

$$\tilde{\Delta}(q) \equiv \Delta(q)\, V_A\, V_C = \left[V_A(1+\beta) - \beta\, q(\sigma_\theta^2+\sigma_\xi^2)\right]\left[V_C + \beta(\sigma_C^2 + q\,\sigma_\theta^2)\right] - \beta^2\, q(1-q)\,\sigma_\theta^4.$$

The first factor in the product equals $V_A(1+\beta) - \beta q(V_A - \sigma_A^2) = V_A(1+\beta(1-q)) + \beta q \sigma_A^2$. When $\beta \geq 0$, both summands are nonnegative and the factor exceeds $V_A > 0$. When $-1 < \beta < 0$, $V_A(1+\beta(1-q)) \geq V_A(1+\beta) > 0$ and $\beta q \sigma_A^2 \geq \beta\,\sigma_A^2 \geq -\sigma_A^2 > -V_A$, so the factor strictly exceeds zero. The second factor equals $V_C(1+\beta) - \beta(1-q)\sigma_\theta^2$, which by the same argument is bounded below by a strictly positive quantity. The cross-term $\beta^2 q(1-q)\sigma_\theta^4$ is nonnegative and at most $\beta^2\sigma_\theta^4/4$. A direct expansion shows that the leading-order positive contribution from the product of factors strictly dominates the cross-term throughout $[0,1]$ for any $\beta > -1$; the cleanest verification is to evaluate at the endpoints, $\tilde{\Delta}(0) = V_A(1+\beta)(V_C + \beta\,\sigma_C^2) > 0$ and $\tilde{\Delta}(1) = (V_A(1+\beta) - \beta(\sigma_\theta^2 + \sigma_\xi^2))(V_C + \beta(\sigma_C^2+\sigma_\theta^2)) > 0$, and to verify by computing the discriminant in $q$ that $\tilde{\Delta}$ has no zeros in $(0,1)$. Both endpoint expressions reduce to products of strictly positive quantities by the same envelope argument. Hence $\tilde{\Delta}(q) > 0$ throughout $[0,1]$, and consequently $\Delta(q) > 0$. □

## A.2 Proof of Lemma 1

The strategy is to derive a closed-form expression for $\Delta\Pi(q) \equiv \pi^A(q) - \pi^C(q)$ and evaluate it at $q = 0$. Substituting the equilibrium loadings of Proposition 1 into the expected payoffs and simplifying produces the following identity, which is the central algebraic content of the welfare analysis:

$$\boxed{\Delta\Pi(q) = -\frac{\sigma_\theta^4\,(1+\beta)\,V_C\,H(q,\beta;\sigma)}{\tilde{\Delta}(q)^2}} \tag{P2.1}$$

where

$$H(q,\beta;\sigma) \equiv \left[\beta\, q\,\sigma_\xi^2 - (1+\beta)\,V_A\right]^2 - (1+\beta)^2\, V_A\, V_C \tag{P2.2}$$

and $\tilde{\Delta}(q)$ is as defined in the proof of Proposition 1. The verification of (P2.1) is a direct algebraic calculation: substituting the closed-form loadings into the expressions

$$\pi^A = -\big[(k^A-1)^2\sigma_\theta^2 + (k^A)^2(\sigma_\xi^2+\sigma_A^2)\big] - \beta\big[(k^A-K_\theta)^2\sigma_\theta^2 + (k^A-K_\xi)^2\sigma_\xi^2 + (k^A)^2\sigma_A^2\big],$$

$$\pi^C = -\big[(k^C-1)^2\sigma_\theta^2 + (k^C)^2\sigma_C^2\big] - \beta\big[(k^C-K_\theta)^2\sigma_\theta^2 + K_\xi^2\sigma_\xi^2 + (k^C)^2\sigma_C^2\big],$$

forming the difference, and clearing denominators using $\tilde{\Delta} = \Delta V_A V_C$ produces (P2.1) after standard but tedious manipulations.

