The Economics of Machine Verification: Verification-Cost Shocks and the Extensive Margin of Monitoring*

Amrit Gill

Independent Researcher

amritbir1@gmail.com

August 2026

Abstract

Economics understands how costly verification disciplines behavior. This paper asks which claims become worth verifying at all when machine intelligence changes the technology of verification. I model verification as a technology vector: fixed and marginal costs, four pipeline reliabilities from detection to enforcement, and a false-positive rate. Three results follow. First, a verification frontier: the minimum economically contestable claim $L^*(\tau, N)$ in claim value and claim volume, with component-wise comparative statics and a false-challenge-dominance condition under which monitoring entry is monotone. Second, recovery is the wrong success metric: under bounded misstatement gains and attainable full deterrence, measured recoveries are hump-shaped in verification effectiveness, and in general a technology that works can reduce measured recoveries. Third, private and social verification frontiers diverge: transfers and externalized dispute costs push private verification toward excess; uninternalized deterrence spillovers push it toward undersupply. The broader implication is institutional. Much of commercial architecture is adaptation to costly checking, and where the pipeline clears, a verification-cost collapse re-selects it. The open question is how much economic activity lies below the human verification frontier but above the machine one.

Keywords: artificial intelligence; verification cost; monitoring; extensive margin; information rents; auditing; deterrence; enforcement.

JEL codes: D82, D83, D86, K42, L14, M42, O33.

*Working paper, August 2026. Comments welcome. Large language models were used extensively in drafting, revising, and stress-testing this paper; all claims, judgments, and errors are the author's own. Disclosure: the author holds commercial interests in automated verification of commercial energy billing.

1 Introduction

Two cost collapses define the economics of artificial intelligence. The first, in the cost of cognitive execution, is well underway: machines draft, code, reconcile, and calculate. The second is a collapse in the cost of establishing whether economically consequential claims are correct, substantiated, and actionable. This paper is about the second collapse—and about the fact that its economics are governed not by whether claims *can* be checked, but by which claims are *worth* checking.

A vast class of economic claims is already verifiable in principle. An invoice line can be reconciled against a contractual formula; a supplier charge can be tested against an indexed pricing term; a reported fact can be compared with its source records. What protects such claims from inspection is rarely logic. It is economics: the all-in cost of detection, substantive verification, evidentiary preparation, and enforcement exceeds the value at stake. A €43 discrepancy that costs €200 to establish is safe—not hidden, just not worth finding. Human verification technologies, with their high fixed and marginal costs, have therefore imposed minimum economic scales of contestation, and the economy's monitoring institutions—sampling, materiality thresholds, episodic audit, tolerated small discrepancies—are the architecture that grew around those scales.

The contribution of this paper is to make the verification frontier itself the object of technological change. Existing auditing and enforcement theory asks how verification disciplines behavior given costly monitoring; this paper asks which classes of economic claims enter the economically contestable set when the verification technology shifts, and how that frontier movement alters behavior, rents, and welfare. The paper does not introduce a new logic of deterrence—that logic runs from Becker (1968) through optimal auditing (Townsend 1979; Border and Sobel 1987; Mookherjee and Png 1989; Khalil 1997) and is confirmed at state scale in the tax-enforcement literature, where credible third-party information and verifiability can generate preventive deterrence before realized audits occur (Kleven et al. 2011; Pomeranz 2015). The divergence between private and social incentives to invoke the legal system is itself well established (Landes and Posner 1975; Shavell 1982, 1997), and Section 2 cites precedents for costly verification, audit deterrence, scale economies in verification, strategic complexity, and verification technology individually. To my knowledge, however, prior work has not jointly modeled a machine-driven, multidimensional verification-cost shock across heterogeneous commercial claim classes, derived the resulting extensive-margin verification frontier, and traced its consequences through deterrence, opacity rents, private and social monitoring incentives, and institutional re-selection. That combined architecture—and what it implies when the frontier moves at scale—is the contribution.

Three results organize the paper.

Result 1 — the verification frontier. Modeling verification as a technology vector $\tau = (c, K, q_D, q_V, q_E, q_L, f)$—marginal cost, fixed architecture cost, four pipeline reliabilities from detection through enforcement, and a false-positive rate—yields a minimum economically contestable claim $L^*(\tau, N)$ in claim value *and claim volume*. Volume matters because fixed costs amortize: a €2 discrepancy repeated two million times can justify verification infrastructure no human would rationally deploy on any individual €2 case, which is why the transition arrives volume-first, in billing, procurement, and standardized transactional layers. The component-wise comparative statics answer a question a scalar "technology improves" index cannot: falling marginal cost moves the frontier for high-volume classes, falling fixed cost unlocks the lower-volume tail, and no improvement in detection can move the frontier at all when the evidentiary and enforcement components are near zero—the formal difference between a dashboard and a verification technology. The frontier's movement is disciplined by a condition that turns out to be economically substantive: because better deterrence enlarges the honest-in-equilibrium population exposed to false positives, monotone entry requires the false-positive rate to fall proportionally faster than the honest population grows. Verification success expands the set of people a system can wrongly accuse.

Result 2 — recovery is the wrong metric. Crossing the frontier changes behavior before it changes recoveries, and eventually instead of them. In the model's no-commitment inspection game, cheaper verification lowers equilibrium misstatement without raising equilibrium inspection frequency: the threat does the work. Consequently, under bounded gains and attainable full deterrence, measured recoveries are hump-shaped in verification effectiveness—rising while the technology harvests the pre-existing stock of misstatements, falling as deterrence shrinks the flow—and in general their sign response to improvement is unrestricted. A successful verification technology should eventually look unproductive on a recovery dashboard. The field evidence is consistent with this shape, most recently in the first credibly identified field study of automated private enforcement, where compliance moved by double digits while the enforcement instrument remained, for most of the affected population, a threat (Merane and Stremitzer 2026).

Result 3 — the private and social frontiers diverge. Deriving welfare from primitives rather than assuming it, the private and social verification thresholds differ, and the sign of the wedge is economically informative. Much of a prevented loss is a transfer, remediation is almost purely redistributive, and part of dispute costs falls on the challenged party—all of which bias private verification entry toward excess: a region of privately profitable, socially wasteful verification races, the machine-age version of Hirshleifer's (1971) overinvestment

in distributive information. Pulling the other way, deterrence spills over to relationships the verifying principal does not own, biasing private enforcement toward undersupply. Both corners are already observable: dispersed-victim automated enforcement produced a market-wide compliance response that no individual victim had an incentive to finance—and the same technology, profit-financed, is recognizably capable of nuisance enforcement against the honest.

These results serve an implication larger than any of them. The paper states it with its conditions attached, because the conditions are what make it credible. Much of the architecture of commerce is not misstatement but adaptation to expensive checking: fixed-price contracts written where cost verification was uneconomic; brands posted as bonds for quality no buyer could verify (Klein and Leffler 1981); grading, standardization, and commoditization as economies of measurement (Barzel 1982); sampling, materiality thresholds, and episodic audit; escrow, letters of credit, and payment-term conventions; intermediaries whose margin is in part a verification wage. Each embeds a shadow price of verification calibrated to the human technology. Where the full pipeline clears, that shadow price collapses—and the prediction is not an audit boom but the re-selection of these institutions, operating through anticipation, unraveling, and contract redesign before realized inspection. Section 10 develops this implication, including why the same claim, made unconditionally for an earlier technology, failed for reasons this paper's own results formalize.

The timing of these questions is not speculative. Merane and Stremitzer (2026) document automated detection and claim generation driving non-compliance with an adjudicated legal rule down by 22.7 percentage points within three months, across claims individually too small ever to have justified human enforcement. Jin, Sokol, and Wagman (2026) show formally that AI-augmented monitoring below an institutional-capacity threshold generates enforcement workload without reducing harm. Catalini, Hui, and Wu (2026) model the scarcity of human verification as the binding constraint of the automation era. Between the bottleneck those authors study and the enforcement episodes now appearing lies the object this paper formalizes: the frontier that determines which claims cross from economically protected to economically contestable when the technology of verification changes. Production-cost collapse gave economics the post-scarcity question; verification-cost collapse poses the corresponding question for trust, developed in Section 10 and reduced to a single measurable quantity: how much economic activity lies below the human verification frontier but above the machine one.

Section 2 positions the paper. Section 3 defines the verification technology. Section 4 presents the model. Section 5 derives the frontier and the rents it protects. Section 6 derives the equilibrium effects of crossing it. Section 7 compares private and social frontiers. Section 8 turns to evidence and measurement. Section 9 collects extensions and boundary conditions;

Section 10 defines the post-opacity transition; Sections 11 and 12 conclude.

2 Related Literature

Costly state verification and optimal auditing. Townsend (1979) made the cost of revealing a privately observed state a determinant of contract form; Gale and Hellwig (1985) derived the debt implications. The auditing tradition made monitoring the choice variable: Border and Sobel (1987) characterize auditing as the instrument limiting an informed agent's rents; Mookherjee and Png (1989) establish random audits; Reinganum and Wilde (1985) and Graetz, Reinganum, and Wilde (1986) develop the compliance game; Khalil (1997) treats auditing without commitment; Varas, Marinovic, and Skrzypacz (2020) and Silva (2020) develop dynamic monitoring and evidence design. In mechanism design, Ben-Porath, Dekel, and Lipman (2014) characterize optimal allocation with costly verification, Halac and Yared (2020) derive threshold rules when verification is costly, and Ball and Kattwinkel (2025) model verification as a probabilistic testing technology whose precision continuously interpolates between private and complete information—the closest theoretical treatment of verification technology as a varying object. This paper takes that technology as the object undergoing change across a population of claim classes and studies the resulting extensive margin.

