The Human Layer Economics: Capital Incentives, Organizational Automation Bias, and the Structural Case for Governance Investment

Ahmad Noureddine

Timer · ORCID 0009-0003-1669-4553

2026-05-09 · DOI 10.5281/zenodo.20096568

Abstract

The first three papers in this series established why the Human Layer matters, how to build it, and how to measure it. This paper addresses the question those papers raise but do not answer: if the evidence for augmentation over automation is this consistent, why does capital continue to flow toward replacement? The answer is not that executives are irrational. It is that the current incentive architecture makes removing the human layer structurally rational at the individual and organizational level, while externalizing the costs to a later period, a different line item, and often a different team. This paper identifies three capital incentive structures driving organizations toward automation: venture capital pricing dynamics, AI vendor pricing model evolution, and executive performance measurement. It introduces the concept of organizational automation bias: the systematic institutional tendency to overweight visible, near-term automation benefits and underweight deferred oversight costs. It then documents the three forces now collapsing the temporal gap that makes this externalization possible: regulatory enforcement becoming actualized, insurance and liability pricing beginning to incorporate AI risk, and institutional market access emerging as a gating function rather than a compliance aspiration. The paper closes with an empirically grounded argument, drawn from the SOX, GDPR, and ISO 27001 precedents, that the Human Layer Score introduced in Paper 3 is positioned to function as a capital markets signal for organizations operating in regulated, trust-dependent environments, and that governance infrastructure of this kind has historically been acquired most advantageously in the window before enforcement concentrates the market toward already-compliant actors.

Published at: ahmad.pt/research

Contents

The first paper in this series asked a straightforward question: if AI systems designed to amplify human judgment outperform those designed for replacement by a factor of three, why does the replacement narrative dominate? [18] Paper 1 established the economic case for augmentation and named five architectural components required for accountable human-AI interaction. Paper 2 delivered the formal specification for those components, defining how decision gates, escalation protocols, accountability structures, override mechanisms, and trust calibration interfaces must be built to function as a dependency graph rather than a checklist. [21] Paper 3 introduced the measurement layer: a scoring instrument that distinguishes whether a human layer is structurally sound, decorative, or absent. [25]

Paper 3 closed with a deliberate deferral. The Human Layer Score quantifies architectural maturity. The economic implications of that score, including compliance cost, risk exposure, and the capital incentive structures that drive organizations toward or away from structural oversight, were reserved for this paper.

This paper delivers that accounting.

1. The Unresolved Question

The augmentation case has not weakened since Paper 1 was published. A Harvard Business School field experiment at Procter and Gamble documented that AI-augmented teams produced breakthrough ideas at three times the rate of teams without AI assistance, with the largest gains accruing to lower-performing workers. [19] A Stanford Graduate School of Business study found that selective AI recommendations, offered only when the human was likely uncertain or incorrect, produced more accurate decisions than either full automation or unassisted human judgment. [7] A 2026 study from Oxford and Stanford, published in Harvard Business Review, confirmed what the earlier data suggested: while automation strategies show early gains relative to augmentation, long-term success is determined by human engagement with new tools, not by labor elimination. [26]

The evidence points consistently in one direction. The capital allocation points in another.

In 2025, AI captured 61 percent of all global venture capital, totaling $258.7 billion of $427.1 billion in total VC investment. [27] More than 90 percent of generative AI funding went to mega-rounds above $250 million. The five largest U.S. cloud and AI infrastructure providers collectively committed between $660 billion and $690 billion in capital expenditure for 2026, with major tech company capex more than doubling in two years to $427 billion by 2025, according to RBC Wealth Management and Futurum Group analysis. [60] That capital is going into data centers, GPU clusters, and networking infrastructure. It is not going into human oversight architecture. Bank of America's Q4 2025 earnings call documented the approach directly: the company's headcount was flat and expected to drift downward as AI handled work previously performed by humans. [61] Meta eliminated approximately 25,000 positions since 2022 while setting 2026 capital expenditure guidance at $115 to $135 billion, nearly double its 2025 outlay, a reallocation pattern explicitly

documented in the company's investor materials. [62] A March 2026 survey of 866 U.S. business leaders conducted by ResumeBuilder.com found that 54 percent had reduced or planned to reduce employee compensation in 2026 specifically to free capital for AI infrastructure spending. [28]

The gap between what augmentation evidence recommends and what capital allocation produces is not a failure of information. Executives and investors read the same research. The gap exists because the incentive architecture that governs capital decisions systematically makes automation rational even when augmentation outperforms it. Understanding that architecture is the prerequisite to changing it.