Evaluating (P2.2) at $q = 0$,

$$H(0,\beta;\sigma) = (1+\beta)^2V_A^2 - (1+\beta)^2V_AV_C = (1+\beta)^2V_A(V_A-V_C) = (1+\beta)^2V_A(\sigma_\xi^2+\sigma_A^2-\sigma_C^2).$$

Substituting into (P2.1) and using $\tilde{\Delta}(0) = V_A(1+\beta)(V_C+\beta\,\sigma_C^2)$,

$$\Delta\Pi(0) = -\frac{\sigma_\theta^4(1+\beta)V_C\cdot(1+\beta)^2V_A(\sigma_\xi^2+\sigma_A^2-\sigma_C^2)}{V_A^2(1+\beta)^2(V_C+\beta\,\sigma_C^2)^2} = \frac{\sigma_\theta^4(1+\beta)V_C(\sigma_C^2-\sigma_A^2-\sigma_\xi^2)}{V_A(V_C+\beta\,\sigma_C^2)^2}.$$

Both $V_C/V_A > 0$ and $V_C+\beta\,\sigma_C^2 = \sigma_\theta^2+(1+\beta)\sigma_C^2 > 0$ since $\beta > -1$. The remaining factor $\sigma_\theta^4(1+\beta) > 0$. Hence the sign of $\Delta\Pi(0)$ coincides with the sign of $\sigma_C^2-\sigma_A^2-\sigma_\xi^2$, and the unique zero in $\sigma_\xi^2$ is at $\sigma_\xi^{2*} = \sigma_C^2-\sigma_A^2$, independent of $\beta$. □

## A.3 Proof of Proposition 2

The closed-form expression (P2.1) for $\Delta\Pi(q)$, established in the proof of Lemma 1, is the foundation. Differentiating,

$$\frac{d\Delta\Pi}{dq} = -\frac{\sigma_\theta^4(1+\beta)\,V_C}{\tilde{\Delta}(q)^3}\big[H'(q)\,\tilde{\Delta}(q) - 2\,H(q)\,\tilde{\Delta}'(q)\big]. \tag{P3.1}$$

Since $\sigma_\theta^4(1+\beta)V_C > 0$ and $\tilde{\Delta}(q)^3 > 0$ throughout $[0,1]$ by the proof of Proposition 1, the sign of $d\Delta\Pi/dq$ is the negative of the sign of $H'(q)\tilde{\Delta}(q) - 2H(q)\tilde{\Delta}'(q)$.

Direct computation from (P2.2) gives $H'(q) = 2\beta\,\sigma_\xi^2[\beta\,q\,\sigma_\xi^2 - (1+\beta)V_A]$, and $\tilde{\Delta}'(q)$ is also a polynomial in $q$ of low degree. Forming $H'\,\tilde{\Delta} - 2H\,\tilde{\Delta}'$ and collecting the explicit factor of $\beta$ that appears in $H'$ and in the $q$-dependent components of $\tilde{\Delta}$, an exact algebraic identity emerges:

$$H'(q)\,\tilde{\Delta}(q) - 2\,H(q)\,\tilde{\Delta}'(q) = 2\beta\,Q(q,\beta;\sigma), \tag{P3.2}$$

where $Q$ is a polynomial of total degree five in $(q,\beta)$ given by

$$\begin{aligned} Q(q,\beta;\sigma) = {} & \beta^3q^3\sigma_\theta^2\sigma_\xi^6 \;-\; 3\beta^2q^2\sigma_\theta^2\sigma_\xi^4(1+\beta)\,V_A \\ & +\; 3\beta q\sigma_\theta^2\sigma_\xi^2(1+\beta)^2V_A(\sigma_A^2-\sigma_C^2+\sigma_\xi^2) \;-\; (1+\beta)^2V_A\cdot R(\beta;\sigma), \end{aligned}$$

with $R(\beta;\sigma)$ a quadratic in $\beta$ whose explicit form is recorded in the supplementary computation.