Enforcement in the field. The deterrence mechanism (Allingham and Sandmo 1972) is empirically established at state scale. Kleven et al. (2011) find evasion near zero for third-party-reported income and substantial for self-reported income; Pomeranz (2015) provides experimental evidence that the VAT paper trail deters misreporting through the prospect of cross-checking, before audits are realized; Carrillo, Pomeranz, and Singhal (2017) document that when one margin becomes verifiable, firms substitute almost fully toward harder-to-verify margins; Naritomi (2019) shows consumers can be enlisted as auditors; Slemrod (2007) surveys the field, including the 1987 episode in which roughly seven million claimed dependents vanished when a verification requirement arrived. Four further findings matter here. A technology shock to verification has a direct precedent: digital encryption of VAT invoices in China raised enforcement by making falsification harder, with large short-run revenue gains that attenuate over time (Fan et al. 2020). Monitoring architecture changes procedural choices: audit protocols that scrutinized auctions more intensely than direct contracting shifted Chilean procurement officials away from auctions (Gerardino, Litschig, and Pomeranz 2024). Enforcement can lose the race to sophisticated avoidance: transfer-pricing regulation in Chile left tax payments essentially unchanged while drastically expanding tax-advisory services (Bustos et al. forthcoming). And the deterrence value of audits substantially exceeds their

direct revenue in US administrative data (Boning et al. 2025). These findings discipline the theory: the behavioral responses modeled below are documented phenomena. The question this paper takes up is what changes when the mechanism escapes the sovereign setting—compelled reporting, standardized filings, administrative sanction—and enters fragmented private markets where every stage of the pipeline must be privately produced and privately profitable, and where verifiability itself is a contracted rather than compelled variable.

Private enforcement and the incentive to invoke the legal system. That private and social incentives to use the legal system diverge is settled. Landes and Posner (1975) show competitive private enforcement can be socially excessive, because the enforcer captures the sanction while society receives only the deterrence benefit net of enforcement cost; Polinsky (1980) compares private and public enforcement of fines; Shavell (1982, 1997) establishes the divergence directly—a plaintiff weighing private gain against private cost while ignoring both the defendant's litigation costs and the deterrence conferred on parties outside the suit. Proposition 5 inherits that structure. What is new here is that the wedge is a function of the verification technology—its components are the false-positive rate, the dispute-cost split, and the pipeline reliabilities—so a verification-cost shock moves the divergence itself rather than operating within a fixed one.

Disclosure and certification. With certifiable information and costless disclosure, unraveling obtains (Grossman 1981; Milgrom 1981); disclosure costs and uncertainty sustain partial opacity (Jovanovic 1982; Verrecchia 1983; Dye 1985); certifiers can strategically coarsen information and capture surplus (Lizzeri 1999; Dranove and Jin 2010). Section 9.4 uses this literature for the second-stage implication of cheap verification: voluntary verifiability.

Technology and verification cost. Catalini and Gans (2020) identify verification cost as a principal economic cost affected by blockchain; Gans (2019) shows cheaper verification of contractual performance enlarges the feasible contract space; Goldfarb and Tucker (2019) place verification among digitization's cost reductions; Alles, Kogan, and Vasarhelyi (2002) insist continuous assurance must be economically demanded, not merely feasible. This paper generalizes from natively digital settings to ordinary commercial claims with heterogeneous data, benchmarks, evidence, and remedies. Closest in spirit within economics itself, Colliard, Hurlin, and Pérignon (2024) develop the economics of computational reproducibility: third-party verification of research results carries fixed and variable costs with economies of scale, and its equilibrium level is suboptimally low because of public-good and externality problems—a private–social divergence in verification provision, established in a scientific

setting. Verification costs, their fixed–variable structure, scale economies, externalities, and the private–social wedge therefore all have prior literatures; the paper claims none of them as new.

The 2026 AI-verification literature. Catalini, Hui, and Wu (2026) model exponentially falling execution costs against biologically bounded human verification—the bottleneck regime in which rents migrate toward ground truth, provenance, and liability. Jin, Sokol, and Wagman (2026) analyze a dynamic inspection game in which regulated firms redesign violations faster than under-resourced monitors recognize them, so AI-augmented monitoring below a regime-shifting investment threshold produces workload without reducing harm. Merane and Stremitzer (2026) provide the first credibly identified field study of automated private enforcement. Zhang (2026) models the make-or-buy of checking AI output, where establishing the verifier's diligence inherits the difficulty of the task verified, so the market for verification exists only on a bounded band whose edge is set by adjudication capacity and liability—the meta-verification problem whose downstream components this paper's q_E and q_L summarize in reduced form. On the empirical side, Shi (2024) finds Medicare audits generated savings predominantly through deterrence and induced upfront compliance investment rather than through recoveries—field evidence for the mechanism of Propositions 3 and 4. On measurement, Kiyani et al. (2026) study escalation policies between weak and strong verification, Wang (2026) argues for statistical limits of auditing in rare-error regimes, and Crescitelli et al. (2026) argue evaluation should price verification cost directly. In the audit industry itself, AI adoption is already measurable: it is associated with higher audit quality, lower fees, and a changed task composition of audit labor (Fedyk et al. 2022; Law and Shen 2025), while auditors discount AI-generated evidence relative to identical human-sourced evidence (Commerford et al. 2022)—a microfoundation for treating evidentiary acceptance q_E as a distinct, slow-moving pipeline component. Existing work, in short, studies verification as a bottleneck, a scarce input, or a statistical problem. This paper studies what happens when the verification technology itself improves enough to move previously uneconomic private claim classes across the monitoring frontier—and which components of the improvement do the moving. Its elements are the machine-driven verification-cost shock across heterogeneous commercial claim classes, the extensive-margin verification frontier, the verification pipeline, opacity rents, and the institutional re-selection developed in Section 10 as the post-opacity transition.

3 The Verification Technology

3.1 A vector, not a scalar

Treating "verification improves" as a one-dimensional index makes better technology synonymous with everything getting better, and the comparative statics inherit the assumption. Machine verification is not one technology. Define a verification technology for claim class i as the vector

$$\tau_i = \Big(c_i,\, K_i,\, q_{D,i},\, q_{V,i},\, q_{E,i},\, q_{L,i},\, f_i\Big), \tag{1}$$

where c_i is the marginal all-in cost per claim monitored, K_i the fixed cost of integration, rule development, and architecture, q_D the probability a relevant discrepancy is detected, q_V the probability it can be substantively established conditional on detection, q_E the probability the result can be assembled into evidence the relevant institution accepts, q_L the probability the institution imposes an economically meaningful consequence, and f the probability an honest claim is falsely challenged. Two derived reliabilities govern behavior and remediation: the agent-consequence probability and the principal-remediation probability,

$$\theta^A = \theta^A(q_D, q_V, q_E, q_L), \qquad \rho^P = \rho^P(q_D, q_V, q_E, q_L), \tag{2}$$

each increasing in every component, with the multiplicative pipeline $\theta^A = q_D\, q_V\, q_E\, q_L$ as the leading case. The multiplicative form is deliberately unforgiving: if any stage is near zero, a hidden discrepancy almost never becomes an economic consequence. Institutions can and do separate the two derived objects—a refund can arrive without the counterparty feeling a sanction, and a sanction can land without the principal being made whole—which is why deterrence and remediation enter the model as distinct probabilities.

3.2 Which components machine systems move

Machine systems compress the information-processing components directly: parsing, retrieval, comparison, formula execution, and reproducible calculation drive c down and q_D, q_V up, while reusable software and foundation models convert bespoke integration into configuration, driving K down. They also compress parts of evidence and enforcement—provenance capture, document assembly, claim drafting, routing—the margin exploited in the automated-enforcement episode of Merane and Stremitzer (2026), where detection and claim generation were mechanized against an already-adjudicated rule. What machine systems do not create by themselves is legal standing, evidentiary admissibility, or collection: q_E and q_L are institutional, and their costs are governed by courts, arbitrators, regulators, and bargaining power. A

Table 1: The verification technology vector: channels, frontier effects, and evidence

Component	Machine channel	Frontier effect	Claim classes / evidence
c (marginal cost)	parsing, retrieval, formula execution	lowers L^* through the numerator; binding where K/N is small	high-volume transactional layers: billing, procurement (Sec. 8)
K (fixed cost)	reusable software; foundation models replace bespoke integration	lowers L^* in proportion to $1/N$	mid- and low-volume classes; continuous assurance (Alles et al. 2002)
q_D (detection)	anomaly search, cross-record matching	raises A; subject to gain dominance (Prop. 2)	web-scale detection (Merane and Stremitzer 2026)
q_V (substantiation)	rule execution against benchmarks	raises A; discounted by correlated error (Sec. 9.1)	formula-governed claims (Kiyani et al. 2026; Wang 2026)
q_E (evidence)	provenance capture, reproducible calculation	without it the dashboard limit binds (Cor. 1)	admissibility and evidence standards
q_L (enforcement)	claim drafting, routing, escalation	without it the dashboard limit binds (Cor. 1)	workload without harm reduction (Jin et al. 2026)
f (false positives)	precision at the chosen operating point	raises L^*; governs entry via FCD	nuisance-enforcement pathologies (Sec. 7.2)

machine-verification shock is therefore a movement of some components of τ, not all—and Section 5 shows the economics differ sharply by component.

A claim class is exposed to such a shock when five conditions coincide: the underlying state is represented in accessible digital data; the obligation or benchmark is formal enough to encode; the class-level economic consequence is material relative to checking cost; the evidence can be reproduced and attributed; and some institution—counterparty, auditor, regulator, insurer, arbitrator, court—can act on the result. These conditions predict the mechanism binds first in contract-to-invoice reconciliation, procurement, standardized claims, utility and telecom billing, logistics, and parts of tax, regulatory, and financial reporting, and only weakly where judgment dominates or states are radically incomplete. Table 1 summarizes the vector: which machine channel moves each component, what it does to the frontier of Section 5, and where the evidence sits.