Before proceeding, the argument requires explicit scoping. This paper does not claim that automation is undesirable in all contexts, nor that augmentation universally outperforms replacement. The evidence for augmentation is strongest, and the governance requirements most consequential, in regulated, trust-dependent, and high-consequence environments where human judgment, contextual reasoning, and accountability for outcomes are load-bearing. In low-consequence, high-repeatability environments with reversible outputs and well-defined parameters (high-frequency trading execution, logistics routing, transactional document classification), automation-first architectures can produce superior efficiency outcomes without materially increasing governance exposure. The argument throughout this paper is correspondingly narrow: in the environments where institutional deployment is most consequential, and where the three-paper framework this series has built is specifically designed to apply, current capital incentive structures appear systematically misaligned with the governance architecture required for durable operation.

---

2. Three Capital Incentive Structures Working Against the Human Layer

2.1 Venture Capital and the Automation Premium

Venture capital pricing in AI rewards narratives of total automation for a structural reason. Software at zero marginal labor cost is more valuable than software with human-in-the-loop requirements, because the unit economics of the former scale without the linear cost growth of the latter. A "fully autonomous" roadmap projects margins that improve as volume increases. An augmentation roadmap projects margins that require the human layer to scale alongside the AI capability. The investor question is not which architecture performs better in regulated environments. The investor question is which architecture produces better returns at scale with predictable cost curves.

This framing is not irrational given the incentives. It is irrational given the full cost picture, which the current accounting framework does not require investors to price in. Gartner's April 2026 survey of 782 infrastructure and operations leaders found that only 28 percent of AI use cases fully succeed and meet ROI expectations, with 20 percent failing outright and 57 percent of leaders reporting at least one failure in their own organizations. [63] RAND Corporation's 2025 meta-analysis of enterprise AI initiatives found that 80.3 percent failed to deliver their intended business value: a third were abandoned before reaching production and nearly a fifth completed deployment without recouping their costs. [64] Neither finding has slowed capital deployment. Capital continues to flow because the automation narrative is priced into VC fund returns

at the portfolio level, where individual project failures are diversified away. The governance failures, liability exposures, and trust erosions that accumulate in the organizations deploying that capital are not diversified away. They concentrate.

The feedback loop this produces: VC rewards automation narratives. Founders pitch automation to attract VC. The pricing signal appears structurally oriented toward architectures that accumulate compliance liability at institutional scale, timed to materialize as regulatory enforcement matures.

2.2 Vendor Pricing and the Structural Incentive Against Human Seats

This is the most underexplored governance argument in the current literature, and it deserves precise treatment.

For two decades, enterprise software revenue scaled with human headcount. Every person using software was a revenue unit. Seat growth tracked hiring growth. The SaaS model was, in structural terms, a financial bet that human labor would remain the primary unit of productive work. The more humans an organization employed to do a function, the more software licenses it purchased. Vendors and customers had aligned incentives: organizational growth produced software revenue growth.

The shift now underway is documented in primary corporate disclosures. Salesforce's Q4 FY2026 SEC filing introduced "Agentic Work Units" as a new product metric, defined as discrete tasks executed by AI agents, measuring work completed without human involvement, not users with access. [52] The company reported 2.4 billion AWUs delivered, growing 57 percent quarter-over-quarter, and positioned AI agents explicitly as "digital labor." In its Q2 FY2026 earnings call, Salesforce management referenced a 40 percent reduction in support headcount achieved through Agentforce deployment as evidence of the platform's value proposition. [53] The pricing signal is legible: the vendor is measuring its success partly by how much human labor its product displaces.