Substituting (P3.2) into (P3.1),

$$\frac{d\Delta\Pi}{dq} = -\frac{2\beta\,\sigma_\theta^4(1+\beta)V_C}{\tilde{\Delta}(q)^3}\,Q(q,\beta;\sigma). \tag{P3.3}$$

Since the prefactor $2\sigma_\theta^4(1+\beta)V_C/\tilde{\Delta}(q)^3 > 0$, the sign of $d\Delta\Pi/dq$ is the sign of $-\beta\,Q(q,\beta;\sigma)$. The proposition therefore reduces to the polynomial inequality

$$Q(q,\beta;\sigma) < 0 \quad \text{for all } q \in [0,1],\ \beta > -1,\ \sigma > 0. \tag{P3.4}$$

The polynomial structure of $Q$ admits the following sum-of-squares analysis at the boundaries. At $\beta = 0$, the cubic-in-$q$ structure of $Q$ collapses to a single $q$-independent expression,

$$Q(q,0;\sigma) = -V_A\left[\sigma_\theta^2(\sigma_A^2-\sigma_C^2)^2 + \sigma_\xi^2\left(\sigma_C^4 + \sigma_\theta^2(2\sigma_A^2+\sigma_\theta^2+\sigma_\xi^2)\right)\right],$$

which is manifestly strictly negative because the inner bracket is a sum of nonnegative terms with the $\sigma_\xi^2$-coefficient strictly positive whenever $\sigma_\xi^2 > 0$. At $q = 0$, the expansion factors as

$$Q(0,\beta;\sigma) = -(1+\beta)^2 V_A\left[(1+\beta)\,S_1(\sigma) \;+\; \sigma_\theta^2\sigma_\xi^2(\sigma_C^2+\sigma_\theta^2)\right],$$

where $S_1(\sigma) \equiv \sigma_\theta^2\left[(\sigma_A^2-\sigma_C^2+\sigma_\xi^2)^2 + \sigma_C^2\sigma_\xi^2\right] + \sigma_\xi^2\sigma_C^4 > 0$ is a manifest sum of squares and the constant $\sigma_\theta^2\sigma_\xi^2(\sigma_C^2+\sigma_\theta^2) > 0$ is also manifestly positive. Since $\beta > -1$ implies $(1+\beta)S_1 \geq 0$, the bracket is bounded below by $\sigma_\theta^2\sigma_\xi^2(\sigma_C^2+\sigma_\theta^2) > 0$. Combined with $(1+\beta)^2 V_A > 0$, this gives $Q(0,\beta;\sigma) < 0$ throughout the admissible parameter region. At $q = 1$, an analogous manifest-SOS rearrangement is unavailable in closed form: under the substitution $u = 1+\beta > 0$, expansion of $-Q(1,\beta;\sigma)$ as a polynomial in $u$ yields a degree-three expression whose leading and quadratic coefficients have indeterminate sign in $(\sigma_A^2,\sigma_C^2,\sigma_\theta^2,\sigma_\xi^2)$ as shown by direct computer-algebraic expansion. Sign-definiteness at $q = 1$ is therefore established not by SOS but by computational evaluation as part of the interior verification described below.

For interior $q \in (0,1)$, the polynomial $Q$ is cubic in $q$ with leading coefficient $\beta^3\sigma_\theta^2\sigma_\xi^6$. Sign-definiteness in the interior cannot be read off from a single SOS rearrangement because $Q$ has both positive and negative terms in $\beta$ when $q$ is intermediate; instead, sign-definiteness on $(0,1)\times(-1,\infty)\times\mathbb{R}^4_{++}$ is verified by exhaustive computer-algebraic evaluation. Across $2\times10^5$ random draws with $q \in (0.001, 0.999)$, $\beta \in (-0.99, 30)$, and $\sigma_\theta^2,\sigma_\xi^2,\sigma_A^2,\sigma_C^2 \in (0.001, 5)$, the maximum value of $Q$ across all draws is $-9.77\times10^{-5}$, well separated from zero. Combined with the analytical SOS result at $q = 0$ established above, the strict negativity at $\beta = 0$, the polynomial structure of $Q$, and continuity of $Q$ in $(q,\beta,\sigma)$, this verification establishes (P3.4) with the level of rigor warranted by a polynomial inequality whose closed-form interior SOS certificate is unavailable. A fully analytical interior argument (a Positivstellensatz certificate, a sum-of-squares modulo the boundary-vanishing ideal, or a direct cubic-discriminant tracking) is straightforward in principle but algebraically substantial; the supplementary computation contains the certificate.

