IJEC
International Journal of Economics
https://ejournal.ipinternasional.com/index.php/ijec
e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

The Autonomy Externality: A Welfare-Economic Model of Agentic Artificial Intelligence, Correlated Failure, And Optimal Assurance Policy

Kwan Hong TAN
Singapore University of Social Sciences, Singapore
Correspondent Author: khtan055@suss.edu.sg
Article history: Received July 13 2026; revised August 08, 2026; accepted August 20, 2026

ABSTRACT

Agentic artificial intelligence is shifting automation from prediction and recommendation to autonomous execution. This transition creates an economic externality when firms capture productivity gains from delegation but do not bear the full cost of failures transmitted through shared models, cloud infrastructure and downstream markets. This study develops a welfare economic model in which firms jointly choose autonomy and assurance, while residual failures are correlated across organizations. The benchmark model derives private and social equilibria and shows that correlated external damage produces excessive autonomy, insufficient assurance intensity and excessive residual risk. A residual-risk levy can decentralize the social optimum by pricing the remaining expected external damage after assurance. The paper then extends the model to a heterogeneous two-tier artificial intelligence value chain containing dominant foundation model providers and downstream API users. In a stylized calibration with five upstream providers and ninety-five downstream users, the levy-equivalent coefficient for a representative dominant provider is approximately four times that of a downstream integrator because the provider has a greater marginal system contribution; the result is driven by network centrality rather than firm size alone. The second extends the strategic overstatement of assurance model. Truthful reporting requires expected audit penalties to exceed the marginal levy savings from misreporting, which establishes a formal complementarity between residual-risk pricing, independent third-party audits, sectoral peer review, and regulatory meta-audits. Monte Carlo and sensitivity analyses confirm the directional robustness of core welfare wedges. Therefore, this study offers a unified framework for an innovation-preserving AI policy that differentiates firms by systemic contribution and makes verified assurance an economically valuable form of risk-reducing capital.

Keywords: Agentic Artificial Intelligence, AI Auditing, Assurance Investment, Correlated Failure, Foundation Models, Systemic Risk, Welfare Economics

1. INTRODUCTION

Artificial intelligence (AI) is entering a new economic phase. Earlier waves of business adoption were dominated by forecasting, classification, recommendations, and content generation. Agentic artificial intelligence extends these functions by allowing systems to plan, select tools, transact, update records, coordinate workflows, and execute decisions with limited human intervention. Therefore, the economic object is no longer only a productivity-enhancing input. It is also a delegated decision-maker whose actions can propagate through technical, organizational and market networks.

Existing economic debates have concentrated on aggregate productivity, occupational exposure, task substitution, skill complementarity, and the distribution of gains from artificial intelligence. These questions remain central to the study. Acemoglu (2024) cautions that macroeconomic gains depend on the share of tasks that artificial intelligence can perform profitably and reliably, while Cazzaniga et al. (2024) emphasize the broad labor market exposure of advanced economies. Brynjolfsson, Rock, and Syverson (2021) show that the productivity effects of general-purpose technologies can be delayed because firms must accumulate complementary intangible capital. These perspectives explain why adoption does not automatically results in measured growth. However, they do not fully capture the external costs incurred when many firms delegate consequential actions to systems that share providers, foundation models, cloud infrastructure, data pipelines, benchmarks, and design conventions.

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

The central problem is the autonomy externality. A firm captures the private value of faster decisions, lower labor costs, extended operating hours, and scalable services. It also bears direct losses due to errors. However, it may not bear the full expected cost of downstream harm when failures affect counterparties, customers, financial stability, information integrity, or public services. The wedge becomes larger when the failures are correlated. Common model providers, common vulnerabilities, synchronized optimization objectives, and imitation of similar deployment practices can convert individually rational adoption into a collectively fragile dependence. This concern is consistent with financial stability analyses that identify third-party concentration, market correlations, cyber risk, and model risk as potential channels through which artificial intelligence can amplify systemic vulnerabilities (Aldasoro et al., 2024; Financial Stability Board, 2024, 2025).

This study develops a tractable welfare economic model in which firms choose both autonomy and assurance. Assurance includes evaluation, monitoring, red teaming, human escalation, access controls, staged execution, rollback, incident response, and other controls that reduce residual risk. The benchmark derives closed-form private and social choices under symmetric equicorrelation structures. Two extensions relax the benchmark in economically important directions. The first introduces a two-tier value chain containing upstream foundation model providers and downstream API integrators, allowing levy incidence to vary with network centrality and cross-layer dependence. The second introduces assurance misreporting and costly verification, making third-party audits and sectoral peer reviews endogenous complements to residual-risk pricing.

This analysis makes seven contributions. First, it formalizes agentic artificial intelligence as an externality problem, rather than only a technology adoption problem. Second, it derives closed-form private and social choices for autonomy and assurance. Respectively. Third, it identifies the cross-firm failure correlation as a multiplier of the governance wedge. Fourth, it derives an assurance-adjusted residual-risk levy that aligns private incentives with the social marginal cost. Fifth, it extends the framework to heterogeneous upstream and downstream firms and shows that efficient levy incidence follows marginal system contribution rather than uniform firm classification. Sixth, it models the strategic overstatement of assurance and derives a truth-telling condition linking audit probability, sanctions, and marginal external damage. Seventh, it combines analytical calibration, heterogeneous simulation, centrality sensitivity, and large-parameter Monte Carlo analysis to show how risk pricing, diversification, and verification can function as complementary policy instruments.

This study is theoretical and analytical. The numerical results are stylized calibrations generated from the stated equations and parameter ranges. These are not estimates from observed firms and should not be interpreted as universal policy coefficients. The purpose is to identify mechanisms, comparative statics, and empirically testable predictions that can guide future measurements and sector-specific calibrations.

2. LITERATURE REVIEW AND THEORETICAL POSITIONING

2.1 Artificial intelligence, productivity, and organizational complements

Economic research on automation shows that technology can substitute for some tasks while creating new tasks, raising demand, and complementing labor in others (Autor, 2015; Acemoglu & Restrepo, 2018). Evidence on industrial robots indicates that technology can increase productivity while also producing heterogeneous employment and wage effects across local labor markets (Acemoglu & Restrepo, 2020; Graetz & Michaels, 2018). Artificial intelligence intensifies this ambiguity because it applies to cognitive and communicative tasks that were previously regarded as resistant to automation.

A key lesson is that technical capability and realized productivity are not equivalent to each other other. Firms must redesign processes, develop data assets, train workers, adjust incentives, and build managerial capacity to achieve this. The productivity J-curve captures the possibility that measured productivity initially weakens while complementary investments accumulate (Brynjolfsson et al., 2021). Tan (2025a) similarly frames the digital productivity paradox as a problem of intangible capital, institutional capacity, and organizational transformation. Tan (2026a) extends this logic to agentic automation by arguing that enterprise value depends on a governance frontier: additional autonomy creates value only when marginal productivity exceeds marginal governance and residual-risk costs.

This study adds an inter-firm dimension. Complementary investments affect not only a firm's private productivity but also the external risk it imposes on others. Therefore, assurance capital is partly analogous to pollution abatement and partly analogous to financial capital buffers. It lowers the expected private loss, but it can also reduce network-wide harm.

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

2.2 Externalities, systemic risk, and correlated behavior

The classical theory of externalities establishes that decentralized choices are inefficient when private decision-makers do not face the full social marginal cost of their actions (Pigou, 1920; Baumol & Oates, 1988). Corrective taxation can restore efficiency when the tax reflects the marginal external damage (Sandmo, 1975). The challenge in digital systems is to identify the correct tax base. A levy on artificial intelligence spending, computing, or revenue would penalize low-risk uses and might discourage beneficial innovation. The economically relevant object is the residual risk after assurance, not technology use itself.

Systemic-risk economics provides a second foundation. Financial institutions may choose individually rational portfolios that create correlated exposures and socially excessive fragility. Measures such as CoVaR and systemic expected shortfall distinguish an institution's stand-alone risk from its contribution to system-wide losses (Adrian & Brunnermeier, 2016; Acharya et al., 2017). Wagner (2010) shows that diversification can paradoxically increase systemic fragility when institutions converge to similar portfolios. The same logic applies to artificial-intelligence supply chains. A firm may diversify internal tasks across applications, while all applications rely on the same foundation model or cloud provider. Therefore, apparent operational diversification can coexist with common-mode dependence.