Zendesk's Chief Revenue Officer, speaking on record in September 2025, stated that AI is "pressuring enterprises to rethink traditional seat-based pricing in favour of usage- and outcome-driven models," and that "smaller teams can manage workloads that would previously have required more staff." [54] Zendesk's dynamic pricing plan explicitly allows customers to shift committed budget between human agent seats and AI-powered "automated resolutions," with the CEO projecting that "100 percent of interactions will involve AI." [55] Workday's own 10-Q filing, submitted to the SEC in 2025, lists "pricing pressures" from AI capabilities reducing the number of human users needed as an explicit business risk factor. [56] These are primary corporate disclosures acknowledging that AI is restructuring the seat-based model on which enterprise software revenue has rested for twenty years.

The governance implication has not been stated with sufficient clarity in any regulatory or standards document. The per-seat-to-outcome shift is structurally aligned with reducing human involvement in operational workflows, because that reduction is what generates the measurable outcome the new pricing model rewards. An organization building on AI infrastructure priced per resolved interaction faces financial incentives to reduce human review steps, shorten decision gate engagement times, and minimize escalation

frequency, because each of those human-layer functions adds cost to the outcome metric. The architecture that Paper 2 specifies as necessary for accountable AI in regulated environments is the architecture that outcome-based pricing structurally penalizes.

The governance architecture and the emerging pricing architecture are therefore in structural tension. Both cannot simultaneously be optimized: governance architecture maximizes the quality of human involvement at consequential decision points; outcome-based pricing architectures reward minimizing that involvement. This tension is not yet widely recognized in AI governance literature, but it is visible in the primary corporate disclosures of the vendors building the infrastructure on which regulated AI systems run.

2.3 Executive and Board Incentive Structures and the Temporal Liability Gap

Quarterly reporting rewards visible, immediate cost reduction. Headcount is the most visible cost line in an income statement. The Human Layer is not a cost line. It is an architectural property of the system, present or absent, measured or unmeasured, but not represented in the standard financial reporting that drives executive performance evaluation and board-level mandates.

This creates what this paper calls the temporal liability gap: the benefits of removing human oversight accrue immediately and visibly, while the costs accrue over time, across reporting periods, and often under different executive leadership than the one that made the removal decision. The executive who eliminates the human review step in Q3 2025 books the headcount reduction in Q4 2025. The compliance failure, the regulatory enforcement action, or the liability lawsuit appears in 2027 or 2028, on someone else's watch, in a line item called "legal and regulatory" rather than "AI governance failure."

The evidence for this mechanism is now substantial, drawn from multiple independent sources.

At the aggregate level, outplacement firm Challenger, Gray and Christmas, the recognized authority on U.S. job cut tracking, documented that in 2025, companies cited AI in 55,000 layoff announcements, twelve times the number attributed to AI just two years earlier. [51] By March 2026, AI had become the leading stated reason for job cuts in a single month, accounting for 25 percent of all planned cuts. [47] Since 2023, when AI-cited job cuts were first tracked, the cumulative total has reached nearly 100,000 announced positions. [47] Andy Challenger, chief revenue officer of the firm, was explicit about the mechanism: "Companies are shifting budgets toward AI investments at the expense of jobs." [47]

That mechanism is confirmed by corporate disclosure. The pattern is consistent across industries, geographies, and financial conditions. Meta reported $201 billion in revenue for 2025 and eliminated approximately 25,000 positions since 2022, simultaneously setting 2026 capital expenditure guidance at $115 to $135 billion, nearly double its 2025 outlay. [62] The workforce reduction was not driven by financial distress. It was driven by capital reallocation. Dell cut approximately 11,000 employees in fiscal 2026, incurring $569 million in severance costs, while its AI server business grew 40 percent in the same period. [65] Atlassian cut 1,600 employees, 10 percent of its workforce, explicitly citing the need to fund AI initiatives. [65] Block's Jack Dorsey, in eliminating 40 percent of the company's workforce, wrote that the reductions were not driven by business failure. [66] Pinterest cut 15 percent of its human workforce in

January 2026 and redirected the savings to AI initiatives. [67] Cisco's CFO, when asked about workforce reductions in 2024, described them as "reallocating versus being in pursuit of cost savings," explicitly framing the human cost as an input to AI capital formation. [68]

Bank of America's approach illustrates the quieter version of the same dynamic. The bank's CFO described AI as saving "about 2,000 people" who would otherwise write code, and characterized the headcount strategy as: "We can just make decisions not to hire and let the headcount drift down." [61] No announcement. No restructuring charge. The Human Layer erodes through attrition rather than elimination, with the same architectural consequence and without the reputational cost of a publicized layoff.