4 The Model

4.1 Claim classes and volume

A continuum of claim classes $i \in I$ with measure μ. A claim class is a stable obligation-rule pair applied to repeated transactions: a type of invoice line, an indexed price component under a contract version, a network charge rule, a claim category. Class i generates N_i claims per period; each misstated and unremedied claim imposes loss $L_i > 0$ on a principal. For each claim, the informed agent draws a private gain from misstatement $b \sim H_i$, continuous with density h_i on $[\underline{b}_i, \bar{b}_i]$, with $\underline{b}_i \leq 0$ permitted: types with $b \leq 0$ are honest regardless of monitoring, so the intrinsically honest share is $H_i(0) \geq 0$, and honest claims are the population on which false positives are generated. Throughout, $b \leq L_i$: the agent cannot gain more from a misstatement than the principal loses; the wedge $L_i - b$ is the real resource cost of the misstatement (misallocation, degraded quality, defensive and dispute waste), a primitive that Section 7 uses and the private calculus below ignores.

The baseline timing is technology-first: before agents act, the principal chooses a monitoring architecture $m_i \in \{0, 1\}$; when active, the class is automatically subjected to the pipeline. The commitment is architectural—reconciliation embedded in workflow—not a promise to remember to inspect; Section 6.1 removes it.

4.2 Behavior and the monitoring decision

Unmonitored, every type with $b > 0$ misstates; the principal's expected loss per claim is $L_i[1 - H_i(0)]$. Monitored, a misstating agent faces expected consequence $\theta_i^A F_i$, where $F_i > 0$ is the enforceable sanction, and misstates if and only if

$$b > \theta_i^A F_i, \tag{3}$$

so the misstating share is $1 - H_i(\theta_i^A F_i)$. Throughout, F_i is the *all-in* expected consequence of a completed pipeline for the agent, inclusive of any restitution or clawback the agent bears; Section 7 prices the restitution component as a transfer. Monitoring benefits the principal through two channels—deterrence and remediation—and costs it false challenges of the honest at rate f_i, each imposing dispute cost φ_i on the principal (a share of a larger social dispute cost; Section 7). The per-claim gross benefit is

$$\beta_i(\tau) = L_i \, A_i(\tau) - f_i \, \varphi_i \, H_i(\theta_i^A F_i), \tag{4}$$

$$\begin{aligned} A_i(\tau) &= [1 - H_i(0)] - (1 - \rho_i^P)\,[1 - H_i(\theta_i^A F_i)] \\ &= [H_i(\theta_i^A F_i) - H_i(0)] + \rho_i^P\,[1 - H_i(\theta_i^A F_i)], \end{aligned} \tag{5}$$

the second line displaying the two channels: deterred mass valued at L_i, plus the remediated share of residual misstatements. The principal monitors when total benefit covers total cost:

$$m_i^*(\tau) = \mathbf{1}\Big\{N_i\,\beta_i(\tau) \geq K_i + N_i\,c_i\Big\}, \tag{6}$$

with net surplus $\Delta_i(\tau) = N_i\beta_i(\tau) - K_i - N_i c_i$. Equation (6) is the extensive margin of monitoring: unlike the bare inequality "benefit exceeds cost," β_i is generated by an explicit strategic response, an explicit pipeline, and an explicit error structure—and everything that follows is comparative statics of (6) in the components of τ.

5 Result 1: The Verification Frontier

5.1 Entry, and when it is monotone

Consider a technological improvement path $t \mapsto \tau(t)$, continuously differentiable, with $dc/dt \leq 0$, $dK/dt \leq 0$, each pipeline component weakly rising (so $d\theta^A/dt \geq 0$ and $d\rho^P/dt \geq 0$), and $df/dt \leq 0$. Along such a path $A_i(\tau)$ is automatically nondecreasing. The false-positive burden is not: define $\Phi_i(t) = f_i(t)\,H_i(\theta_i^A(t)F_i)$, the expected challenged-honest mass per claim. Its derivative contains opposing forces—

$$\frac{d\Phi_i}{dt} = \frac{df_i}{dt}\,H_i(\theta^A F) + f_i\,h_i(\theta^A F)\,F_i\,\frac{d\theta^A}{dt}, \tag{7}$$

the first term the precision gain, the second the growth of the honest-in-equilibrium population that improved deterrence itself creates. Better verification manufactures more people to falsely accuse.

Definition 1 (False-challenge dominance)**.** The path satisfies false-challenge dominance (FCD) at t if $d\Phi_i/dt \leq 0$; equivalently, if the false-positive rate falls proportionally at least as fast as the honest population grows: $-\,d\ln f_i/dt \;\geq\; d\ln H_i(\theta^A F_i)/dt$.

Proposition 1 (Verification entry)**.** *(a)* Existence: *if Δ_i is continuous in t with $\Delta_i(t_L) < 0 < \Delta_i(t_H)$, a crossing $t^* \in (t_L, t_H)$ exists at which monitoring becomes privately rational. (b)* Monotone entry: *if in addition FCD holds along the path, and at least one of $\{c$ falls, K falls, A rises, Φ falls$\}$ is strict on every interval, then Δ_i is strictly increasing, the crossing is unique, and class i is unmonitored below t^* and monitored above. (c)* Failure: *absent FCD, Δ_i*

can be locally decreasing, and monitoring can switch on and then off along an unambiguously improving technology path.

Part (c) is not a pathology to be assumed away; it is economics. A technology that raises consequence probabilities faster than it improves precision converts previously misstating types into honest ones and then taxes the principal for challenging them. The condition is not a technicality: whether "better verification technology" reliably means "more verification" is itself an economic question, decided by the race between precision and deterrence.

5.2 The frontier in value and volume

Rearranging (6), class i is monitored if and only if its per-claim exposure exceeds

$$L^*(\tau, N) = \frac{c + K/N + f\,\varphi\,H(\theta^A F)}{A(\tau)}, \tag{8}$$

the minimum economically contestable claim.

Proposition 2 (The frontier map)**.** *$L^*(\tau, N)$ satisfies: (i) $\partial L^*/\partial N = -K/(N^2 A) < 0$—volume amortizes fixed cost; (ii) $\partial L^*/\partial c = 1/A > 0$ and $\partial L^*/\partial K = 1/(NA) > 0$—marginal cost governs the frontier for high-volume classes, fixed cost for low-volume classes; (iii) $\partial L^*/\partial f = \varphi H/A > 0$—precision is a frontier variable, not a detail; (iv) an increase in any pipeline component q lowers L^* whenever a gain-dominance condition holds—the deterrence-and-remediation gain in A dominates the added false-positive exposure in the numerator; this condition is logically independent of FCD, and strictly lowers it when $f = 0$; absent that condition its effect is ambiguous.*

Corollary 1 (The dashboard limit)**.** *If remediation requires the evidentiary and enforcement stages, then, for $c > 0$, as $q_E\, q_L \to 0$, $\theta^A \to 0$ and $\rho^P \to 0$, hence $A \to 0$ and $L^*(\tau, N) \to \infty$ for any values of c, K, q_D, q_V, f. No finite claim is economically contestable, however cheap and accurate detection becomes.*

The map in Proposition 2 is the payoff of the vector formulation, because it tells us *which* dimension of machine progress unlocks *which* region of the claim space. Cheap inference ($c \downarrow$) moves the frontier where K/N is already negligible: the high-volume transactional layers, which is why the transition arrives volume-first—a €2 discrepancy repeated two million times crosses before an idiosyncratic €5,000 dispute. Reusable architecture ($K \downarrow$) unlocks the mid- and low-volume tail, converting bespoke integration into configuration. Scale economies in verification are themselves established (Alles, Kogan, and Vasarhelyi 2002; Colliard, Hurlin, and Pérignon 2024); what the amortization argument adds is an explicit frontier in claim

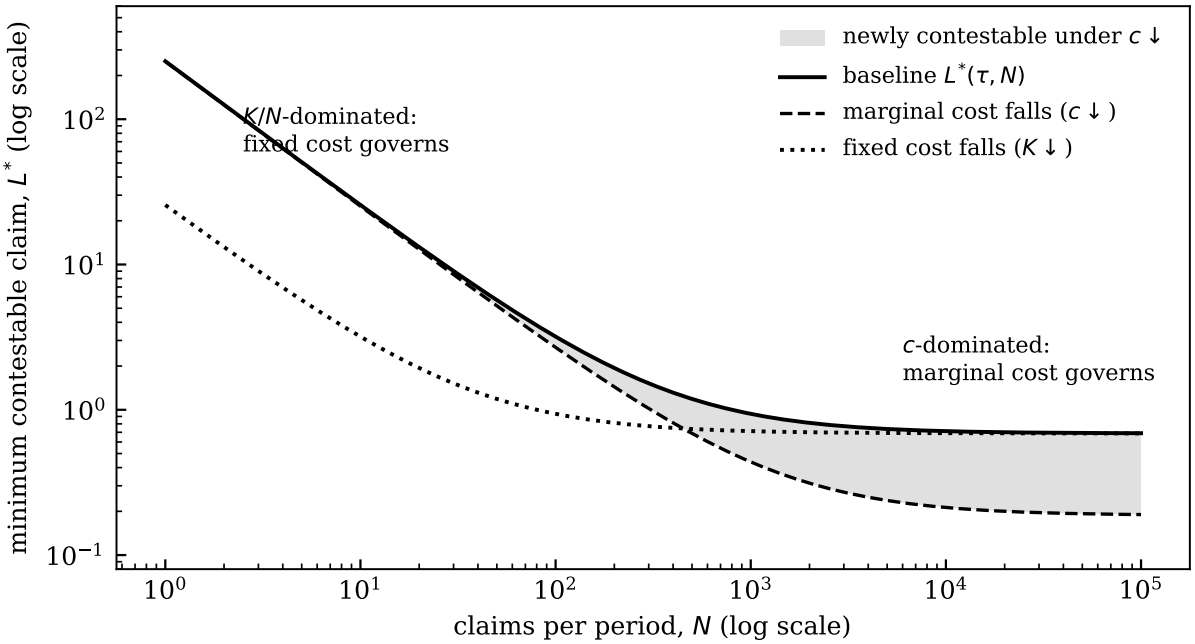

Figure 1: The verification frontier $L^*(\tau, N)$ of equation (8) and its component-wise shifts. A fall in marginal cost c lowers the high-volume asymptote; a fall in fixed cost K rotates the low-volume branch down; a fall in the false-positive rate f operates like c through the numerator. Claim classes between the old and new frontiers become economically contestable. Parameter values are illustrative.

value and claim volume, determining which classes become economically contestable after a technology shock. Precision ($f \downarrow$) is a first-class frontier instrument, and Corollary 1 states formally what practitioners learn expensively: detection perfection cannot substitute for enforceability. A system with 99.9% anomaly detection and no evidentiary or institutional consequence is a dashboard, and its frontier is at infinity—the private-market twin of the regulatory result that AI-augmented monitoring below an institutional-capacity threshold produces workload without harm reduction (Jin, Sokol, and Wagman 2026). Figure 1 draws the frontier and its two characteristic shifts.