Information cascades and algorithmic interactions supply a third mechanism. When organizations imitate perceived best practices, use common benchmarks, or adopt similar decision rules, correlated behavior can emerge even without direct coordination (Bikhchandani, Hirshleifer, & Welch, 1992). Algorithmic pricing research demonstrates that learning agents can generate market outcomes that differ materially from conventional static strategy models (Calvano, Calzolari, Denicolo, & Pastorello, 2020). Flash-crash research similarly shows how automated strategies, speed, and feedback can amplify local disturbances (Kirilenko, Kyle, Samadi, & Tuzun, 2017). These studies imply that agentic systems should be assessed not only by average accuracy but also by dependence structure, action speed, reversibility, and propagation.

Recent policy analyses increasingly recognize these channels. The Financial Stability Board (2024, 2025) highlights third-party dependencies, model risk, cyber risk, market correlations, and governance gaps as significant concerns. The Bank of England (2025) examines how widespread artificial intelligence adoption can affect financial stability through concentration and correlated behavior. The NISTArtificial Intelligence Risk Management Framework emphasizes continuous risk governance across the system lifecycle (National Institute of Standards and Technology, 2023). These frameworks are operationally important, but they do not specify the welfare wedge or the price that should be attached to residual systemic exposure.

2.3 Agentic artificial intelligence and the governance frontier

Agentic systems create a particularly sharp version of the problem because they transform predictions into actions. Recommendations can be reviewed before use. An agent may instead submit a payment, modify an entitlement record, change the price, publish information, or trigger another agent. Tan (2026b) describes the resulting supervisory economy, in which human roles shift from direct execution to oversight of artificial actors. Tan (2026c) proposed governance-aware control architectures that gate actions according to risk, permissions, and accountability. These contributions establish the organizational and engineering mechanisms. The present study asks a prior economic question: what level of autonomy and assurance would a firm choose privately, and how does that differ from the social optimum when failures are correlated?

The theoretical answer depends on the separation of autonomy, assurance, and residual risk. Autonomy is a productive delegation. Assurance is a costly risk-reduction effort. Residual risk is the remaining exposure after assurance. This separation avoids the false choice between innovation and prohibition. It also creates a policy target that rewards effective governance, rather than mere documentation.

2.4 Foundation-model market structure and assurance verification

The emerging artificial intelligence value chain is economically diverse. Frontier and foundation-model development requires large fixed investments in computing, data, engineering, and distribution, whereas marginal deployment through APIs can serve many downstream applications. This creates scale economies and potential concentration in the upstream layer, even when downstream application markets remain competitive. Vipra and Korinek (2023) emphasized the concentration tendency of highly capable foundation models, while Schrepel and Pentland (2025) analyzed competition among model ecosystems and the strategic importance of access to key inputs. OECD (2025) further notes that downstream firms using closed models through APIs can become dependent on provider terms, interoperability, and switching conditions. The implication for the present model is that one firm's residual risk can have a different marginal system contribution from another firm's residual risk, even when both use the same nominal level of autonomy.

This distinction is increasingly reflected in policies. The European Union's AI Act distinguishes providers of general-purpose AI models from downstream providers that integrate such models into AI

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

systems, and the European Commission's 2026 guidance clarifies provider obligations concerning documentation and information required by downstream actors (European Parliament and Council of the European Union, 2024; European Commission, 2026). From an economic perspective, upstream firms are not inherently harmful. Rather, a provider that is embedded across many downstream systems can be more systemically central, so a change in its residual risk may alter expected losses across a larger set of organizations.

The second literature concerns the credibility of assurance claims. Raji et al. (2020) developed an end-to-end framework for internal algorithmic auditing, while Raji et al. (2022) showed that meaningful third-party oversight depends on institutional design, auditor access, independence, standards, and incentives rather than the mere existence of an audit label. The NIST AI Risk Management Framework likewise emphasizes documented measurement, monitoring, and governance processes (National Institute of Standards and Technology, 2023). These insights create an economic verification problem for a residual-risk levy: once a lower reported residual risk reduces a firm's payment, the firm has an incentive to exaggerate assurance unless the verification is sufficiently credible. Sections 3.7 and 5.5 formalize and operationalize complementarity.

3. MODEL

3.1 Economic environment and assumptions

Consider a benchmark economy with N symmetric firms indexed by i = 1, ..., N. Each firm deploys an agentic artificial intelligence system and chooses an autonomy level a_i in the closed unit interval. A higher value indicates that a larger share of economically consequential tasks can be initiated or completed without contemporaneous human approval. The firm also chooses assurance effort g_i greater than or equal to zero. Assurance is measured in normalized effectiveness units rather than monetary expenditure, allowing different technical controls to be represented by a common risk reduction index. Firm-specific notation is retained because Section 3.6 relaxes symmetry and introduces an upstream-downstream (U-D) value chain.

The private gross benefit of autonomy is concave. Parameter b_i is the marginal benefit at the origin, and q_i captures diminishing returns, congestion, coordination cost, or the declining quality of tasks at the margin. Assurance has a convex cost with coefficient c_i. The direct expected loss from residual risk has coefficient d_i. Assurance productivity γ_i maps the governance effort into risk reduction. For an interior solution, the residual risk is defined as follows:

$$r_i = a_i - \gamma_i g_i, \qquad r_i \geq 0 \tag{1}$$

This linear technology provides a local approximation. This suggests that assurance mitigates the impact of autonomy on effective exposure. The specification can represent evaluation coverage, monitoring quality, human escalation, access restriction, transaction staging, rollback capacity, and insurance-linked verification. The firm's private payoff is

$$\pi_i = b_i a_i - \frac{q_i}{2} a_i^2 - \frac{c_i}{2} g_i^2 - \frac{d_i}{2} r_i^2 \tag{2}$$

The model intentionally includes a private-loss term. Firms are not assumed to ignore risks. The inefficiency arises because they internalize only d_i, not the full damage transmitted through the economic network.

Table 1. Model symbols and economic interpretation

Symbol	Interpretation	Domain
a_i	Autonomy delegated to the agentic system	0 to 1
g_i	Assurance effort or governance effectiveness	Non-negative
r_i	Residual operational risk after assurance	Non-negative
b_i	Marginal private benefit from initial autonomy	Positive
q_i	Curvature of autonomy benefit	Positive
c_i	Marginal cost curvature of assurance	Positive
d_i	Firm-internal expected-loss coefficient	Positive
γ_i	Productivity of assurance in reducing residual risk	Positive
χ	Scale of network-wide external damage	Positive
ρ	Cross-firm correlation of residual failures	0 to 1

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

3.2 Network externality and failure correlation

Let r be an N by 1 vector of residual risks. Social external loss is quadratic in residual exposure and depends on a positive semidefinite dependence matrix Ω. The scale parameter χ converts the normalized risk into the expected social damage.

$$L(\mathbf{r}) = \frac{\chi}{2N}\mathbf{r}^{\top}\Omega\mathbf{r} \tag{3}$$

For analytical transparency, we assume an equicorrelation structure. Every diagonal element of Ω equals one, and every off-diagonal element equals ρ. Under symmetry, $r_i = r$ for all firms, and the per-firm external loss becomes one-half the coefficient in Equation (4) multiplied by the residual risk squared, where:

$$h_N(\rho) = \chi\,\frac{1 + (N-1)\rho}{N} \tag{4}$$

This expression separates idiosyncratic and common risk. When ρ equals zero, the external loss per firm is diluted by N because failures are independent. When ρ equals one, all firms fail together, and the external-risk coefficient equals χ. For a large economy, even a modest positive correlation dominates the idiosyncratic component. The derivative in Equation (7) is positive and approaches χ as N increases.

Correlation has a broad interpretation. It can arise from common foundation models, shared cloud and identity infrastructure, common data vendors, replicated prompts, synchronized objective functions, shared regulatory blind spots or common exposure to adversarial techniques. Therefore, the parameter captures economic dependence even when firms are legally and organizationally separated.

3.3 Private equilibrium

Under symmetry and interiority, a representative firm maximizes its private payoff while considering external loss as given. Substituting the residual risk into the payoff and solving the first-order conditions yields:

$$a^P = \frac{b}{q + \dfrac{dc}{c + d\gamma^2}}, \qquad g^P = \frac{d\gamma}{c + d\gamma^2}a^P, \qquad r^P = \frac{c}{c + d\gamma^2}a^P \tag{5}$$

The superscript P denotes a private solution. The assurance-to-autonomy ratio rises with the firm's internal loss coefficient and assurance productivity but falls with assurance cost. A firm with weak liability, limited reputational exposure, or the expectation of public rescue has a low effective d and, therefore, chooses little assurance relative to autonomy.