The Bloomberg Intelligence C-Suite AI Survey, conducted across 604 senior executives in nine major sectors in late 2025, documented the structural mismatch directly. [48] Sixty-six percent of respondents reported making job cuts over the preceding twelve months as a result of AI deployment. More than 60 percent simultaneously expected headcount to increase on a three-year view. Separately, a survey of more than 1,000 executives found that 60 percent had made headcount reductions in anticipation of AI efficiencies, while only 2 percent reported large reductions as a result of actual AI implementation. [49] The pattern is consistent: capital reallocation decisions are being made before the evidence arrives, against a long-term expectation that remains unverified, on a compensation structure where financial measures govern 75 percent of executive incentive plans and short-term incentives reward annual performance rather than multi-year governance outcomes. [50] The cost reduction is visible in the current reporting period. The governance failure is not.

This is not a failure of management intelligence. It is a rational response to how performance is measured. Gartner documented in January 2026 that executives face board-level mandates to deliver cost savings through headcount reduction, and that AI is frequently positioned as the mechanism to make those reductions painless. [16] The Gartner analysis also found that AI productivity gains are rarely sufficient to enable frictionless reductions without operational risk, and that through 2028, AI investments are more likely to produce net headcount increases than decreases in knowledge worker roles, as new roles emerge alongside eliminated ones. [16] That finding has not altered the board mandate. The mandate operates on a different timeline than the evidence.

The result is an incentive structure in which individually rational capital decisions, made by executives responding appropriately to their performance environments, appear structurally oriented toward a collectively suboptimal outcome: governance-thin AI infrastructure deployed at institutional scale in regulated markets, positioned to generate compliance exposure precisely as enforcement matures.

---

3. Organizational Automation Bias

Paper 3 documented individual automation bias at length: the systematic tendency of humans to defer to automated recommendations even when override options exist, even when they have evidence the system is wrong. [5] The phenomenon is robust across 35 studies involving 19,774 participants. [5] It operates regardless of the human's experience level, and it intensifies under the conditions of high workload, time pressure, and cognitive load where it matters most.

This paper introduces organizational automation bias as a distinct phenomenon operating at the capital allocation level. An organization, as a collective decision-making entity, exhibits systematic tendencies that parallel individual automation bias. The option to preserve human oversight exists. The incentive structure makes exercising that option appear irrational. And by the time the system fails, the humans nominally responsible for oversight are in a moral crumple zone, positioned as accountable but structurally unable to have controlled the outcome. [17]

Organizational automation bias requires formal delimitation from adjacent concepts with which it might be confused. It is not ordinary short-termism, which is a general preference for near-term returns over long-term value that applies across investment categories. It is not the principal-agent problem, which describes the misalignment between an agent's interests and the principal's interests in any delegated decision. It is not quarterly capitalism or managerialism, which are structural critiques of how public markets incentivize short reporting cycles over long-term value creation. These are real phenomena and organizational automation bias intersects with all of them. The distinction is specific. Organizational automation bias is the systematic institutional mispricing of governance degradation as a cost category: not merely discounting the future, but failing to represent governance failures in the cost accounting at all. The three conditions that make it distinct are: first, the benefits of removing human oversight are directly attributable to specific decisions made by identifiable actors in a defined reporting period; second, the costs of that removal are not merely deferred but structurally fragmented across timelines, departments, reporting entities, and executive tenures such that no single actor's performance evaluation registers them; third, when failures occur, the accountability attribution defaults to technical failure rather than governance failure, which suppresses the organizational learning that would otherwise correct the bias. A company that discounts long-term returns is exhibiting short-termism. A company that systematically does not account for governance degradation as a cost because that cost is unrepresented in the measurement architecture is exhibiting organizational automation bias. The distinction matters because the remedies differ: short-termism is addressed by changing time horizons in performance measurement; organizational automation bias is addressed by making governance degradation visible as a cost, which is precisely what the Human Layer Score and its compliance trace are designed to do.

The mechanism operates through four channels.