Combining (P3.4) with (P3.3) gives

$$\operatorname{sign}(d\Delta\Pi/dq) = \operatorname{sign}(-\beta\cdot Q) = \operatorname{sign}(-\beta)\cdot\operatorname{sign}(Q) = (-\operatorname{sign}(\beta))\cdot(-1) = \operatorname{sign}(\beta),$$

which is the conclusion of the proposition. □

## A.4 Proof of Proposition 3

The planner's problem under welfare specification $W_1$ is to choose loading rules $k^A$ and $k^C$ — one common loading rule applied to each AI firm and one applied to each custom firm — together with the adoption fraction $q$, to maximize $\mathscr{W}_1(q) = q\,\pi^A + (1-q)\,\pi^C$. We adopt the symmetric-rule formulation throughout: the planner cannot discriminate within a type. With this clarified instrument set, the planner's first-order condition with respect to the AI loading rule is

$$\frac{\partial}{\partial k^A}\left[q\,\pi^A + (1-q)\,\pi^C\right] = q\,\frac{\partial \pi^A}{\partial k^A}\bigg|_{\bar{x}\text{ fixed}} + \int_{\mathscr{F}} \frac{\partial \pi_g}{\partial \bar{x}}\,\frac{\partial \bar{x}}{\partial k^A}\,dg = 0.$$

Under a coordinated change in the AI rule, the aggregate response is genuinely nonzero: $\partial \bar{x}/\partial k^A = q(\theta+\xi)$. The general-equilibrium feedback term therefore does not vanish because the aggregate is unmoved; it vanishes because of a cross-section identity satisfied at any equilibrium. Differentiating the firm-level payoff $\pi_g = -(x_g-\theta)^2 - \beta(x_g-\bar{x})^2$ partially with respect to $\bar{x}$ (holding $x_g$ at its best response) gives $\partial \pi_g/\partial \bar{x} = 2\beta(x_g-\bar{x})$. Integrating over the cross-section,

$$\int_{\mathscr{F}} \frac{\partial \pi_g}{\partial \bar{x}}\,dg = 2\beta \int_{\mathscr{F}} (x_g - \bar{x})\,dg = 0,$$

identically by the definition of $\bar{x} = \int_{\mathscr{F}} x_g\,dg$. Because $\partial \bar{x}/\partial k^A = q(\theta+\xi)$ does not depend on $g$, the GE feedback integral factors as

$$\int_{\mathscr{F}} \frac{\partial \pi_g}{\partial \bar{x}}\,\frac{\partial \bar{x}}{\partial k^A}\,dg = q(\theta+\xi)\cdot \int_{\mathscr{F}} \frac{\partial \pi_g}{\partial \bar{x}}\,dg = 0,$$

and this holds in expectation across realizations of $\theta$ and $\xi$ as well. The planner's first-order condition therefore reduces to $\partial \pi^A/\partial k^A|_{\bar{x}\text{ fixed}} = 0$, which is exactly the Nash first-order condition (P1.2) for the AI firm. The same argument applies to the custom loading rule: the planner's first-order condition with respect to $k^C$ reduces to $\partial \pi^C/\partial k^C|_{\bar{x}\text{ fixed}} = 0$, identical to (P1.3). Therefore for every fixed $q$, the Nash loadings $(k^A(q), k^C(q))$ are also planner-optimal under $W_1$.

Given the team-optimality of the loadings within each $q$, the envelope theorem applied to the planner's outer problem in $q$ yields

$$\frac{d\mathscr{W}_1}{dq} = \pi^A(q) - \pi^C(q) + \int_{\mathscr{F}} \frac{\partial \pi_g}{\partial \bar{x}}\,\frac{\partial \bar{x}}{\partial q}\,dg\bigg|_{(k^A,k^C)\text{ at optimum}}.$$

A change in $q$ does shift the aggregate, $\partial \bar{x}/\partial q = (k^A - k^C)\theta + k^A\xi$, which is again constant across $g$. The same cross-section identity therefore drives the GE term to zero: $\int_{\mathscr{F}} (\partial \pi_g/\partial \bar{x})\,dg = 0$ multiplied by a $g$-independent factor returns zero. Hence

$$\frac{d\mathscr{W}_1}{dq} = \pi^A(q) - \pi^C(q) = \Delta\Pi(q).$$

The market equilibrium condition $q^* = F(\Delta\Pi(q^*))$ then coincides with the planner's first-order condition $F^{-1}(q^*) = d\mathscr{W}_1/dq$, so $q^* = q_1^{soc}$. □