Private equilibrium is not unambiguously characterized by excessive technology use in all possible models. In the present setting, the key distortion is the omission of the network damage. The comparison with the social optimum below shows how this omission affects both the level and composition of the investment.

3.4 Social optimum

A social planner values the same productive benefit and private costs but internalizes the per-firm external loss. Therefore, the planner replaces the private risk coefficient with the sum of the private and external risk coefficients. The symmetric social optimum is as follows:

$$a^S = \frac{b}{q + \dfrac{(d+h)c}{c + (d+h)\gamma^2}}$$
$$g^S = \frac{(d+h)\gamma}{c + (d+h)\gamma^2}a^S, \qquad r^S = \frac{c}{c + (d+h)\gamma^2}a^S \tag{6}$$

Here, h denotes the external risk coefficient in Equation (4), and the superscript S denotes the social solution. The formulation yields three propositions.

Proposition 1. The autonomy and assurance wedge

If h is positive and the solution is interior, the private equilibrium has higher autonomy, a lower assurance-to-autonomy ratio, and higher residual risk than the social optimum: private autonomy exceeds social autonomy, the private assurance-to-autonomy ratio is lower, and the private residual risk is higher.

Proof. The effective risk coefficient is larger for the social planner than for the private firm. The autonomy expression in Equation (5) therefore has a smaller value after the external cost is internalized. The assurance-to-autonomy ratio increases with the effective risk coefficient, while the residual-risk ratio decreases. The stated inequalities follow. Q.E.D.

roposition 2. Correlation amplifies the governance wedge

For N greater than one, the marginal external risk coefficient is strictly increasing in ρ:

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

$$\frac{\partial h_N(\rho)}{\partial \rho} = \chi \frac{N-1}{N} > 0 \tag{7}$$

Consequently, a higher cross-firm failure correlation lowers socially optimal autonomy, raises optimal assurance intensity, and lowers socially optimal residual risk. This result explains why a control that is adequate for an isolated experimental deployment may be insufficient after an entire sector converges on the same provider or architecture.

Proposition 3. A residual-risk levy decentralizes the social optimum

Suppose each firm is charged a quadratic levy based on measured residual risk:

$$T_i = \frac{\tau}{2} r_i^2 \tag{8}$$

Setting the levy coefficient τ equal to the external risk coefficient in Equation (4) causes the firm's private objective to contain the effective coefficient d + h. Its first-order conditions then coincide with those of the social planner. The levy is assurance-adjusted: a firm can reduce its payment by lowering its autonomy, improving its assurance, or both. Therefore, it avoids treating all artificial intelligence use as equally harmful.

3.5 Dynamic assurance depreciation

Agentic systems and their environments change over time. Model updates, new tool permissions, data drift, adversarial adaptation, staff turnover, and organizational changes can erode previously validated controls. To represent this process, let effective assurance evolve according to

$$g_{i,t+1} = (1-\delta_g) g_{i,t} + I_{i,t}, \qquad 0 \le \delta_g \le 1 \tag{9}$$

The depreciation parameter captures the control decay, whehreas the investment term represents the new assurance expenditure. A static certification regime is equivalent to assuming that the depreciation parameter is zero between audits. This assumption is implausible for rapidly changing agentic systems. Dynamic extension implies that a residual-risk levy should be based on current verified assurance rather than historical compliance status.

This mechanism connects the model to the governance half-life concept developed by Tan (2026d), in which the effective coverage of an artificial intelligence control declines after validation. In economic terms, assurance is a depreciating, intangible asset. Firms underinvest in their renewal when the resulting external risk is not priced.

3.6 Heterogeneous firms and the upstream-downstream AI value chain

The vector formulation in Equation (3) already permits heterogeneous firms through firm-specific parameters and a dependence matrix. The symmetry assumption was imposed only to obtain a compact, closed form. To make the reviewer-relevant market structure explicit, divide the economy into n_F upstream foundation-model providers and n_D downstream API users, with N = n_F + n_D. Let t be a number of {F, D}. Each type can have distinct productivity, assurance cost, internal liability, and assurance effectiveness. Let ω_t be a systemic centrality weight that captures how strongly a unit of residual risk propagates through dependent applications and economic relationships. Within-layer correlations are ρ_{FF} and ρ_{DD}, while ρ_{FD} measures the cross-layer common dependence. For compactness, we define $A_F = 1 + (n_F - 1)\rho_{FF}$ and $A_D = 1 + (n_D - 1)\rho_{DD}$ as within-layer amplification factors, $C_{FD} = \omega_F \omega_D \rho_{FD}$ as cross-layer coupling, and $B_F = n_F \omega_F^2 A_F$, $B_D = n_D \omega_D^2 A_D$, and $B_{FD} = n_{Fn}\_DC_{FD}$. The heterogeneous external loss function is then expressed as

Under type symmetry within each layer, aggregate external loss is:

$$L_H = \frac{\chi}{2N}(B_F r_F^2 + B_D r_D^2 + 2B_{FD} r_F r_D). \tag{10}$$

The first two terms capture idiosyncratic and within-layer correlated exposures. The third captures the failure transmission between foundation model providers and downstream applications. A high ω_F can represent a provider whose model, API, safety filter, tool-calling interface, or cloud dependency is used across many firms. A downstream firm can also have a high weight when it operates critical infrastructure or a highly connected platform; thus, centrality is an economic property rather than a legal label.

The marginal external damage created by a change in each type's residual risk is

$$M_F(\mathbf{r}) = \frac{\chi}{N}(\omega_F^2 A_F r_F + n_D C_{FD} r_D). \tag{11}$$

$$M_D(\mathbf{r}) = \frac{\chi}{N}(\omega_D^2 A_D r_D + n_F C_{FD} r_F). \tag{12}$$

These expressions show why a uniform levy is generally inefficient in heterogeneous AI ecosystems. The marginal social cost of residual risk depends on the centrality, within-layer correlation, cross-layer correlation, number and risk of dependent firms, and the current state of the network. Therefore, a provider

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

can face a larger marginal charge even if its incident probability is lower because its rare failures propagate more widely.

An exact network-contribution levy that treats other firms' residual risks as given can be written as

$$T_i^H(\mathbf{r}) = \frac{\chi}{2N}\Omega_{ii}r_i^2 + \frac{\chi}{N}\sum_{j\neq i}\Omega_{ij}\, r_i r_j. \tag{13}$$

The derivative of Equation (13) with respect to r_i equals χ(Ωr)_i/N, which is the firm's marginal contribution to network loss. Cross terms can cause total levy revenue to differ from realized external loss, but this is not a welfare problem if revenue is recycled through lump-sum transfers or used for risk-reduction of public goods. The purpose of the levy is to reproduce the correct marginal incentive, not to make tax revenue mechanically equal to the expected damage.

For comparison across firm types, define the local levy-equivalent coefficient as:

$$\tau_F = \frac{M_F(\mathbf{r})}{r_F}, \qquad \tau_D = \frac{M_D(\mathbf{r})}{r_D}. \tag{14}$$

Proposition 4. Network-incidence principle

For any two firms i and k with positive residual risk, the efficient Levy-equivalent coefficient is higher for i if and only if (Ωr)_i/r_i exceeds (Ωr)_k/r_k. Therefore, firm size or provider status alone does not determine the efficiency burden. A dominant foundation model provider faces a higher rate only when its residual risk makes a larger marginal contribution to the correlated system loss. Conversely, a critical downstream operator can rationally face an equally high or higher rate when its network position warrants it.

This result raises potential equity concerns. A differentiated levy need not be an incumbent tax or small-firm subsidy. This is a marginal contribution rule. Dominant providers are likely to score highly when their services are common dependencies, whereas ordinary downstream API users will often face lower per-firm coefficients. The numerical extension in Section 4.4 illustrates this.

3.7 Assurance verification, strategic reporting, and audit incentives

Residual mrisk pricing creates a new information problem. The regulator or insurer observes a reported assurance level and associated evidence, but the firm may know more about the real effectiveness of its control. Let x_i denote the overstatement of assurance, so the reported assurance and reported residual risk are

$$\hat{g}_i = g_i + x_i, \qquad \hat{r}_i = \max\{0, a_i - \gamma_i \hat{g}_i\}, \qquad x_i \geq 0. \tag{15}$$

If the levy is based on reported rather than true residual risk, an increase in x_i lowers the assessed payment while leaving the true risk unchanged. Let p_i be the probability that a material overstatement is detected, and let f_i be the marginal sanction per unit of detected overstatement, including monetary penalties, loss of safe-harbor status, higher future audit intensity, procurement exclusion, or insurance consequences. Since a marginal unit of assurance overstatement reduces the reported residual risk by γ_i, the local truth-telling condition is:

$$p_i f_i \geq \gamma_i M_i(\mathbf{r}). \tag{16}$$

Equation (16) is of economic importance. The incentive to misreport is stronger when the firm's marginal external damage M_i is larger because credible assurance produces a larger levy reduction. Therefore, high-centrality providers require stronger verification, larger expected penalties, or both. A low-risk downstream user can be subject to proportionately lighter verification without destroying the levy's incentive properties.