The first is anchoring to headcount as the primary ROI variable. AI return on investment is measured against labor costs avoided. The denominator of the ROI calculation is the cost of human labor that the AI replaces. The numerator is the value the AI produces. Compliance exposure, liability accumulation, trust erosion, and market access restriction are not in the denominator. They are not counted as costs of automation. They are counted, if counted at all, as separate line items that accrue later. An ROI calculation that excludes the deferred costs of removing the human layer will consistently overstate the return to automation.

The second is framing augmentation as cost and automation as efficiency. The language of capital allocation encodes the bias. Automation is described as efficiency gain: costs fall, throughput rises, margins improve. Augmentation is described as residual cost: the human layer that remains after the AI has been deployed. The framing obscures the correct comparison, which is automation plus deferred compliance liability versus augmentation plus structural trust value. A system designed to amplify human judgment in a Tier 3 regulated

environment generates verifiable oversight, institutional market access, and liability defense that a system designed to eliminate human involvement does not. None of those properties appear in the efficiency framing.

The third is discounting deferred costs at the wrong rate. Standard financial analysis discounts future liabilities at the cost of capital. But regulatory and reputational failures are not normally distributed risks that smooth out over time. They are tail events: sudden, concentrated, and reputationally compounding. A discounted cash flow model applied to AI compliance risk systematically underweights these tails. The $11.3 million average cost of a failed AI project in financial services, documented in 2025, does not represent the distribution of outcomes; it represents the central case. [34] The EU AI Act fine structure, at up to 7 percent of global annual turnover, represents the tail. Standard discounting treats those tails as negligible. Actuarial experience with low-probability, high-consequence events suggests they are not.

The fourth is misattributing failure. When AI systems produce consequential errors in regulated environments, the failure attribution defaults to "the model" or "the data." The governance architecture is not identified as the missing variable. Health insurer AI systems operating in Medicare Advantage generated an 82 percent overturn rate on appealed denials, documented in the Health Affairs study cited above. [57] The post-mortems on these failures focused on model accuracy and data quality. The absence of structural human oversight at the decision gate was not the primary analytical frame, which means the organizational learning from those failures did not produce governance architecture improvements. It produced model retraining. The root cause remained.

Organizational automation bias differs from individual automation bias in one critical respect. Individual automation bias is correctable through architectural intervention: cognitive forcing functions, decision gates, verification sampling, and trust calibration interfaces, as specified in Paper 2. Organizational automation bias is not corrected by system architecture. It is corrected by changing the incentive structure. The three forces documented in the next section describe how that change is arriving.