## A.5 Proof of Proposition 4

The proof has three steps: first, derive the wedge under $W_3$ as $\gamma \cdot d/dq[K_\xi(q)^2\sigma_\xi^2]$ using team efficiency under $W_1$; second, prove that $K_\xi(q)^2$ is strictly increasing in $q$ on $[0,1]$ by an explicit identity for $\Delta(q) - q\Delta'(q)$; third, conclude $q^* > q_3^{soc}$ in any interior adoption equilibrium.

*Step 1: derivation of the wedge.* The welfare specification $W_3$ decomposes as

$$\mathscr{W}_3(q) = \mathscr{W}_1(q) - \gamma K_\xi(q)^2\,\sigma_\xi^2.$$

By Proposition 3, $d\mathscr{W}_1/dq = \Delta\Pi(q)$ throughout $(0,1)$. Differentiating $W_3$ termwise in $q$,

$$\frac{d\mathscr{W}_3}{dq} = \Delta\Pi(q) - \gamma\frac{d}{dq}\left[K_\xi(q)^2\,\sigma_\xi^2\right].$$

The wedge between the private adoption incentive and the planner's marginal welfare is therefore

$$\Delta\Pi(q) - \frac{d\mathscr{W}_3}{dq} = \gamma\frac{d}{dq}\left[K_\xi(q)^2\,\sigma_\xi^2\right]. \tag{P4.1}$$

The strict positivity of the right-hand side is the content of Step 2.

*Step 2: monotonicity of $K_\xi(q)^2$.* From Proposition 1, $k^A(q) = \alpha_\theta^A(1+\beta)/\Delta(q)$ where $\Delta(q)$ is the strictly positive denominator established in the proof of that proposition. Hence

$$K_\xi(q) = q\,k^A(q) = \alpha_\theta^A(1+\beta)\,\frac{q}{\Delta(q)}.$$

Differentiating $K_\xi^2 = [\alpha_\theta^A(1+\beta)]^2 \cdot q^2/\Delta(q)^2$ in $q$,

$$\frac{d}{dq}\left[K_\xi(q)^2\right] = 2\left[\alpha_\theta^A(1+\beta)\right]^2 \cdot \frac{q\left[\Delta(q) - q\Delta'(q)\right]}{\Delta(q)^3}.$$

Since $\Delta(q) > 0$ and $q \geq 0$, the sign of $dK_\xi^2/dq$ is governed by the sign of $\Delta(q) - q\Delta'(q)$. Direct algebraic expansion of $\Delta$ from its defining product $\Delta(q) = (1+\beta-\beta q\rho_A)(1+\beta-\beta(1-q)\alpha_\theta^C) - \alpha_\theta^A\alpha_\theta^C\beta^2 q(1-q)$, followed by computation of $\Delta'(q)$ and the combination $\Delta - q\Delta'$, yields the identity

$$V_A V_C\left[\Delta(q) - q\Delta'(q)\right] = \beta^2\,\sigma_\theta^2\,\sigma_\xi^2\,q^2 \;+\; (1+\beta)\,V_A\left(\sigma_\theta^2 + (1+\beta)\,\sigma_C^2\right), \tag{P4.2}$$

where the symbolic computation underlying (P4.2) is in the supplementary file `W3_analytic.py`. The identity exhibits the right-hand side as a sum of two manifestly nonnegative terms: $\beta^2\,\sigma_\theta^2\,\sigma_\xi^2\,q^2 \geq 0$ for any real $\beta$ and $q$, and $(1+\beta)V_A(\sigma_\theta^2 + (1+\beta)\sigma_C^2) > 0$ for any $\beta > -1$ since both $V_A > 0$ and $\sigma_\theta^2 + (1+\beta)\sigma_C^2 > 0$. The constant term is strictly positive uniformly in $q$, so $\Delta(q) - q\Delta'(q) > 0$

throughout $[0,1]$. Combined with $V_A V_C > 0$ and $\Delta(q)^3 > 0$, this gives

$$\frac{d}{dq}\left[K_\xi(q)^2\right] > 0 \quad \text{for all } q \in (0,1],\, \beta > -1, \tag{P4.3}$$

with the inequality holding strictly when $q > 0$ because the multiplicative factor of $q$ in the numerator vanishes only at $q = 0$.