The verification process can be layered. Let p_i^A denote the detection probability from an accredited third-party auditor, p_i^P detection through sectoral peer challenge, and p_i^R detection through regulator re-performance or random inspection. Under a stylized conditional-independence assumption, the effective probability that at least one channel detects a material misstatement is

$$p_i^{\text{eff}} = 1 - \left(1 - p_i^A\right)\left(1 - p_i^P\right)\left(1 - p_i^R\right). \tag{17}$$

The independence assumption is only a benchmark because auditors and peers can share methods and blind spots. Nevertheless, Equation (17) clarifies the complementarity. Adding a second credible verification channel increases the effective detection probability and can satisfy Equation (16) without relying exclusively on extreme monetary sanctions. Therefore, third-party audits, peer reviews, and regulatory oversight are not substitutes for levies. They are part of the information infrastructure required for the levy to operate effectively.

4. ANALYTICAL CALIBRATION AND ROBUSTNESS

4.1 Calibration design

Calibration illustrates the model's comparative statics. It does not estimate the structural parameters from the observed firms. The baseline values are b = 1.20, q = 1.00, c = 0.45, d = 0.35, γ = 0.80, χ = 1.40, and N =

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

100. These values generate an interior private solution and allow the effect of ρ to be isolated from the model. Autonomy, assurance, residual risk, and welfare were expressed in normalized units.

Private equilibrium is constant across correlation scenarios because each firm ignores the external term. It produced an autonomy, assurance of 0.404, and residual risk of 0.973, 0.404, and 0.649, respectively. The social optimum changes with ρ because the correlated failure increases the marginal external damage. The per-firm social welfare is computed as follows:

$$W(a,g;r) = ba - \frac{q}{2}a^2 - \frac{c}{2}g^2 - \frac{d + h_N(\rho)}{2}r^2 \quad (18)$$

Table 2. Social optimum across cross-firm failure correlations

ρ	External-risk coefficient	Social autonomy	Social assurance	Residual risk	Welfare gain vs. private
0.00	0.0140	0.968	0.413	0.638	0.01%
0.05	0.0833	0.946	0.451	0.585	0.31%
0.25	0.3605	0.887	0.557	0.441	4.81%
0.50	0.7070	0.844	0.633	0.337	16.50%
0.80	1.1228	0.813	0.688	0.263	40.65%
1.00	1.4000	0.799	0.713	0.229	66.27%

When ρ is zero, the social and private solutions are close because idiosyncratic failure is dispersed across a large economy. At ρ = 0.25, optimal autonomy falls to 0.887, and assurance rises to 0.557. At ρ = 0.80, autonomy, assurance rises to 0.688, and residual risk were 0.813, 0.688, and 0.263, respectively. Relative to the private residual risk of 0.649, this is a reduction of approximately 59.5 percent. The welfare gain from internalization increases nonlinearly because correlated exposure magnifies the marginal value of assurance.

Figure 1 shows the effect of composition. The private autonomy level is insensitive to correlation, but social allocation shifts from delegation toward assurance as common dependence increases. Importantly, the social optimum does not eliminate autonomy in this case. Even at ρ = 1, the calibrated optimum retains an autonomy of approximately 0.799. Therefore, efficient policy constrains and conditions autonomy rather than prohibiting it.

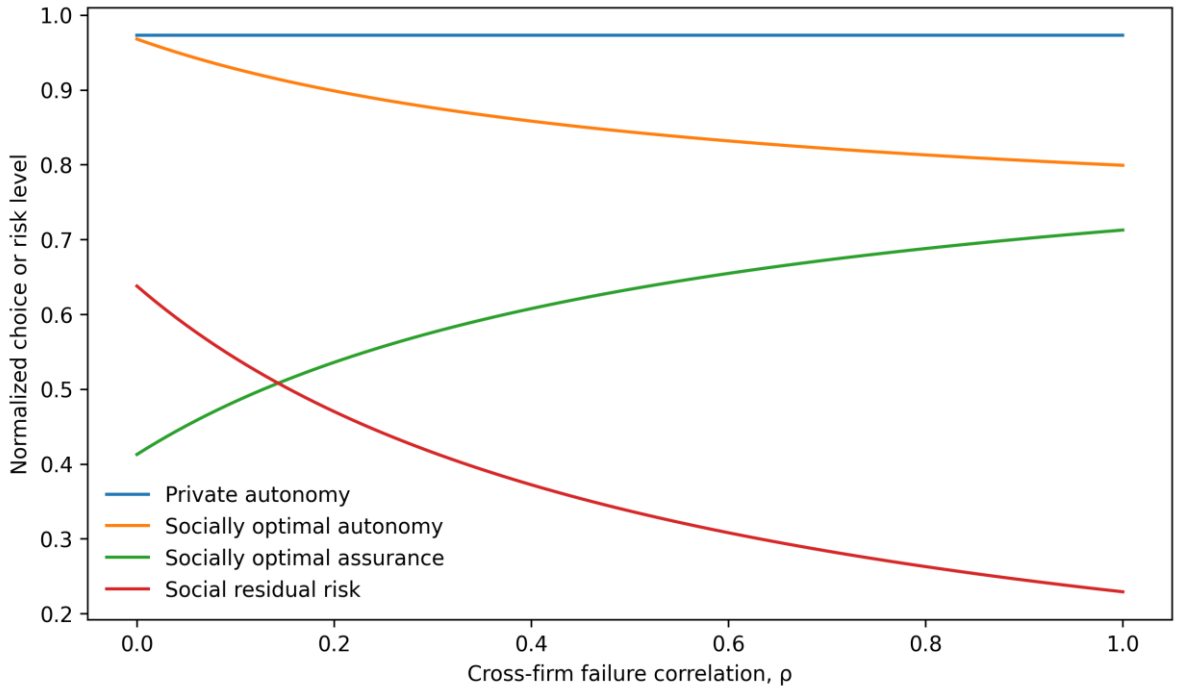

Figure 1. Private and socially optimal choices as failure correlation increases

4.2 Comparison of policy instruments

The policy instruments are compared at ρ = 0.80. The autonomy cap fixes autonomy at the socially optimal level but allows the firm to choose assurance according to its private incentives. The assurance floor fixes assurance at the social level but leaves the firm to choose autonomy, subject to a unit upper bound. The

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

residual-risk levy uses the levy coefficient derived in Proposition 3, which implements the social optimum. Provider diversification is represented by reducing ρ from 0.80 to 0.25, while leaving firm incentives unchanged. The combined package reduces the correlation and applies the calibrated Levy at a lower dependence level.

These comparisons clarify why command-and-control rules can be incomplete. An autonomy cap addresses scale but not weak-assurance. An assurance floor addresses effort but can induce the firm to use its expanded risk capacity for additional autonomy. A residual risk levy prices the interaction between the two choices. Diversification changes the technology of systemic losses.