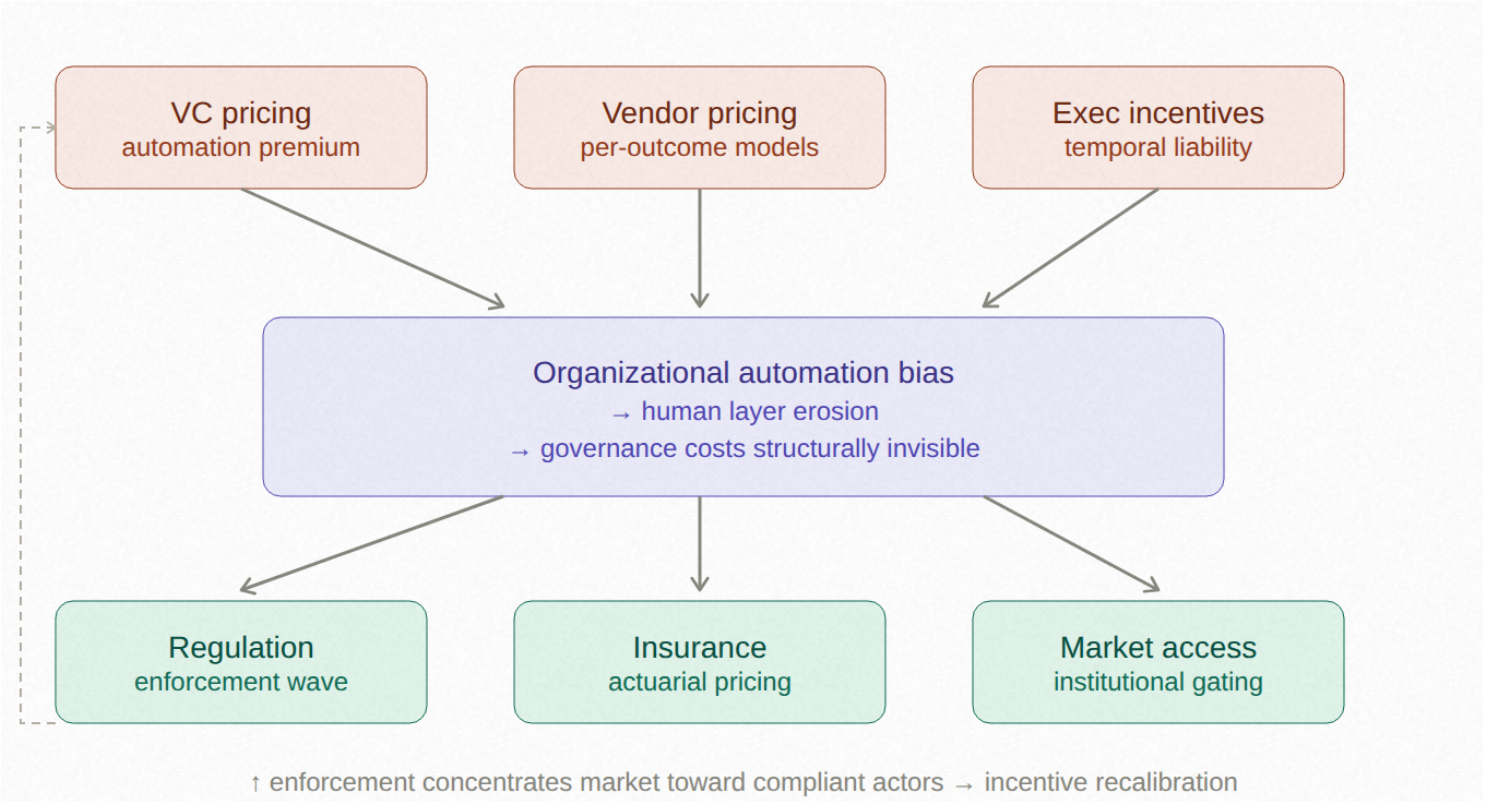

Figure 1. The Organizational Automation Bias Loop. Three capital incentive structures (VC pricing, vendor pricing, and executive incentives) drive organizational automation bias through human layer erosion and the structural invisibility of governance costs in standard accounting frameworks. Regulatory enforcement, insurance pricing, and institutional market access gating collapse the temporal liability gap that enables this pattern. The dashed path indicates how enforcement pressure concentrates the market toward compliant actors, reshaping incentives for remaining organizations.

---

4. Three Forces Collapsing the Temporal Liability Gap

The temporal liability gap that enables organizational automation bias depends on the deferred costs of removing the human layer remaining invisible to capital markets, regulators, and institutional counterparties during the period when the automation decision is made. All three of those conditions are changing.

4.1 Regulatory Enforcement: From Theoretical to Actualized

The EU AI Act entered into force on August 1, 2024. The prohibition on unacceptable-risk AI practices became enforceable on February 2, 2025. Governance obligations for general-purpose AI took effect on August 2, 2025. The most material enforcement deadline for organizations deploying AI in high-risk categories, including employment, credit decisions, healthcare, and law enforcement, is August 2, 2026. [3] That deadline is three months away as this paper publishes.

The fine structure is not a nuisance cost. Violations involving prohibited AI practices carry penalties of up to 35 million euros or 7 percent of global annual turnover, whichever is higher. For organizations with material AI exposure in regulated markets, 7 percent of global annual turnover is an existential number. The EU's Digital Omnibus proposal of late 2025 introduced the possibility of a delay to December 2027 for Annex III high-risk systems. Organizations should not build compliance plans on that possibility. The prudent deadline remains August 2026, and the enforcement regime it activates is real.

The temporal liability gap operates in the regulatory dimension only so long as enforcement remains theoretical. Across multiple regulatory transitions, from GDPR to securities regulation reform, research has consistently found that improvements in supervisory regime credibility produce capital-market effects before the first enforcement action takes place. [41] The market prices enforcement probability, not enforcement actuality. The EU AI Act's enforcement infrastructure, including national competent authorities designated by August 2025 and the European AI Board, is operational. The enforcement probability is no longer speculative.