Aggregation follows from the frontier's position in the claim population: with G the distribution of per-claim exposures among comparable classes, coverage is $V = 1 - G(L^*)$, and the aggregate effect of a given technological shift depends on the density of claim classes just below the pre-shock frontier. Historically tolerated small discrepancies are not evidence that they do not matter; they are evidence that the cost of challenge exceeded the private gain from correction. The macroeconomic significance of machine verification is an integral over precisely that mass.

5.3 Opacity rent

Define the informed party's surplus as $\Pi_A = \Pi_A^M + \Pi_A^O$: a productive component (compensation for cost, quality, risk-bearing, differentiation, relationship capital) and an opacity component sustained by the counterparty's inability to economically inspect. The opacity rent of class i is defined counterfactually,

$$R_{O,i}(\tau) = \Pi_{A,i}(\tau) - \Pi_{A,i}(\tau^{FI}), \tag{9}$$

where τ^{FI} is a benchmark technology at which the same claims are verifiable at negligible all-in cost, holding the legal and institutional environment fixed. The model of Section 4 normalizes $\Pi_A^M = 0$—the agent's only modeled surplus is opportunistic—so (9) identifies the opacity component exactly; in richer settings the decomposition must be carried explicitly, and (9) is designed so that market power, differentiation, and risk-bearing difference out. Opacity rent needs no fraud: it can live in an overcharge, a wider bargaining margin, a lower claim payout, or a lower probability of contractual challenge. And because $\theta^A, \rho^P \leq 1$ even at τ^{FI}, opacity rent is strictly narrower than the total gains from misstatement: what a verification-cost shock contests is the component protected by inspection cost, not every informational advantage.

Two scope statements discipline the object. First, (9) is a structural definition—a causal estimand, not an observable. Once prices, participation, and contract design respond to verification, $\Pi_{A,i}(\tau^{FI})$ is not directly observed; empirical identification requires exogenous variation in verification cost and explicit assumptions about what is held fixed, and Section 8 states the corresponding designs. Second, an ex-post gap between what a counterparty paid and the cheapest feasible alternative is not itself opacity rent. It decomposes into a *computation gap* (the arrangement was dominated at signing given information available then—the component sustained by verification cost, and the object above); a *forecasting gap* (the arrangement was ex-ante optimal and the state realized differently—risk rather than opacity, though cheap re-verification shortens exposure to it); and *compensated unmeasured attributes* (service, flexibility, risk-bearing priced on dimensions the comparison omits—not a rent at all, and by Section 6.3 the predicted destination of relocated opacity). Only the first component is the model's object, and measurement that does not separate the three overstates it. The distinction gives the technology two objects rather than one. *Conformity claims* verify a bill against its own contract: the counterfactual is the contract, and the mechanisms are the deterrence and remediation channels of this paper. *Optimality claims* verify the contract against the feasible set—every available arrangement repriced against the counterparty's realized data—where nothing is misstated, the rent is an inertia rent sustained

by the historical cost of computing the counterfactual, the operative mechanism is exit and repricing rather than sanction, and switching frictions play the enforcement-floor role that q_L plays for conformity claims. The frontier machinery applies to both: an optimality claim class has its own L, N, c, and K, crosses the same threshold (8), and disciplines pricing through the threat of exit exactly as Proposition 3's threat channel disciplines misstatement—before anyone switches. The institutional consequences of the second object, where intermediaries' margins were the verification wage for carrying K, are developed in the companion paper, which grades the counterfactual itself—best obtainable contract, best contract assemblable from already-priced components, and economically feasible benchmark—in ascending order of ambition and descending order of enforceability.

Corollary 2 (Rent compression)**.** *On a monitored class, the agent's opportunistic surplus per claim falls from* $\mathbb{E}[b\cdot\mathbf{1}\{b>0\}]$ *to* $\mathbb{E}[(b-\theta^A F)_+]$*, a reduction of* $\mathbb{E}[\min\{b_+,\theta^A F\}]>0$ *whenever* $\theta^A F>0$ *and positive mass lies above zero. Along improvement paths satisfying false-challenge dominance—so that the monitored set is nondecreasing—aggregate opacity rent weakly falls absent endogenous obfuscation, strictly when a positive-measure set of classes crosses the frontier with* $\theta^A F>0$*. The restriction is necessary: where a rising challenged-honest burden drives a class's monitoring to exit, that class's opacity rent is restored, and the aggregate claim fails.*

6 Result 2: Crossing the Frontier Changes Behavior Before It Changes Recoveries

6.1 The threat does the work: verification without commitment

The architectural commitment of the baseline is natural for embedded machine monitoring, but the frontier does not depend on it. In the one-shot simultaneous inspection game—the agent misstates for gain b or not; the principal inspects at cost c or not; inspection prevents the loss L—so a caught agent forfeits the gain—and imposes consequence F; $0<c<L$ and $b, F>0$—the unique interior equilibrium is mixed, with the principal's indifference pinning misstatement and the agent's indifference pinning inspection:

$$p^*=\frac{c}{L},\qquad q^*=\frac{b}{b+F}. \tag{10}$$

Proposition 3 (Deterrence without inspection)**.** *In the interior equilibrium, a reduction in verification cost c lowers equilibrium misstatement one-for-one relative to L, while equilibrium inspection frequency* q^* *is unchanged. If* $c\geq L$ *inspection is dominated and misstatement is*

certain; c crossing below L creates the inspection equilibrium. The extensive-margin threshold therefore survives without commitment, and the observable response to cheaper verification is a change in counterparty behavior, not in audit counts.

The game is the classic inspection game (Tsebelis 1989; Graetz, Reinganum, and Wilde 1986); Khalil (1997) is the contracting treatment of the commitment problem. The game is deliberately minimal—one claim, homogeneous gain, perfect inspection, no fixed cost, volume, pipeline, or heterogeneity—so what survives without commitment is the extensive-margin logic of a cost threshold, not the full frontier (8) or its component comparative statics; extending the no-commitment analysis to the paper's primitives is open. Its empirical redirection is the point: credibility, rather than realized inspection frequency, is the operative channel—consistent with the Danish audit evidence (Kleven et al. 2011), the VAT paper trail (Pomeranz 2015), and of the automated-enforcement episode in which compliance moved by double digits while realized enforcement remained, for most of the population, a threat (Merane and Stremitzer 2026).

6.2 Recovery non-monotonicity

Verification systems are habitually evaluated by recovered money. Conditional on monitoring, expected detected-and-remedied loss per claim is

$$D = \rho^P \, L \, [1 - H(\theta^A F)], \tag{11}$$

rising in remediation effectiveness, falling in deterrence. For a technology path raising both derived reliabilities,

$$dD = L \, [1 - H(\theta^A F)] \, d\rho^P - \rho^P L F \, h(\theta^A F) \, d\theta^A, \tag{12}$$

with unrestricted net sign.

Proposition 4 (Recovery non-monotonicity)**.** *Suppose misstatement gains have bounded support ($b \leq \bar{b}$) and the technology path attains full deterrence ($\theta^A F \geq \bar{b}$ at its endpoint). Then measured recoveries are hump-shaped along the path: zero without monitoring, zero again at full deterrence, positive between—so they first rise and eventually fall, while avoided loss rises throughout. Short of those conditions, the sign of dD in (12) is unrestricted: recoveries need not rise with verification effectiveness, and an improvement can reduce observed recoveries even while reducing expected loss and opacity rent. In either case recovery is not a sufficient success metric; coverage, incidence, avoided loss, and counterparty behavior must be measured alongside it.*

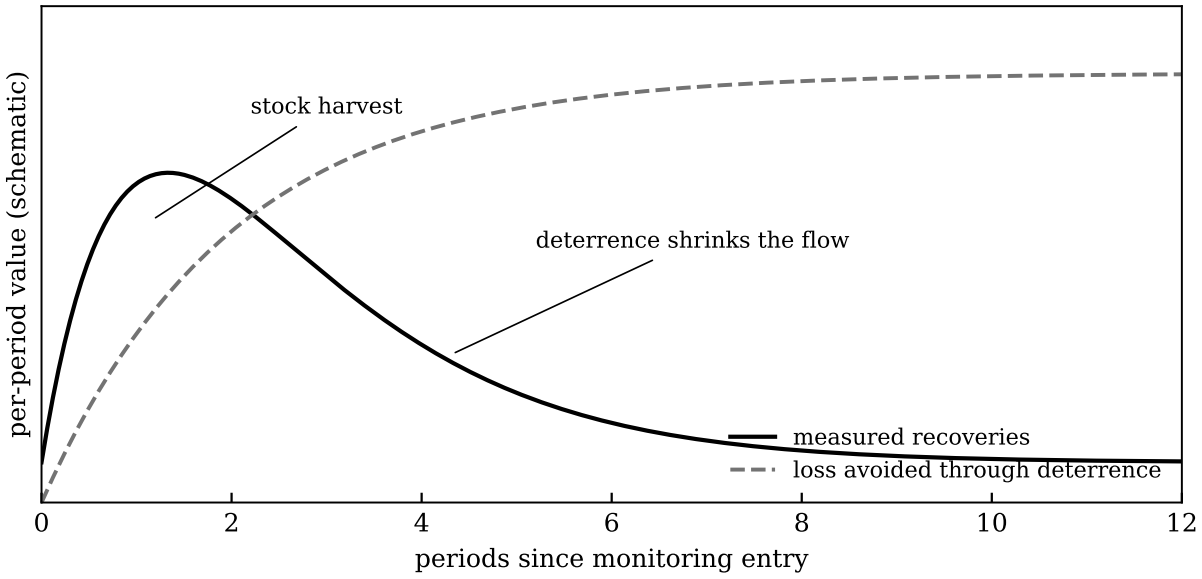

Figure 2: Recovery non-monotonicity (Proposition 4). At monitoring entry, measured recoveries rise while the pre-existing stock of misstatements is harvested, then fall as deterrence shrinks the flow; loss avoided through deterrence rises throughout. Evaluating a verification system by recovered money alone therefore misreads success as failure. Schematic.