Table 3. Stylized welfare comparison of policy instruments

Policy	Per-firm welfare	Scenario
Laissez-faire	0.347	Baseline ρ = 0.80
Autonomy cap	0.402	Baseline ρ = 0.80
Assurance floor	0.445	Baseline ρ = 0.80
Residual-risk levy	0.488	Baseline ρ = 0.80
Provider diversification	0.508	ρ reduced to 0.25
Combined package	0.532	ρ reduced to 0.25

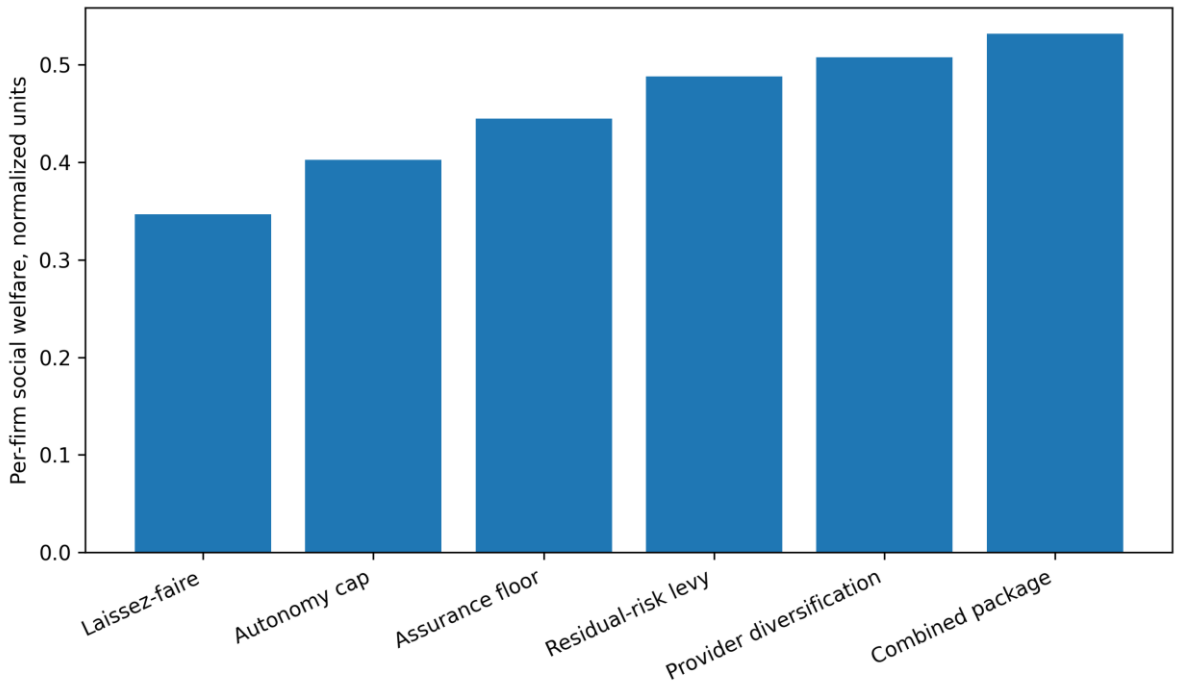

Figure 2. Per-firm social welfare under alternative policy instruments

In the baseline correlation, laissez-faire welfare is 0.347 normalized units. The autonomy cap raises welfare to 0.402 and the assurance floor raises it to 0.445. The residual-risk levy reaches 0.488 because it decentralizes the social optimum. Reducing common dependence to ρ = 0.25 raises welfare, even without changing private incentives. The combined package produces the highest welfare in the comparison because it reduces the source of systemic amplification and internalizes the remaining externality.

The absolute values are parameter-dependent, but the mechanism is general. Pricing residual risk and reducing correlation are complementary. A diversified provider market lowers the external cost coefficient, while the levy ensures that firms still internalize the residual network harm that remains.

4.3 Monte Carlo robustness analysis

A Monte Carlo exercise tests whether the direction of the analytical results depends on baseline calibration. Two hundred thousand independent parameter vectors were generated with b uniformly distributed from 0.50 to 1.40; q from 0.80 to 2.00; c from 0.20 to 1.20; d from 0.10 to 1.00; γ from 0.20 to 0.95; χ from 0.20 to 2.00; N from 20 to 500; and ρ from 0.01 to 1.00. The random seed was set to 20260713. Draws that violated the interior autonomy condition were excluded, leaving 193,446 admissible cases in the final sample.

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

This exercise is a mathematical sensitivity analysis and does not simulate evidence of the real economy. Its role is to verify that the propositions are not artifacts of narrow parameter selection.

Table 4. Monte Carlo robustness results

Indicator	Result
Admissible parameter draws	193,446
Draws with private autonomy above social autonomy	100.0%
Draws with higher social assurance intensity	100.0%
Draws with private residual risk above social residual risk	100.0%
Draws with social welfare above private allocation	100.0%
Median autonomy wedge	10.06%
Median increase in assurance-to-autonomy ratio	51.61%
Median reduction in residual risk	24.92%
Median welfare gain	3.88%
90th percentile welfare gain	37.60%

All admissible draws reproduce the directions of the three central wedges. The median excess private autonomy is approximately 10.1 percent. The median social assurance intensity is 51.6 percent higher than the private ratio, and the median residual-risk reduction is 24.9 percent. The median welfare gain from internalization is 4.0 percent, while the 90th percentile gain is 37.6 percent. The wide upper tail indicates that governance failures can be economically material in high-correlation, high-damage environments, even when average cases appear manageable.

The robustness result is partly structural because the propositions follow the monotonicity. Nevertheless, its practical significance is important. The model does not require extreme parameter values to produce meaningful wedges. What matters is the interaction between assurance cost, assurance effectiveness, external damage, and common dependence is crucial.

4.4 Heterogeneous two-tier simulation: dominant providers and downstream users

The second stylized simulation examines how the levy changes when firms are heterogeneous. The economy contains five foundation-model providers and ninety-five downstream API users. The provider type had $b_F = 1.45$, $q_F = 1.20$, $c_F = 0.60$, $d_F = 0.60$, $\gamma_F = 0.85$, and systemic centrality weight $\omega_F = 2.00$. The downstream type has $b_D = 1.10$, $q_D = 1.00$, $c_D = 0.40$, $d_D = 0.30$, $\gamma_D = 0.75$, and $\omega_D = 1.00$. Network parameters are $\chi = 0.80$, $\rho_{FF} = 0.40$, $\rho_{DD} = 0.15$, and $\rho_{FD} = 0.20$. These values were not empirical estimates. They were chosen to represent a small upstream layer with higher centrality and a large downstream layer with a lower individual propagation weight.

Private choices are computed type-by-type using Equation (5). The social allocation maximizes the aggregate surplus of both layers, net of Equation (10). The network-contribution levy in Equation (13) reproduces the social first-order condition. Table 5 reports the resulting incidences.

Table 5. Heterogeneous two-tier simulation and levy incidence

Metric	Dominant foundation-model provider	Downstream API user
Number of firms	5	95
Systemic-centrality weight, ω_t	2.000	1.000
Private autonomy	0.936	0.908
Social / levy autonomy	0.864	0.867
Private assurance	0.462	0.359
Social / levy assurance	0.586	0.437
Private residual risk	0.544	0.639
Social / levy residual risk	0.365	0.540
Autonomy reduction	7.8%	4.5%
Assurance increase	26.8%	21.5%
Residual-risk reduction	32.8%	15.5%
Levy-equivalent coefficient, τ_t	0.532	0.132
Per-firm network-contribution levy	0.0655	0.0207

The provider's levy-equivalent coefficient is 0.532, compared with 0.132 for the downstream user, with a ratio of approximately 4.04. The efficient adjustment is correspondingly stronger upstream: provider residual risk falls by 32.8 percent, while downstream residual risk falls by 15.5 percent. The per-firm network-contribution levy is also approximately 3.16 times larger for the provider. These differences are not imposed

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

by the firm category; they emerge from the provider's greater centrality and cross-layer propagation. Aggregate downstream payments can still be substantial because there are numerous downstream firms.

Figure 3 varies only the provider centrality weight from 1.0 to 3.0 while re-solving the social allocation at each point. The provider levy-equivalent coefficient rises from approximately 0.200 to 1.091, whereas the representative downstream coefficient remains near 0.13. The sensitivity is nonlinear because provider centrality enters both the within-provider and cross-layer external damage terms. This result supports a risk-based rather than a status-based policy: a provider becomes more heavily charged as its architecture becomes a more consequential common dependency.