In the United States, the absence of a federal AI framework has not produced a regulatory vacuum. State-level AI legislation is expanding rapidly: Colorado's SB 205 took effect in February 2026, New York City's Local Law 144 is in full enforcement, and more than half of U.S. states have introduced or passed AI legislation. [37] Organizations that have built structural human layers compliant with the EU AI Act's human oversight requirements have, by construction, built to a standard that satisfies the most demanding state-level requirements in the U.S. The regulatory architecture is converging, and the convergence endpoint already exists in the EU AI Act.

4.2 Insurance and Liability Pricing: The Actuarial Human Layer

The second force closing the temporal liability gap is actuarial. Insurance and liability pricing are beginning to incorporate AI risk explicitly, which means the costs that organizations have treated as deferred and theoretical are beginning to appear as present and priced.

The health insurance sector provides the clearest early evidence. AI denial systems operating in Medicare Advantage generated an 82 percent overturn rate on appeal, documented in a January 2026 Health Affairs study by Mello et al. that examined AI use in prior authorization and claims processes, and prompted litigation that federal courts allowed to advance into discovery. [57] A Congressional report found that AI-driven denial rates at UnitedHealthcare, CVS, and Humana's Medicare Advantage plans increased significantly following expanded algorithmic deployment. [58] The NAIC's 2025 Health AI/ML Survey Report, completed by the Big Data and Artificial Intelligence Working Group across sixteen states, found that nearly one-third of health insurers do not regularly test their AI models for bias or discrimination, a documented governance gap already in the regulatory queue. [59]

The actuarial implication is direct. Insurers offering AI liability coverage cannot price that coverage without understanding the human oversight architecture of the insured organization. An organization with a Decorative human layer, scoring in the 5 to 9 band on the Human Layer Score, has documented evidence of nominal oversight without enforcement. That evidence, producible through the compliance trace Paper 3

requires, is simultaneously the audit evidence a regulator examines and the underwriting input an insurer evaluates. The absence of a structural persistence layer, the absence of logged decision deltas, and the absence of identity-anchored accountability records are not invisible to an insurer's diligence process. They are legible as elevated tail risk.

The trajectory follows the pattern established in cybersecurity. Cyber liability insurance became a mainstream product category in the late 2010s. By the early 2020s, insurers were requiring evidence of specific security controls (multi-factor authentication, endpoint protection, and incident response plans) as conditions of coverage. Organizations without those controls faced either premium surcharges or coverage exclusions that exceeded the cost of implementing the controls themselves. AI liability insurance is at the beginning of the same trajectory. The Human Layer Score is positioned to function as the AI governance equivalent of a SOC 2 Type II report: an independently verified assessment of governance architecture that gives an insurer a basis for pricing rather than a basis for declining.

4.3 Institutional Market Access: The Premium Market Is Gated

The third force is market structure. Tier 3 institutional markets are not equally accessible to all AI system architectures. They are accessible to systems with structural human oversight and not accessible to systems with nominal oversight, regardless of technical performance metrics.

This is already operational. A Mordor Intelligence analysis of the EU AI Act governance market in October 2025 found that 90 percent of European financial institutions use AI but only 9 percent feel equipped to meet the AI Act's standards. [43] That preparedness gap represents the current market structure: the vast majority of AI deployments in regulated European markets are operating without the governance architecture the market will require within months. The firms that close that gap become the credible counterparties. The firms that do not become the counterparty risk their institutional clients are required to manage.

Enterprise procurement is driving this ahead of regulation. Enterprise AI buyers in regulated markets are already requiring counterparties to demonstrate governance architecture. The question is moving from "does this system have human oversight?" toward "what is the Human Layer Score, and does it meet the tier minimum for our procurement standard?" The premium market, the market that commands institutional margins in regulated industries, is concentrating around the governance distribution. Organizations scoring in the Structural and Institutional bands are increasingly positioned to access Tier 2 and Tier 3 markets. Organizations scoring in the Decorative or Operational bands may face growing barriers to those markets, regardless of their AI capability. The 3x augmentation multiplier documented in Paper 1 is, in part, a premium market access effect: the markets where augmentation outperforms automation most consistently are precisely the markets that require structural human oversight as a condition of operation.