The two-stage time path is the framework's most practically consequential prediction, because it inverts the natural evaluation instinct: a verification technology that is working should eventually look unproductive on a recovery dashboard. The existing evidence is consistent with this mechanism rather than a direct test of it: the deterrence value of US tax audits substantially exceeds their direct revenue (Boning et al. 2025; see also Kleven et al. 2011; Pomeranz 2015), and China's computerized VAT enforcement produced large short-run revenue gains that attenuated over time (Fan et al. 2020)—the time profile the proposition predicts. Section 8's designs are built to observe the full path in private markets. Figure 2 draws the two curves whose divergence is the point.

6.3 Obfuscation: the frontier fights back

If the agent can make verification harder—fragmented data, contractual complexity, rebundling, ambiguous categories, plausible but noisy documentation—improvement invites response. A related mechanism appears in Biais and Landier (2020), where agents endogenously adopt more complex technologies because complexity generates larger agency rents; that result motivates, but is distinct from, the verification-obfuscation extension developed here, in which complexity is chosen to degrade a counterparty's inspection pipeline. Strategic

complexity is itself established: Ellison and Wolitzky (2012) model sellers raising consumers' search costs to preserve pricing rents. The extension here is the target and the consequence: complexity is invested specifically to preserve verification-dependent opacity rents, and because it can push net verification surplus below zero, the frontier itself becomes endogenous. Let obfuscation $o \geq 0$ cost $K_o(o)$, convex increasing, with $\partial\theta^A/\partial o < 0$. A monitored agent obfuscates to the point where marginal complexity cost equals the marginal opacity rent preserved:

$$K_o'(o^*) = -F\left[1 - H(\theta^A F)\right] \frac{\partial \theta^A}{\partial o}, \tag{13}$$

and near the frontier there is a discrete motive to invest just enough complexity to push Δ_i below zero and deter verification entry altogether. Equation (13) is the tractable static core of a genuinely dynamic contest—Jin, Sokol, and Wagman (2026) show that when violators redesign faster than monitors recognize, persistent-redesign equilibria exist in which monitoring produces workload without harm reduction—and its empirical signature is substitution: when one margin becomes checkable, activity migrates to harder-to-verify margins, documented at national scale by Carrillo, Pomeranz, and Singhal (2017). A credible evaluation of machine verification must therefore search for margin substitution, not only error reduction; obfuscation expenditure itself is pure waste and reappears in the welfare ledger of Section 7.

7 Result 3: Private and Social Frontiers Diverge

7.1 Welfare from primitives

Verification consumes real resources to change, in large part, who holds money. Whether crossing the frontier is efficient therefore cannot be assumed; it must be derived. Recall the primitive of Section 4.1: a misstatement of type b transfers b to the agent and destroys $L - b \geq 0$ of real value. Deterring a claim of type b saves society the waste $L - b$; the transfer component nets out. Remediation, by contrast, reverses the transfer after the waste is sunk: its direct social value is approximately zero, net of the process costs already counted in c. Sanctions F are transfers. Let $\varphi_S \geq \varphi_P$ be the full social cost of a false challenge—both parties' escalation, relationship, and procedural costs—of which the principal internalizes φ_P. Verification resource costs K and c are identical in both ledgers. The social per-claim benefit of monitoring class i is then

$$\beta_{S,i}(\tau) = \int_0^{\theta^A F} (L_i - b)\, dH_i(b) \;-\; f_i\, \varphi_{S,i}\, H_i(\theta^A F) \;+\; S_i(\tau), \tag{14}$$

where $S_i \geq 0$ collects deterrence spillovers the principal does not capture: the same monitoring architecture, or the same agent's induced compliance, improving behavior in parallel relationships, class-wide markets, or rule-following generally. The planner monitors when $N\beta_S \geq K + Nc$, defining a social frontier $L_S^*(\tau, N)$ by the analogue of (8).

7.2 The divergence and its sign

Comparing (14) with the private benefit (4)–(5):

$$\beta_{P,i} - \beta_{S,i} = \int_0^{\theta^A F} b\, dH_i(b) \;+\; \rho_i^P\, L_i\, [1 - H_i(\theta^A F)] \;+\; (\varphi_{S,i} - \varphi_{P,i})\, f_i\, H_i(\theta^A F) \;-\; S_i. \quad (15)$$

Proposition 5 (Frontier divergence)**.** *The private and social verification frontiers coincide only by accident. The wedge (15) decomposes into three forces pushing toward excessive private verification—the transfer component of deterrence (the principal counts deterred types' full L as savings; society counts only the waste $L - b$), remediation (a pure transfer the principal values at $\rho^P L$), and externalized dispute costs of false challenges—against one force pushing toward insufficient private verification: uninternalized deterrence spillovers. Accordingly: (i) when transfers dominate (b close to L, remediation-heavy monitoring, dispersed dispute costs), $L_P^* < L_S^*$: a region of privately profitable, socially wasteful verification—a verification race in Hirshleifer's (1971) sense, machine-financed; (ii) when spillovers dominate (dispersed victims, class-wide compliance responses, small per-principal stakes), $L_P^* > L_S^*$: private enforcement is undersupplied; (iii) efficient entry is the knife-edge at which spillovers exactly offset the transfer terms.*

Remark (relation to prior work). The existence and direction of the wedge in (15) are not new: the transfer term and the externalized dispute cost are machine-age instances of Shavell's (1982, 1997) divergence, and the over-enforcement region is Landes and Posner's (1975). Three things are new. The divergence is located at the extensive margin, determining which claim classes are monitored at all rather than litigation intensity conditional on a dispute. The wedge is technology-dependent, so improvements in τ can change its sign. And part of the wedge is generated by the verification technology's own error structure—the false-challenge term $(\varphi_S - \varphi_P)\, f\, H$ has no counterpart in fixed-technology models of the divergence—so the divergence moves when the technology moves.

Both corners are already observable in the field. The dispersed-victim corner is the automated-enforcement episode: one enforcer's mechanized threat produced a market-wide, double-digit compliance response that no individual website visitor had an incentive to finance—textbook undersupply of general deterrence, resolved in that case by a profit motive

(Merane and Stremitzer 2026). And the same profit motive, pointed at imperfect precision and asymmetric dispute costs, finances the opposite corner: automated challenge of the honest, profitable in practice because settlement is cheaper than defense. The present model prices false challenges only as costs—no challenger here collects settlement revenue—so profitable nuisance is an interpretation the framework points at rather than an equilibrium it contains; the challenger's settlement income and the resulting nuisance region are formalized in the companion paper (Gill 2026b). It remains the recognized pathology of automated cease-and-desist industries, and the reason those authors conclude the technology pressures policymakers to recalibrate legal rules: the $f\varphi$ term, monetized by a party this model does not yet pay.

Two honest caveats bound the result. First, the static ledger understates deterrence's social value: in competitive and repeated settings, anticipated opportunism is priced ex ante, distorting participation, quality provision, and defensive investment; deterrence then restores surplus the transfer accounting misses, so the over-verification bias in (i) is an upper bound whose tightening is left to future work. The same refinement applies within the accounting itself: a refund is restitution (a transfer), repair of defective performance restores real value, compliance investment can be productive rather than deadweight, and process costs are real resource losses—the frontier of (15) prices these with single parameters, and finer accounting moves the wedge's magnitude without changing its structure. Second, the policy levers map one-to-one onto the model's objects, which is what makes the welfare analysis usable: statutory materiality and de minimis thresholds raise L^* directly; fee-shifting and bad-faith sanctions reprice the $f\varphi$ asymmetry that finances nuisance; evidence standards set q_E; safe harbors for certified claims accelerate the benign second-stage equilibrium of Section 9.4. Post-frontier economies will need those levers, because Proposition 5 says the market will not find the social frontier on its own.

8 Evidence and Measurement

8.1 An illustrative case: the Google Fonts episode

The framework's closest existing field evidence is a case the model organizes well—read here as illustration rather than test, since the framework was developed with the episode in hand. After a German court ruled that dynamically embedding Google Fonts violated the GDPR, an Austrian lawyer deployed automated detection and claim generation against websites at scale; non-compliance in Austria fell 22.7 percentage points—nearly half—within three months (Merane and Stremitzer 2026). In the model's terms: a single adjudicated

rule set $q_V \approx 1$ and supplied q_L; web crawling collapsed c and raised q_D for a claim class of enormous N and individually trivial L; automated document assembly compressed the evidentiary and claim-generation components. The technology shock was concentrated in exactly the components Proposition 2 says unlock high-volume, low-value classes—and the episode is consistent with the framework's qualitative structure: claims that sat below the human frontier crossed it (Result 1); behavior moved economy-wide through threat rather than realized enforcement (Proposition 3), with the enforcement instrument scarcely fired relative to the compliance response (Proposition 4's mechanism); the market-wide response to one enforcer's investment instantiates the undersupply corner of Proposition 5, while the same episode's disruptive-enforcement controversy instantiates the nuisance corner. What the episode does not test is the general framework across heterogeneous commercial claim classes with amortized fixed costs—the setting the model is built for, and the gap the following design addresses.