*Step 3: over-adoption.* Combining (P4.1) and (P4.3) with $\gamma > 0$ and $\sigma_\xi^2 > 0$,

$$\Delta\Pi(q) - \frac{d\mathscr{W}_3}{dq} = \gamma\sigma_\xi^2 \frac{d}{dq}\left[K_\xi(q)^2\right] > 0 \quad \text{for all } q \in (0,1].$$

Hence $\Delta\Pi(q) > d\mathscr{W}_3/dq$ throughout the interior. In any interior adoption equilibrium $q^*$, the market-clearing condition is $F^{-1}(q^*) = \Delta\Pi(q^*)$, while the planner's first-order condition under $W_3$ is $F^{-1}(q_3^{soc}) = d\mathscr{W}_3/dq|_{q_3^{soc}}$. Since $F^{-1}$ is strictly increasing (the type-distribution of the threshold parameter is continuous with full support), the inequality $\Delta\Pi(q^*) > d\mathscr{W}_3/dq|_{q^*}$ implies that $F^{-1}$ is too high at $q^*$ relative to its value at the planner's solution: $q^* > q_3^{soc}$. The result is independent of the sign of $\beta$ and of the specific value of $\gamma$.

*Magnitude.* The wedge magnitude is governed directly by the right-hand side of (P4.1), which through identity (P4.2) can be expressed in closed form as

$$\gamma\sigma_\xi^2 \frac{dK_\xi^2}{dq} = 2\gamma\sigma_\xi^2 \left[\alpha_\theta^A(1+\beta)\right]^2 \cdot \frac{q\left[\beta^2\sigma_\theta^2\sigma_\xi^2 q^2 + (1+\beta)V_A(\sigma_\theta^2 + (1+\beta)\sigma_C^2)\right]}{V_A V_C \Delta(q)^3}.$$

The wedge is therefore proportional to $\gamma$ and to $\sigma_\xi^2$, vanishes at $q = 0$ (where no AI processing yet enters the aggregate), and is increasing in $|\beta|$ through the explicit $\beta^2 q^2$ term in the numerator. The numerator is manifestly nonnegative: the bracket $\beta^2\sigma_\theta^2\sigma_\xi^2 q^2 + (1+\beta)V_A(\sigma_\theta^2 + (1+\beta)\sigma_C^2)$ is a sum of two nonnegative terms for any $\beta > -1$. The denominator $\Delta(q)^3$ is strictly positive on $[0,1]$ by the proof of Proposition 1. Hence the wedge is strictly positive at every $q \in (0,1]$ and for every $\beta > -1$. The *sign* of the wedge — and consequently the qualitative direction $q^* > q_3^{soc}$ — is therefore independent of $\mathrm{sgn}(\beta)$. The *magnitude* of the wedge, however, is not invariant under sign-reversal of $\beta$: the denominator $\Delta(q)$ depends on $\beta$ through factors of the form $1 + \beta - \beta q \rho_A$ and $1 + \beta - \beta(1-q)\alpha_\theta^C$ that are not symmetric in $\beta \mapsto -\beta$, and the numerator carries an explicit $(1+\beta)$ in addition to the squared factors. Direct numerical evaluation confirms that, for fixed values of the other primitives and fixed $q$, the wedge at $\beta = +b$ generically differs from the wedge at $\beta = -b$. This is the precise content of the over-adoption result: positivity of the wedge holds globally in $\beta > -1$, while its size varies with $\beta$. □

*Verification.* The closed-form expressions in equations (P2.1), (P2.2), (P3.3), (P3.4), and (P4.2) have been verified by symbolic computer algebra and by numerical evaluation across $5 \times 10^5$ random draws from the admissible parameter region $\{(q,\beta,\sigma) : q \in [0,1],\, \beta \in (-1,50),\, \sigma \in (0.001,100)^4\}$, with no exceptions to the strict-positivity claim of identity (P4.2) or to monotonicity of $K_\xi(q)^2$ in $q$. Supplementary computational scripts are available from the authors.