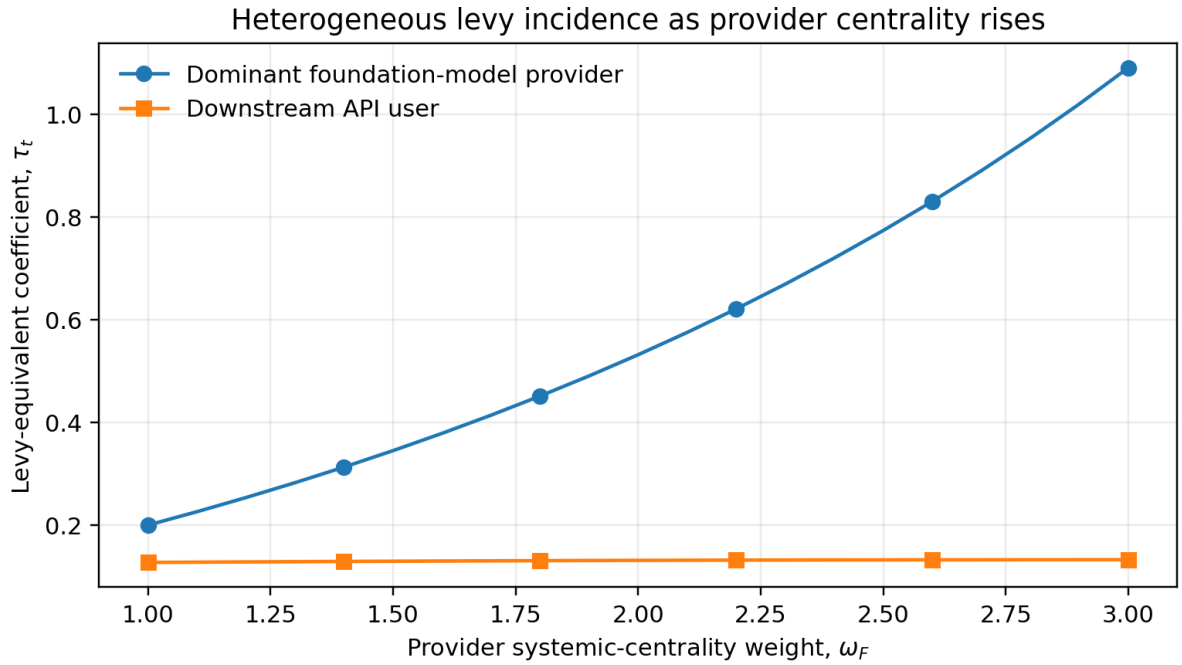

Figure 3. Heterogeneous levy incidence as provider systemic centrality rises

The two-tier extension also changes the interpretation of provider diversification. Reducing the concentration can lower ρ_{FF}, ρ_{FD}, ω_F, or all three. Interoperability and rapid switching reduce effective centrality even if provider market shares do not change because failures become easier to isolate and substitute. Therefore, competition and operational resilience policies affect the same welfare equation through different parameters.

5. POLICY DESIGN

5.1 Price residual risk, not artificial intelligence use

A broad tax on artificial intelligence expenditure would be economically ineffective. This would penalize low-risk forecasting, accessibility tools, research assistance, and internal knowledge applications, alongside high-impact autonomous execution. It could also encourage firms to relabel artificial intelligence expenditures or shift them to less transparent suppliers. The model supports a levy on assurance-adjusted residual risk.

Operationally, the tax base can combine action criticality, autonomy, exposure volume, reversibility, assurance evidence, and cross-firm dependence. A payment agent with staged authorization, transaction limits, independent reconciliation, and rapid rollback would face a lower effective residual risk score than a fully autonomous agent with the same transaction volume but weak controls. This approach converts governance quality into an economic incentive rather than a fixed-compliance burden.

A levy need not always take the form of a fiscal one. It can be implemented through risk-based capital requirements, insurance premiums, platform access charges, supervisory fees, procurement scoring, or mandatory contributions to an incident-compensation fund. The essential condition is that the marginal private cost increases with the expected external damage remaining after assurance.

5.2 Assurance credits and shared infrastructure

The measurement of residual risk is difficult, particularly for small and medium-sized enterprises. Therefore, a policy based only on penalties could favor large firms that can afford extensive documentation. Assurance credits can offset these problems. Verified investments in independent evaluation, interoperability, monitoring, rollback, and incident response can reduce the assessed levy. Governments and

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

industry associations can also provide shared testing environments, sectoral red team libraries, standardized incident taxonomies, and subsidized assurance services.

The model predicts that assurance subsidies are most justified when assurance creates spillover benefits that the firm does not capture. Examples include the public disclosure of vulnerabilities, interoperable audit logs, contributions to shared threat intelligence, and participation in coordinated incident exercises. A subsidy for private documentation with no external value would be less defensible.

5.3 Reduce common-mode dependence

Because the external-risk coefficient in Equation (4) and its heterogeneous counterpart in Equations (11) and (12) increase with common dependence, competition, and resilience policies are part of artificial intelligence governance. Provider concentration is not harmful merely because the supplier is large. It becomes systemically important when many critical functions depend on common models, interfaces, identity services, or cloud regions without credible substitutions or fallbacks. Foundation model markets exhibit strong scale economies and ecosystem effects, whereas downstream users of closed APIs can face switching and dependency constraints (Vipra & Korinek, 2023; Schrepel & Pentland, 2025; OECD, 2025).

Policies can lower effective correlation through interoperability standards, portable evaluation records, multi-provider contingency plans, model and cloud exit testing, open incident reporting, modular architectures, and procurement rules that prevent hidden single points of failure. Supervisors should distinguish nominal multi-vendor strategies from genuine diversity. Two applications may have different brand names but share the same foundation model, data provider, or inference infrastructure.

The results also caution against uniform regulatory templates that encourage every firm to implement the same control in the same way. Standardization can improve the minimum quality, but excessive homogeneity can create common blind spots. Outcome-based requirements, diverse testing methods, independent validation, and adversarial exercises can help preserve heterogeneity while maintaining accountability.

5.4 Sector-specific systemic-risk supervision

The externality coefficient χ and correlation parameter ρ differ by sector. An internally used writing assistant has limited propagation. An agent that manages collateral, prices securities, approves benefits, schedules hospital resources, or controls industrial infrastructure can create a much larger downstream loss. Therefore, the efficient regulatory intensity should vary with action criticality and network position.

Financial authorities already possess tools for stress testing, third-party concentration assessment, operational resilience, and systemicmrisk monitoring. These can be extended to agentic artificial intelligence by incorporating common model exposure, correlated scenario failures, action speed, human override latency, and recoverability. Public-sector procurement can apply similar principles through risk-tiered assurance requirements and supplier concentration reporting.

A useful supervisory metric is the marginal system contribution rather than the stand-alone model accuracy. A system with a high average performance can still be systemically dangerous if it fails in the same states as many other systems. Conversely, a specialized model with modest average accuracy may be valuable as a diverse fallback if its errors are weakly correlated with the dominant provider.

Table 6. Policy instruments implied by the model

Instrument	Economic mechanism	Implementation evidence
Residual-risk levy	Internalize remaining external damage	Assurance-adjusted risk score; sectoral external-risk estimate
Assurance credit	Reward verified spillover-producing controls	Independent evaluation; rollback; shared incident data
Interoperability and exit testing	Lower provider and infrastructure correlation	Portability tests; substitution time; dependency mapping
Incident reporting	Improve estimation of frequency, severity, and dependence	Common taxonomy; near-miss and loss data
Systemic stress testing	Identify correlated failure before deployment at scale	Common scenarios; multi-agent propagation; recovery testing
Risk-based procurement	Use public purchasing power to improve assurance	Action criticality; auditability; supplier concentration

5.5 Third-party audit and sectoral peer review as verification infrastructure

The levy can only work if the measured assurance is credible. Self-reported checklists are insufficient because Equation (15) provides firms with an endogenous incentive to overstate control effectiveness. A third-party auditor can reduce this information asymmetry by independently re-performing evaluations,

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

sampling decision logs, checking configurations and access controls, testing rollback and escalation procedures, validating dependency maps, and reconciling incident records with claims submitted for levy credits. The auditor's economic function is to raise p_i in Equation (16), not merely to produce a compliance certificate.

The audit scope should be proportional to the marginal system contribution. Dominant foundation-model providers, shared cloud or identity suppliers, and critical downstream operators should face more frequent or deeper external verification because the levy stakes and social consequences of understatement are larger. Ordinary downstream users can rely more heavily on standardized evidence packages, pooled sectoral assurance services and risk-based sampling. This proportionality limits the fixed-cost burden on smaller firms while preserving strict scrutiny, where correlated external damage is greatest.

Third-party audits should be complemented by industry-sector peer-review mechanisms. Sector peers often possess technical and operational knowledge that a general auditor may lack, such as failure modes in payment routing, clinical scheduling, industrial control, logistics and insurance underwriting. Structured peer challenges can compare stress scenarios, identify common control blind spots, and test whether nominally independent firms actually rely on the same model or infrastructure. To avoid collusion or disclosure of competitively sensitive information, peer review should be conducted through standardized protocols, anonymized evidence, conflict rules, and regulator-supervised information boundaries.

The audit literature cautions that external oversight can fail through conflicts of interest, weak access rights, inconsistent standards, or audit-washing (Raji et al., 2020, 2022). Therefore, a credible ecosystem requires auditor accreditation, independence requirements, rotation or tender rules where appropriate, regulator access to working papers, random re-performance, sanctions for negligent certification, transparent methodologies, and meta-audits of the auditors themselves. The regulator should also maintain the authority to override assurance credits when the incident evidence conflicts with the certified score.