8.2 The measurement framework

The unit of observation is the claim class—a stable obligation-rule pair over repeated transactions—mapping directly to i. The minimum measurement set is: all-in verification cost per class, separated into fixed and marginal components and inclusive of human escalation and enforcement; monitored status and intensity; true-discrepancy and false-positive rates at stated sensitivity and specificity; verified discrepancy value, realized recovery, and the pipeline stage of failure; and subsequent counterparty pricing, documentation, error, and complexity behavior. The design objects are the ex ante verification ratio and its distribution,

$$VR_i = \frac{N_i \, \beta_i}{K_i + N_i \, c_i}, \tag{16}$$

with a verification-cost shock predicting a positive mass of classes crossing $VR = 1$, concentrated where the pre-shock density near the frontier is greatest; the coverage change $\Delta \Pr(m_i = 1)$; and—because Proposition 4 makes recovery non-monotonic—a realization ratio (recovered or avoided loss over verified discrepancy value) reported alongside incidence, with the two-stage time path of Section 6.2 as the sharpest test. Credible designs randomize verification availability or intensity across comparable suppliers, divisions, or claim classes, or exploit staggered rollout with plausibly exogenous timing; the standard is set by the difference-in-differences design of Merane and Stremitzer (2026). Because the theory predicts deterrence and substitution, pre-registration of claim classes and outcomes—with conformity and optimality claim classes logged as separate objects from the first audit onward, since the decomposition cannot be reconstructed retroactively—disclosure of false-positive and

dispute outcomes, and reproducible evidence trails are not refinements; they are the difference between measuring the mechanism and advertising a tool.

8.3 Energy billing as the structured proving ground

Commercial energy billing satisfies all five exposure conditions with unusual cleanliness: bills are generated from explicit contracts, tariff formulas, metered consumption, network charges, taxes, and market indices, so a discrepancy ties to a specific formula, period, index, meter series, or network component, and the path from claim to evidence is short. Claim classes are numerous, individually small, and high-volume—precisely the $L^*(\tau, N)$ region Proposition 2 says crosses first—and the institutional layer (billing correction, contractual remedy, regulated network components) supplies q_E and q_L at low cost. The hypothesis is not that reconciliation is technically novel; it is that a sufficiently large fall in all-in cost moves claim classes previously ignored—because expected recovery was smaller than manual effort—across $VR = 1$. The design measures cost per verified billing component, the share of components economically monitored, verified leakage, recovery, false-positive and dispute outcomes, and supplier behavior before and after automated verification: the distribution of VR_i crossing one, observed in the field, with Proposition 4's time path and Section 6.3's substitution as the follow-on tests.

9 Extensions and Boundary Conditions

9.1 Correlated error and effective verification quality

A second model checking a first model is not automatically independent verification: shared training data, retrieval sources, architectural biases, and benchmarks correlate errors. The effective q_V in (2) is conditional reliability after accounting for correlation, not nominal accuracy; C_V is a cost-quality frontier over sensitivity, specificity, confidence, and domain. Wang (2026) argues for statistical limits of auditing in rare-error regimes—as systems improve, verifying them becomes harder at a fundamental rate—and Kiyani et al. (2026) propose optimal escalation between cheap weak and costly strong verification. A technology that lowers nominal checking cost by tolerating more false positives has lowered c while raising f, and equation (8) prices both. More generally, a market flooded with cheap pseudo-verification raises measured reliability while lowering effective reliability, which is Corollary 1 territory.

9.2 The enforcement bottleneck

If $q_E\, q_L$ is near zero, Corollary 1 applies regardless of detection quality: the dashboard economy, discrepancies everywhere and consequences nowhere. The regulatory version is formalized by Jin, Sokol, and Wagman (2026)—below an institutional-investment threshold, AI-augmented monitoring raises workload without reducing harm—and the private version predicts where machine verification disappoints: weak remedies, hostile data access, evidence standards machine output cannot yet satisfy. The prediction is testable at the pipeline-stage level: Section 8.2's failure-stage variable exists to locate the binding constraint.

9.3 Verifiability as a contracted variable

The deepest institutional difference from public enforcement is that a sovereign compels reporting while private parties must contract for verifiability: data provenance, calculation formulas, structured evidence fields, audit rights, escalation rules, liability allocation. Cheap verification raises the return to writing obligations in checkable form, so contracts become part of the verification architecture—and ambiguity becomes a purchasable instrument of opacity, priced at the discount attached to unverifiable claims. The predicted complementarity—verification technology and contractual machine-readability moving together—is the transition's cleanest leading indicator, observable in new contract vintages before any dispute occurs.

9.4 Certification as a second-stage implication

The agent's strategy set exceeds misstate-or-obfuscate: the agent can pre-certify, because a verification-cost shock lowers the cost of *being verified* as much as the cost of verifying. The disclosure literature supplies the consequence. With certifiable information and negligible certification cost $\kappa(\tau)$, $\kappa' < 0$, unraveling logic applies (Grossman 1981; Milgrom 1981): once κ falls below the price or bargaining discount attached to unverifiable claims, honest types certify, non-certification becomes informative of misstatement, and the deterrence cutoff for uncertified claims rises; positive κ and uncertainty sustain partial certification (Jovanovic 1982; Verrecchia 1983; Dye 1985). Where certification requires an intermediary with recognized standing, the certifier can strategically coarsen information and capture surplus (Lizzeri 1999; Dranove and Jin 2010)—opacity rents migrating toward certification, liability-bearing, and recognized assurance, the distributional prediction shared with Catalini, Hui, and Wu (2026). An operating instance already exists in research itself: third-party certification of computational reproducibility, in which an independent agency verifies results on behalf of journals and readers (Pérignon et al. 2019). The arc—opacity, then adversarial

contestability, then voluntary verifiability—is a research agenda rather than a theorem here. The companion paper (Gill 2026b) develops the institutional half of that agenda as results rather than conjectures: certifiers lose the amortization function and—because adoption of a shared certifier is strategically complementary—can collapse by tipping rather than decay; brand value migrates rather than dies, its bonding component compressing toward the cost of credibly attested verification while search and taste components persist, with the online-review era as the closest historical analogue; optimality rents—mismatch between a relationship's contract and its attainable counterfactual, the second object defined above—compress through repricing and exit down to switching friction, with incumbents repricing before customers switch; cheap accusation expands nuisance challenge unless rebuttal costs fall symmetrically or fee-shifting restores the balance; and opacity relocates toward unverifiable attributes. A third paper is to carry the measurement program of Section 8 in commercial energy markets, logging conformity and optimality claim classes separately with the construction-cost series $K(t)$ as the novel measured object. This paper supplies the technology and the frontier; the companion supplies the institutions; the measurement supplies the test. Its empirical trace is already visible in an adjacent domain: the open-science reforms—pre-registration, registered reports, mandated data deposit—are a knowledge-market population choosing verifiability once being checkable became cheaper than being challenged.

10 The Post-Opacity Transition

The broader regime deserves a name and a precise definition, because the analogy it invokes is exact and the imprecise version of the claim is wrong. Post-scarcity names the regime implied by collapsing production costs—never universal abundance, but a movement of the binding constraint. Post-opacity, as defined here, names the regime implied by collapsing verification costs: not universal transparency, but a movement of the binding constraint on trust—from the cost of checking to the institutions of evidence, enforcement, and certification. Formally, with T the set of technically verifiable claims and $S(\tau)$ the economically monitored set of equation (6), the wedge $T \setminus S(\tau)$ is latent verifiability: what an economy can check but rationally does not. A post-opacity transition is a contraction of that wedge caused by verification technology—defined claim class by claim class, which is what makes it measurable rather than metaphorical, and conditional on the pipeline, which is what keeps it from being futurism.

What the transition re-selects

The deepest consequence of a verification-cost collapse is not more verification. It is the re-selection of institutions that exist because verification was expensive. A large share of commercial architecture is adaptation to costly checking rather than a response to misstatement: fixed-price contracts chosen where cost verification was uneconomic; brands posted as forfeitable bonds for quality no buyer could inspect (Klein and Leffler 1981); grading, standardization, and commoditization as economies of measurement (Barzel 1982); statistical sampling, materiality thresholds, and episodic audit as rationing devices for scarce verification; escrow, letters of credit, and payment-term conventions as trust infrastructure priced at the cost of checking; intermediaries whose margin is, in part, a verification wage. Each embeds a shadow price of verification calibrated to the human technology. Where the full pipeline clears, that shadow price collapses, and the structures built on it lose their economic foundation. They re-form through anticipation, unraveling, and contract redesign rather than through realized inspection. The model says why the transition takes this form: behavior moves on threat rather than frequency (Proposition 3), monitoring volume is a transient (Proposition 4), and voluntary verifiability is the attractor once being checkable is cheaper than being challenged (Section 9.4). Domain-specific precedents already exist: audit design shifted procurement officials' choice of procedure in Chile (Gerardino, Litschig, and Pomeranz 2024), and AI adoption is reorganizing audit work itself (Fedyk et al. 2022; Law and Shen 2025). Post-opacity generalizes such responses into a prediction: systematic re-selection across commercial institutions when the underlying verification technology experiences a sufficiently broad cost shock. Where the full evidence and enforcement pipeline clears, arrangements selected under expensive checking lose part of their economic rationale—opacity rents compress, contracting and certification adapt, and institutions are re-selected toward structures designed to be economically verifiable. The visible signature of a post-opacity transition is institutional change, not an audit boom.