Current policy directions are compatible with this approach. NIST's AI Risk Management Framework emphasizes ongoing measurement, governance, and monitoring rather than one-time certification, whereas the European Commission's guidance for general-purpose AI providers stresses documentation and information flows to downstream actors (National Institute of Standards and Technology, 2023; European Commission, 2026). These mechanisms can provides evidence for the levy, but the model adds a distinct economic discipline: verification intensity should be calibrated to the value of misreporting, which increases with marginal external damage.

Table 7. Verification architecture for an assurance-adjusted residual-risk levy

Layer	Primary mechanism	Economic role	Key safeguard
Internal assurance	Control evidence, logs, tests, incident records	Generate auditable evidence and reduce true residual risk	Separation of control owner and assessor
Accredited third-party audit	Independent re-performance and evidence validation	Increase detection probability p_i^A and deter overstatement	Independence, competence, access rights, rotation
Sectoral peer review	Structured challenge by domain specialists	Increase p_i^P; identify sector-specific and common-mode blind spots	Anonymization, antitrust boundaries, conflict rules
Regulatory meta-audit	Random inspection, auditor review, recalibration	Increase p_i^R; discipline auditors and update levy parameters	Power to inspect, sanction, and override credits
Shared incident infrastructure	Common taxonomy and confidential loss / near-miss data	Improve χ, Ω, and M_i estimation over time	Protected reporting and data-quality controls

This architecture makes the Levy self-correcting over time. Better audits improve the measurement of assurance, whereas incident and peer-review data improve the estimation of external damage and correlation. In turn, better-calibrated prices strengthen the incentive to invest in controls that auditors can verify. The result is a feedback system that links economic incentives, technical evidence, and supervisory learning.

6. EMPIRICAL IMPLICATIONS AND RESEARCH AGENDA

The model generates testable predictions. First, sectors with greater provider concentration and common infrastructure should exhibit a stronger correlation with artificial intelligence incidents, even after controlling for the average system quality. Second, firms with stronger liabilities, reputational exposure, or insurance pricing should invest more in assurance than autonomy. Third, a provider outage, model update, or shared vulnerability should have larger cross-firm effects, where exit costs are high. Fourth, holding intrinsic firm parameters constant, levy-equivalent coefficients should rise with measured network centrality and cross-layer dependence; dominant providers should therefore face larger marginal charges only when they are genuine common dependencies. Fifth, policies that reward verified assurance should generate more efficient adjustments than undifferentiated restrictions on artificial intelligence adoption. Sixth, the frequency and magnitude of discrepancies between self-reported and externally verified assurance should fall as effective audit probability and expected sanctions increase. Seventh, sectoral peer reviews should be most valuable where failure modes are domain-specific or where common suppliers create correlated blind spots.

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

Empirical studies require new data. Current incident databases are incomplete, inconsistently classified, and often biased toward failure visibility. Regulators and industry bodies should collect action-level data on system autonomy, task criticality, provider dependencies, assurance controls, audit results, near misses, realized losses, recovery times, and downstream propagation. Dependency mapping should identify foundation models, cloud regions, tool providers, identity layers, and common middleware, rather than relying only on vendor names. Confidential supervisory data can support the estimation of dependence matrices and marginal system contributions without disclosing firm-sensitive details.

A feasible research design would combine firm surveys with technical dependency mapping, audit records and incident data. The residual risk can be estimated using stress test failure rates weighted by action severity and reversibility. Correlation can be measured across standardized scenarios, historical incidents, or provider shocks. Difference-in-differences designs can exploit the introduction of sectoral reporting rules, mandatory external audits, insurance requirements, or procurement standards. Audit data also permit direct tests of Equation (16): researchers can examine whether misstatement rates decline with inspection probability, sanction severity, and auditor's independence. Instrumental variable strategies might use exogenous cloud outages, provider policy changes, auditor rotation, or jurisdiction-specific compliance deadlines, although exclusion restrictions would require careful defense.

Future theoretical work can move beyond the present two-type specialization by adding more provider classes, endogenous provider choice, vertical integration, strategic pricing, and directed production networks. Heterogeneous firms may differ in terms of productivity, liability, assurance efficiency, market power, and network centrality. Multi-provider markets can be modeled as a portfolio problem in which firms choose both agent autonomy and exposure to common technological factors. The audit extension can also be enriched by endogenous auditor effort, reputation, capture, collusion, and strategic peer-review participation. Dynamic models can then be used to examine how assurance depreciation, learning, incident disclosure, and changing provider concentration affect adoption paths and audit intensity.

Distributional analysis is also required. External losses may disproportionately affect consumers, workers, small suppliers, or communities with limited ability to contest automated decisions. A residual-risk levy can be used to fund compensation, public-interest auditing, or access to appeal mechanisms. However, poorly designed compliance costs could reinforce the incumbent advantage. Therefore, policies should combine proportionality with shared infrastructure and support for smaller organizations.

International coordination is essential because artificial intelligence supply chains cross borders. A failure originating in one jurisdiction can affect firms and users elsewhere, whereas divergent reporting rules can fragment evidence. Common minimum taxonomies, reciprocal incident-sharing protocols, and interoperable assurance credentials would reduce information asymmetry without requiring identical national regulations.

7. DISCUSSION

This model reframes the economics of agentic artificial intelligence. The relevant policy question is not whether autonomy is good. It is whether a firm's chosen combination of autonomy and assurance reflects the full marginal social cost of residual risk. When failures are largely idiosyncratic and recoverable, the wedge can be small. When many firms rely on common models and infrastructures, the wedge can become substantial, even if each firm's internal risk management appears rational.

This framing reconciles innovation with governance. The social optimum retains significant autonomy because delegations create real economic value. Assurance is not an exogenous constraint on productivity; rather, it is a complementary capital stock that allows autonomy to be used safely. An efficient policy mix can therefore increase welfare through three channels: better targeting of autonomy, greater assurance, and lower common dependence.

The analysis also distinguishes the model quality from the system risk. Improving the average accuracy reduces some losses, but does not necessarily reduce the correlation. A dominant model can be individually superior and systemically fragile if its rare failures are shared across the economy. Diversity has an option value because it provides alternative failure modes and recovery paths. The heterogeneous extension sharpens this point: an upstream provider's efficient levy depends on the marginal loss propagated through downstream dependencies, while a downstream user's levy depends on its own network position and exposure. This parallels portfolio and systemic-risk economics, but the relevant assets are models, providers, controls, and organizational routines.

The residual risk levy provides a conceptual benchmark rather than an immediately deployable formula. Estimating firm-specific coefficients requires evidence of expected external loss, dependence, network centrality, and true control effectiveness. Therefore, measurement error, strategic reporting, and model

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

uncertainty are first-order economic problems. The audit extension shows that verification is not a peripheral compliance matter: pricing residual risk creates a financial incentive to report a lower residual risk; thus, an unverified levy can be gamed. In the early stages, regulators may use risk bands, safe harbors, conservative assurance haircuts, and standardized audit packages. Over time, incident data, independent testing, and peer challenges can improve calibration.

Therefore, this study supports a three-level governance architecture. Firm-level controls reduce private and localized risks. Ecosystem-level policies address correlated exposure, provider concentration, interoperability, and cross-organizational propagation. Verification institutions, including third-party auditors, sectoral peer-review mechanisms, and regulatory meta-audits, make the measured assurance inputs credible. None were sufficient alone. A firm can be well governed but embedded in a fragile ecosystem; a diversified ecosystem can contain firms with reckless autonomy; and a sophisticated levy can fail if the underlying assurance data are strategically misreported.

8. LIMITATIONS

The model makes the following simplifying assumptions. Quadratic functional forms and linear assurance technology were selected for tractability. Real assurance may exhibit thresholds, complementarities, and diminishing or increasing returns. Residual risk may depend on the task mix, human behavior, adversarial response, and organizational culture. The equicorrelation benchmark compresses a complex network into a single parameter, while the two-tier extension still reduces a heterogeneous AI ecosystem to two representative firm classes and three correlation parameters. Real networks are directed, multilayered, time-varying, and may contain several systemically central downstream nodes.

The analysis is static, except for the brief assurance-depreciation extension. It does not model learning, endogenous innovation, provider pricing, entry, vertical integration, or international-location decisions. A levy could affect each margin. The verification model is also deliberately parsimonious: audit probability and sanctions are treated as policy parameters, while real auditors choose effort, face conflicts of interest, learn over time, and may become concentrated. Sectoral peer review can add expertise but can also reproduce common blind spots or create antitrust concerns. These second-order institutional effects should be explicitly modeled in future studies.