Why this is not the previous trust prophecy

A version of this claim has been made before, unconditionally, and it failed. That failure is why the conditions in this paper are the contribution rather than the hedge. The blockchain literature identified verification cost as the economic object a decade ago, and its popular wing forecast that collapsing trust costs would reorganize commerce (Catalini and Gans 2020 is the serious statement of the economics). The forecast aged badly for two reasons this paper formalizes: verifiability-in-principle required migrating economic activity onto new rails, and detection without an evidentiary and enforcement pipeline changes nothing—the dashboard

limit of Corollary 1. Machine verification differs in the decisive respect that it reads the existing substrate: contracts, invoices, ledgers, and meter feeds as they are, attacking the fixed and marginal costs of checking legacy claims without migration. The same conditions bound the present claim: the transition is gated by the pipeline (Corollary 1), contested by strategic complexity (Section 6.3), stalled where enforcement is congested (Section 9.2), and confined to T—formal, data-represented claims—leaving judgment and tacit quality untouched. Within those bounds, and only within them, the dissolution of opacity-dependent practices is the equilibrium prediction, not a hope.

Where the mass is

The applications follow the frontier's logic to wherever formal claims accumulate in volume. In private trade: contract form, the basis of intermediation, and the pricing of unverifiable claims (Section 9.3). In insurance: claim categories that are individually small, formula-governed, and repeated. In tax administration: the transition already realized at sovereign scale (Section 2). And in public spending, where the stakes are stated in official accounts: U.S. federal agencies alone reported an estimated $186 billion in improper payments for fiscal year 2025 across 64 programs, with cumulative estimates of roughly $3 trillion since 2003 (GAO 2026). That these are overwhelmingly formula-eligible, document-based claim classes of the kind the frontier governs is this paper's reading of their composition, not GAO's conclusion. Improper-payment control at claim-class scale is the public-sector face of the same transition.

The research program reduces to one quantity, stated with the refinement the pipeline forces:

How much economic activity lies below the human verification frontier, above the machine verification frontier—and above the enforcement floor that Corollary 1 defines?

Every parameter of the answer—c, K, f, the pipeline components, the density of G near L^*—is an object Section 8's design estimates in real transactional domains.

11 Limitations

Seven boundaries deserve explicit statement. The monitoring architecture is one-shot and technology-first; richer dynamics—history-dependent inspection, reputation, learning, and the adaptive redesign contest of Jin, Sokol, and Wagman (2026)—are not modeled. The false-challenge-dominance condition is sufficient, not necessary, and the paper does not

characterize optimal monitoring when it fails, where mixed architectures (monitor, then relax) may arise. Gain and loss distributions are treated as known; real systems learn them, and selection into monitoring complicates inference. Verifier error may correlate with the systems generating claims, and statistical limits bind in rare-error regimes (Wang 2026). The welfare analysis prices false positives with a single dispute-cost parameter and treats spillovers in reduced form; the industrial organization of nuisance enforcement and the general-equilibrium pricing of deterrence are richer than equation (15). The certification implication of Section 9.4 is not a solved equilibrium. And not every economically valuable fact is objectively verifiable: preferences, long-horizon strategy, artistic value, and much medical and legal judgment resist cheap verification under any technology—the frontier moves within T, and T is not everything.

12 Conclusion

The first economic effect of modern AI is a collapse in the cost of doing cognitive work. The second, slower and more institutional, is a collapse in the cost of checking claims that were always checkable in principle. Many information asymmetries survive not because truth is unknowable but because establishing it costs more than it returns; those are not states of ignorance but states of economically protected opacity, and they have financed a component of intermediation, billing complexity, and small-scale misstatement for as long as inspection has been expensive.

This paper's claim is deliberately specific. AI changes the extensive margin of economically rational verification: it turns previously non-contestable small claims into contestable ones, especially where claim volume amortizes fixed verification infrastructure; which claims cross depends on which components of the technology vector move, and no amount of detection substitutes for evidence and enforceability; incentives change before enforcement occurs, so the technology's success ultimately shows up as fewer recoveries rather than more; and because much of what verification prevents is a transfer, the market alone will not find the socially right frontier—private verification will race past it where transfers dominate and stall short of it where deterrence is a public good. Beneath the specific claim sits the larger one that Section 10 states with its conditions: commerce is dense with institutions that exist because checking was expensive—contract forms, brands, grades, thresholds, intermediaries—and a technology that collapses the cost of checking legacy claims re-selects them. The general-equilibrium version of that statement, in which verification technology redraws contract form, intermediation, and the boundaries of firms, is a separate paper; Baker and Hubbard (2004) document the closest historical instance, in which on-board computers changed asset

ownership in trucking by changing what was contractible. Here it is enough to have built the frontier that makes the question tractable.

The frontier is equation (8); measuring its movement is the research program this paper defines.

A Proofs

Proof of Proposition 1. (a) $\Delta_i(t) = N_i\beta_i(t) - K_i(t) - N_i c_i(t)$ is continuous; $\Delta_i(t_L) < 0 < \Delta_i(t_H)$ gives a crossing by the intermediate value theorem. (b) Write $N^{-1}\, d\Delta/dt = L\, dA/dt - \varphi\, d\Phi/dt - dc/dt - N^{-1} dK/dt$. Along an improvement path $dA/dt = (d\rho^P/dt)[1 - H(\theta^A F)] + (1-\rho^P)\, h(\theta^A F)\, F\, (d\theta^A/dt) \geq 0$; FCD gives $d\Phi/dt \leq 0$; $dc/dt \leq 0$ and $dK/dt \leq 0$. All four terms are nonnegative, and strictness of any one on every interval makes Δ strictly increasing, giving uniqueness of the crossing and the stated monitoring pattern. (c) Take $dc = dK = df = d\rho^P = 0$ and $d\theta^A/dt > 0$ with $f > 0$. Then $dA/dt = (1-\rho^P)\, h(\theta^A F)\, F\, (d\theta^A/dt)$ and $d\Phi/dt = f\, h(\theta^A F)\, F\, (d\theta^A/dt)$, so $N^{-1}\, d\Delta/dt = h(\theta^A F)\, F\, (d\theta^A/dt)\, [\, L(1-\rho^P) - \varphi f\,]$. For $\varphi f > L(1-\rho^P)$—cheap-to-trigger, costly disputes alongside near-complete remediation—Δ is strictly decreasing along an unambiguously improving path, establishing failure of monotonicity. The construction takes ρ^P locally unresponsive to the moving component, permitted under weak monotonicity; for strictly increasing ρ^P the same failure obtains whenever $\varphi f h F\, (d\theta^A/dt) > L\, [\, (d\rho^P/dt)(1-H) + (1-\rho^P)\, hF\, (d\theta^A/dt)\,]$. □

Proof of Proposition 2 and Corollary 1. Rearranging (6) with (4)–(5) gives $LA \geq c + K/N + f\varphi H(\theta^A F)$, i.e. (8). Signs (i)–(iii) are immediate partial derivatives. For (iv), let a pipeline component q rise, moving θ^A and ρ^P up: $dL^*/dq = [\, \varphi f h F \theta^A_q\, A - (c + K/N + f\varphi H)\, A_q\,]/A^2$, negative iff $(c + K/N + f\varphi H)\, A_q > \varphi f h F \theta^A_q\, A$, the gain-dominance condition; at $f = 0$ the numerator effect vanishes and $dL^*/dq = -L^* A_q/A < 0$. For Corollary 1, ρ^P and θ^A inherit the factor $q_E q_L$ (remediation requiring the same stages), so $q_E q_L \to 0$ forces $A \to 0$ while the numerator of (8) stays bounded away from zero ($c > 0$), hence $L^* \to \infty$. □

Proof of Corollary 2. On a monitored class the agent's surplus per claim is $\mathbb{E}[(b - \theta^A F)_+]$; the reduction from $\mathbb{E}[b \cdot \mathbf{1}\{b > 0\}]$ equals $\mathbb{E}[\min\{b_+, \theta^A F\}] > 0$ under the stated conditions, and $\frac{d}{d\theta^A}\mathbb{E}[(b - \theta^A F)_+] = -F[1 - H(\theta^A F)] \leq 0$ gives weak decline along improvement paths within monitored classes; classes crossing the frontier take the discrete reduction. False-challenge dominance ensures no monitored class exits along the path (Proposition 1), so no class's rent is restored; without it, exiting classes revert to $\mathbb{E}[b \cdot \mathbf{1}\{b > 0\}]$ and the aggregate can rise. □

Proof of Proposition 3. Principal: inspect $-c$ versus not-inspect $-pL$; indifference $p^* = c/L$. Agent: because inspection blocks the gain and imposes F, misstating yields $(1-q)b - qF$

versus 0; indifference gives $q^* = b/(b+F)$. Under $0 < c < L$ and $b, F > 0$ both lie in $(0,1)$; $\partial p^*/\partial c = 1/L > 0$; q^* is independent of c. If $c \geq L$, inspection is weakly dominated at $c = L$ and strictly above, and misstatement is certain. □

Proof of Proposition 4. Recoveries per claim are zero before monitoring entry and $D(t) = \rho^P(t)L[1 - H(\theta^A(t)F)]$ after it. At entry $D > 0$: the pre-existing stock of misstatements is verified at rate ρ^P. If the path attains $\theta^A F \geq \bar{b}$, then $1 - H(\theta^A F) = 0$ there and $D = 0$ again. D is continuous on the monitored segment and positive on its interior, so it attains an interior maximum: recoveries rise from zero and return to zero, the asserted hump. Avoided loss $L\,H(\theta^A F)$ is nondecreasing along the path throughout. Without bounded support or full deterrence, equation (12)—the differential of (11)—has two terms of opposite sign and unrestricted relative magnitude, which is the ambiguity statement. The stock-and-flow narration is the interpretation of the two segments of this static comparative-statics path, not an additional dynamic model. □

Proof of Proposition 5. Subtract (14) from $\beta_P = L[H(\theta^A F) - H(0)] + \rho^P L[1 - H(\theta^A F)] - f\varphi_P H(\theta^A F)$, using $\int_0^{\theta^A F}(L - b)\,dH = L[H(\theta^A F) - H(0)] - \int_0^{\theta^A F} b\,dH$: the deterrence terms differ by $\int_0^{\theta^A F} b\,dH \geq 0$; the remediation term $\rho^P L[1 - H(\theta^A F)] \geq 0$ appears only privately; the false-positive terms differ by $(\varphi_S - \varphi_P)fH \geq 0$; spillovers S appear only socially. Equation (15) follows, and the frontier comparisons follow from monotonicity of the threshold in the per-claim benefit. □

References

Alles, Michael G., Alexander Kogan, and Miklos A. Vasarhelyi. 2002. "Feasibility and Economics of Continuous Assurance." *Auditing: A Journal of Practice & Theory* 21(1): 125–138.