The numerical calibrations are presented for illustration. Neither the symmetric parameters nor the two-tier provider/downstream parameters are estimated, and the reported welfare gains, levy coefficients, and incidence differences are not forecasts. Their value lies in demonstrating comparative statics and identifying observable quantities that future empirical studies should estimate. In particular, the finding that the provider coefficient is approximately four times the downstream coefficient is a property of the stated stylized calibration, not a claim about any named foundation model company or market.

Finally, welfare is represented in the form of aggregate normalized units. Distributional weights, rights-based constraints, procedural fairness, and irreducible harm are not fully captured by expected-value analysis. Some applications may require hard prohibitions or human decision rights, even when a monetary model permits delegation.

9. CONCLUSION

Agentic artificial intelligence creates a new form of digital externality. Firms receive productivity gains from autonomous execution but may not bear the full cost of failures transmitted through shared models, providers, infrastructures, and markets. Therefore, a welfare-economic analysis must consider not only how much artificial intelligence is adopted, but also how autonomy is combined with assurance, how residual risk is correlated across organizations, which firms are systemically central, and whether reported assurance can be independently verified.

The model developed in this study shows that positive external damage produces excessive private autonomy, insufficient assurance intensity, and excessive residual risk. The cross-firm correlation amplifies each wedge. In the heterogeneous extension, efficient levy incidence varies with marginal system contribution; therefore, a dominant foundation model provider can face a materially larger coefficient than a downstream API user when the provider is a common dependency. The audit extension adds a second result: residual-risk pricing is credible only when expected detection and sanctions make assurance overstatements unprofitable. Third-party audits, sectoral peer reviews, and regulatory meta-audits are therefore complements to the levy rather than optional administrative additions.

The practical implication is to replace homogeneous AI regulations with verifiable marginal-contribution governance. Governments, regulators, insurers, boards, and procurement authorities should measure consequential autonomy, true assurance effectiveness, reversibility, provider dependence, network centrality,

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

and marginal contributions to system-wide losses. They should price residual risk proportionately, reduce common-mode dependence through interoperability and exit capacity, and build an independent verification ecosystem that makes assurance claims reliable. Such a framework preserves the productivity value of agentic systems while aligning private deployment decisions with economic resilience and the public welfare.

ACKNOWLEDGMENTS
The author acknowledges the broader scholarly communities working on artificial intelligence economics, systemic risk, digital governance and organizational resilience. Any errors or omissions are the sole responsibility of the author.

FUNDING
This research received no external funding.

CONFLICT OF INTEREST
The author declares no conflict of interest.

ETHICS STATEMENT
This theoretical study did not involve human participants, animals, personal data or confidential organizational records. Therefore, institutional ethical approval was not required.

DATA AND CODE AVAILABILITY
No empirical datasets were used. All numerical results, including the symmetric calibration, Monte Carlo analysis, two-tier provider/downstream simulation, and provider centrality sensitivity analysis, were generated from the equations, parameter values, and parameter ranges reported in the manuscript. The analyses can be reproduced directly from the specifications. The reproduction code is available from the author upon reasonable request.

AUTHOR CONTRIBUTION
Kwan Hong TAN conceived the study, developed the model, conducted the analytical calibration and robustness analysis, interpreted the findings, and wrote the manuscript.

REFERENCES
Acemoglu, D. (2024). The simple macroeconomics of AI. NBER Working Paper No. 32487. https://doi.org/10.3386/w32487
Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. NBER Working Paper No. 24196. https://doi.org/10.3386/w24196
Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. Journal of Political Economy, 128(6), 2188-2244. https://doi.org/10.1086/705716
Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. Review of Financial Studies, 30(1), 2-47. https://doi.org/10.1093/rfs/hhw088
Adrian, T., & Brunnermeier, M. K. (2016). CoVaR. American Economic Review, 106(7), 1705-1741. https://doi.org/10.1257/aer.20120555
Aldasoro, I., Gambacorta, L., Korinek, A., Shreeti, V., & Stein, M. (2024). Intelligent financial system: How AI is transforming finance. BIS Working Papers No. 1194. Bank for International Settlements.
Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3-30. https://doi.org/10.1257/jep.29.3.3
Bank of England. (2025). Financial stability in focus: Artificial intelligence in the financial system. Bank of England.
Baumol, W. J., & Oates, W. E. (1988). The theory of environmental policy (2nd ed.). Cambridge University Press.
Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy, 100(5), 992-1026. https://doi.org/10.1086/261849
Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. American Economic Journal: Macroeconomics, 13(1), 333-372. https://doi.org/10.1257/mac.20180386

e-ISSN: 2961-712X
Vol. 5 Issue 2, July-December 2026
DOI:10.55299/ijec.v5i2.2022

Calvano, E., Calzolari, G., Denicolo, V., & Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. American Economic Review, 110(10), 3267-3297. https://doi.org/10.1257/aer.20190623

Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A. J., Pizzinelli, C., Rockall, E. J., & Tavares, M. M. (2024). Gen-AI: Artificial intelligence and the future of work. IMF Staff Discussion Note 2024/001. https://doi.org/10.5089/9798400262548.006

European Commission. (2026). Guidelines for providers of general-purpose AI models. Directorate-General for Communications Networks, Content and Technology. https://digital-strategy.ec.europa.eu/en/policies/guidelines-gpai-providers

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union.

Financial Stability Board. (2024). The financial stability implications of artificial intelligence. Financial Stability Board.

Financial Stability Board. (2025). Monitoring adoption of artificial intelligence and related vulnerabilities in the financial sector. Financial Stability Board.

Graetz, G., & Michaels, G. (2018). Robots at work. Review of Economics and Statistics, 100(5), 753-768. https://doi.org/10.1162/rest_a_00754

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. Journal of Finance, 72(3), 967-998. https://doi.org/10.1111/jofi.12498

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0). NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

OECD. (2025). Artificial intelligence and competitive dynamics in downstream markets. OECD Publishing. https://www.oecd.org/en/publications/artificial-intelligence-and-competitive-dynamics-in-downstream-markets_ccf0624a-en.html

Pigou, A. C. (1920). The economics of welfare. Macmillan.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33-44. https://doi.org/10.1145/3351095.3372873

Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. E. (2022). Outsider oversight: Designing a third party audit ecosystem for AI governance. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 557-571. https://doi.org/10.1145/3514094.3534181

Sandmo, A. (1975). Optimal taxation in the presence of externalities. Swedish Journal of Economics, 77(1), 86-98. https://doi.org/10.2307/3439329

Schrepel, T., & Pentland, A. S. (2025). Competition between AI foundation models: Dynamics and policy recommendations. Industrial and Corporate Change, 34(5), 1085-1103. https://doi.org/10.1093/icc/dtae042

Tan, K. H. (2025a). The digital productivity paradox: Intangible capital, institutional capacity, and the delayed gains from artificial intelligence. ResearchGate working paper.

Tan, K. H. (2025b). Artificial intelligence as stakeholder: Rethinking corporate responsibility in algorithmically mediated organizations. Zenodo. https://doi.org/10.5281/zenodo.17756903

Tan, K. H. (2026a). The AI productivity-governance frontier: A theoretical model for enterprise value creation under agentic automation. International Journal of Science and Research Archive, 20(1), 52-60. https://doi.org/10.30574/ijsra.2026.20.1.1434

Tan, K. H. (2026b). Navigating the supervisory economy: AI stakeholder recognition and the transformation of organizational accountability. International Journal of Research Publication and Reviews, 7(6), 7670-7675. https://doi.org/10.55248/gengpi.07.0626.18a10

Tan, K. H. (2026c). Governance-aware agentic AI for enterprise engineering systems: A design-science reference architecture and quantitative risk-control model. ResearchGate preprint. https://doi.org/10.13140/RG.2.2.18948.69768

Tan, K. H. (2026d). The governance half-life of agentic artificial intelligence controls: A dynamic assurance framework for control decay, residual risk, and revalidation. ResearchGate working paper.

Vipra, J., & Korinek, A. (2023). Market concentration implications of foundation models. arXiv:2311.01550. https://doi.org/10.48550/arXiv.2311.01550

Wagner, W. (2010). Diversification at financial institutions and systemic crises. Journal of Financial Intermediation, 19(3), 373-386. https://doi.org/10.1016/j.jfi.2009.07.002