Allingham, Michael G., and Agnar Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1(3–4): 323–338.

Baker, George P., and Thomas N. Hubbard. 2004. "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking." *Quarterly Journal of Economics* 119(4): 1443–1479.

Ball, Ian, and Deniz Kattwinkel. 2025. "Probabilistic Verification in Mechanism Design." *Theoretical Economics* 20(4): 1247–1284.

Barzel, Yoram. 1982. "Measurement Cost and the Organization of Markets." *Journal of Law and Economics* 25(1): 27–48.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76(2): 169–217.

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman. 2014. "Optimal Allocation with Costly Verification." *American Economic Review* 104(12): 3779–3813.

Biais, Bruno, and Augustin Landier. 2020. "Endogenous Agency Problems and the Dynamics of Rents." *Review of Economic Studies* 87(6): 2542–2567.

Boning, William C., Nathaniel Hendren, Ben Sprung-Keyser, and Ellen Stuart. 2025. "A Welfare Analysis of Tax Audits Across the Income Distribution." *Quarterly Journal of Economics* 140(1).

Border, Kim C., and Joel Sobel. 1987. "Samurai Accountant: A Theory of Auditing and Plunder." *Review of Economic Studies* 54(4): 525–540.

Bustos, Sebastián, Dina Pomeranz, Juan Carlos Suárez Serrato, José Vila-Belda, and Gabriel Zucman. Forthcoming. "The Race Between Tax Enforcement and Tax Planning: Evidence from a Natural Experiment in Chile." *American Economic Journal: Economic Policy*. NBER Working Paper 30114.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal. 2017. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics* 9(2): 144–164.

Catalini, Christian, and Joshua S. Gans. 2020. "Some Simple Economics of the Blockchain."

Communications of the ACM 63(7): 80–90. Earlier version: NBER Working Paper 22952 (2016).

Catalini, Christian, Xiang Hui, and Jane Wu. 2026. "Some Simple Economics of AGI." MIT Sloan Research Paper, February. arXiv:2602.20946.

Colliard, Jean-Edouard, Christophe Hurlin, and Christophe Pérignon. 2024. "The Economics of Computational Reproducibility." HEC Paris Research Paper 1521. SSRN 3418896.

Commerford, Benjamin P., Sean A. Dennis, Jennifer R. Joe, and Jenny W. Ulla. 2022. "Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence." *Journal of Accounting Research* 60(1): 171–201.

Crescitelli, Viviana, Generoso Immediato, Fabio Persia, and Stefania Costantini. 2026. "AI Evaluation Should Measure Verification Cost, Not Correctness Alone." arXiv:2608.08709.

Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48(4): 935–963.

Dye, Ronald A. 1985. "Disclosure of Nonproprietary Information." *Journal of Accounting Research* 23(1): 123–145.

Ellison, Glenn, and Alexander Wolitzky. 2012. "A Search Cost Model of Obfuscation." *RAND Journal of Economics* 43(3): 417–441.

Fan, Haichao, Yu Liu, Nancy Qian, and Jaya Wen. 2020. "The Effects of Computerizing VAT Invoices in China." NBER Working Paper 24414.

Fedyk, Anastassia, James Hodson, Natalya Khimich, and Tatiana Fedyk. 2022. "Is Artificial Intelligence Improving the Audit Process?" *Review of Accounting Studies* 27(3): 938–985.

Gale, Douglas, and Martin Hellwig. 1985. "Incentive-Compatible Debt Contracts: The One-Period Problem." *Review of Economic Studies* 52(4): 647–663.

Gans, Joshua S. 2019. "The Fine Print in Smart Contracts." NBER Working Paper 25443.

Gerardino, Maria Paula, Stephan Litschig, and Dina Pomeranz. 2024. "Distortion by Audit: Evidence from Public Procurement." *American Economic Journal: Applied Economics* 16(4): 71–108.

Goldfarb, Avi, and Catherine Tucker. 2019. "Digital Economics." *Journal of Economic Literature* 57(1): 3–43.

Gill, Amrit. 2026b. "Machine Verification and the Institutions of Trust: Reputation, Certification, and Contract when Checking Becomes Cheap." Working paper.

Graetz, Michael J., Jennifer F. Reinganum, and Louis L. Wilde. 1986. "The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement." *Journal of Law, Economics,*

and Organization 2(1): 1–32.

Grossman, Sanford J. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics* 24(3): 461–483.

Halac, Marina, and Pierre Yared. 2020. "Commitment versus Flexibility with Costly Verification." *Journal of Political Economy* 128(12).

Hirshleifer, Jack. 1971. "The Private and Social Value of Information and the Reward to Inventive Activity." *American Economic Review* 61(4): 561–574.

Jin, Ginger Zhe, D. Daniel Sokol, and Liad Wagman. 2026. "Adaptive Enforcement with AI-Augmented Monitoring." NBER Working Paper 35010.

Jovanovic, Boyan. 1982. "Truthful Disclosure of Information." *Bell Journal of Economics* 13(1): 36–44.

Khalil, Fahad. 1997. "Auditing without Commitment." *RAND Journal of Economics* 28(4): 629–640.

Kiyani, Shayan, Sima Noorani, George Pappas, and Hamed Hassani. 2026. "When to Trust the Cheap Check: Weak and Strong Verification for Reasoning." arXiv:2602.17633.

Klein, Benjamin, and Keith B. Leffler. 1981. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy* 89(4): 615–641.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica* 79(3): 651–692.

Landes, William M., and Richard A. Posner. 1975. "The Private Enforcement of Law." *Journal of Legal Studies* 4(1): 1–46.

Law, Kelvin K. F., and Michael Shen. 2025. "How Does Artificial Intelligence Shape Audit Firms?" *Management Science*.

Lizzeri, Alessandro. 1999. "Information Revelation and Certification Intermediaries." *RAND Journal of Economics* 30(2): 214–231.

Merane, Jakob, and Alexander Stremitzer. 2026. "Automated Private Enforcement: Evidence from the Google Fonts Case." *Journal of Legal Analysis* 18(1): 182–206.

Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics* 12(2): 380–391.

Mookherjee, Dilip, and Ivan Png. 1989. "Optimal Auditing, Insurance, and Redistribution." *Quarterly Journal of Economics* 104(2): 399–415.

Naritomi, Joana. 2019. "Consumers as Tax Auditors." *American Economic Review* 109(9): 3031–3072.

Pérignon, Christophe, Kamel Gadouche, Christophe Hurlin, Roxane Silberman, and Eric Debonnel. 2019. "Certify Reproducibility with Confidential Data." *Science* 365(6449): 127–128.

Polinsky, A. Mitchell. 1980. "Private versus Public Enforcement of Fines." *Journal of Legal Studies* 9(1): 105–127.

Pomeranz, Dina. 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review* 105(8): 2539–2569.

Reinganum, Jennifer F., and Louis L. Wilde. 1985. "Income Tax Compliance in a Principal-Agent Framework." *Journal of Public Economics* 26(1): 1–18.

Shavell, Steven. 1982. "The Social versus the Private Incentive to Bring Suit in a Costly Legal System." *Journal of Legal Studies* 11(2): 333–339.

Shavell, Steven. 1997. "The Fundamental Divergence between the Private and the Social Motive to Use the Legal System." *Journal of Legal Studies* 26(2): 575–612.

Shi, Maggie. 2024. "Monitoring for Waste: Evidence from Medicare Audits." *Quarterly Journal of Economics*.

Silva, Francisco. 2020. "The Importance of Commitment Power in Games with Imperfect Evidence." *American Economic Journal: Microeconomics* 12(4): 99–113.

Slemrod, Joel. 2007. "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives* 21(1): 25–48.

Townsend, Robert M. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21(2): 265–293.

Tsebelis, George. 1989. "The Abuse of Probability in Political Analysis: The Robinson Crusoe Fallacy." *American Political Science Review* 83(1): 77–91.

U.S. Government Accountability Office (GAO). 2026. *Payment Integrity: Agencies' Estimated Improper Payments Increased to $186 Billion in Fiscal Year 2025.* GAO-26-108694. Washington, DC, April.

Varas, Felipe, Iván Marinovic, and Andrzej Skrzypacz. 2020. "Random Inspections and Periodic Reviews: Optimal Dynamic Monitoring." *Review of Economic Studies* 87(6): 2893–2937.

Verrecchia, Robert E. 1983. "Discretionary Disclosure." *Journal of Accounting and Economics* 5: 179–194.

Wang, Jason Z. 2026. "The Verification Tax: Fundamental Limits of AI Auditing in the Rare-Error Regime." arXiv:2604.12951.

Zhang, Hong. 2026. "The Division of Verification: Adjudication, Liability, and the Boundary of the Firm in Checking AI." SSRN Working Paper 7086002.