---

5. The Human Layer Score as a Capital Signal

Paper 3 introduced the Human Layer Score as a governance instrument. This section argues, with reference to empirical evidence from prior regulatory transitions, that it may also function as a capital markets signal for organizations operating in regulated, trust-dependent environments, under conditions where AI governance verification becomes institutionally standardized, as analogous governance standards have in adjacent domains. This claim requires precision, and it requires honest acknowledgment of what the evidence does and does not yet demonstrate.

5.1 What the Empirical Record Shows About Governance and Capital Markets

Three prior regulatory transitions provide relevant evidence, and the evidence is more specific than a general claim that compliance creates value.

Sarbanes-Oxley, enacted in 2002 following the Enron and WorldCom scandals, provides the first anchor point. Research using SOX as a quasi-natural experiment in corporate governance found that firms more compliant with SOX provisions prior to enactment received more positive stock-price reactions when the legislation passed. [36] The market rewarded existing governance quality when the enforcement regime became credible. SOX also produced larger improvements in operational efficiency in concentrated industries than in non-concentrated ones, with the gains most pronounced among firms with weaker governance prior to enactment. [36] This is the honest version of the precedent: pre-compliance is advantageous for firms with institutional market exposure; for smaller actors without concentrated market position, the compliance burden can be a net negative. The governance standard concentrated the market toward actors who could absorb the governance infrastructure requirement.

The GDPR provides the second anchor. Research examining the GDPR's firm-level effects documents that enforcement concentrating market access toward already-compliant vendors is a documented outcome of credible regulatory enforcement. [54] After GDPR enforcement became credible, the market for technology vendors serving enterprise customers became more concentrated: larger, already-compliant vendors captured the business of clients who needed to reduce third-party data risk, while non-compliant or smaller actors faced market exit pressure. [55] The GDPR also had measurable negative effects on compliance-cost-sensitive firms' profits, which establishes that the costs are real and should not be dismissed. Non-compliance before GDPR enforcement was a viable strategy precisely because enforcement was low. That viability collapsed as enforcement became credible. [56]

The third body of evidence concerns the general mechanism linking supervisory regime quality to capital-market effects. Research on EU securities regulation reform found that "improvements of countries' supervisory regimes have immediate capital-market effects even before the first enforcement action took place." [41] The market prices in governance quality in anticipation of enforcement. The implication for AI governance is direct: institutional investors are pricing AI governance quality now, before August 2026, because they understand the enforcement landscape. A 2025 survey of European financial executives found

that 90 percent of institutions use AI, but only 9 percent feel equipped to meet the AI Act's standards. [43] That preparedness gap has a capital-market expression, and it is not symmetric: the prepared organizations price in the governance premium; the unprepared organizations accumulate the liability tail.

ISO 27001 provides the fourth precedent, at the level of certification as market access mechanism. A long-term event study of 143 ISO 27001-certified firms against matched non-certified controls found improved profitability, labor productivity, and sales performance in certified organizations. [42] More operationally: ISO 27001 moved from a voluntary standard to a non-negotiable prerequisite for institutional contracts in finance, healthcare, and government procurement. The firms that certified during the voluntary window acquired market access that, once certification became a prerequisite, was the cost of entry rather than a differentiator. The window during which a governance standard functions as a differentiator is the period in which it is acquired most advantageously.

5.2 What the Human Layer Score Is and Is Not Claiming

The Human Layer Score as a capital signal is not a blanket claim that building governance infrastructure generates returns exceeding its cost for all organizations in all contexts. The evidence from GDPR and SOX is clear that compliance costs are real, that they fall unevenly across firm sizes, and that the returns to governance investment are concentrated among firms with institutional market exposure, regulated industry presence, and Tier 2 or Tier 3 deployment contexts.

The precise claim is narrower and more defensible: for organizations operating in regulated, trust-dependent, or high-stakes environments where the Human Layer framework defines Tier 2 and Tier 3 deployment, the same environments where the augmentation multiplier is most consistently documented, structural human layer investment at this moment carries the profile of governance infrastructure acquired in the differentiator window, before it becomes an enforcement-mandated prerequisite.

Three specific mechanisms support this claim.

The first is market access gating. Tier 3 institutional markets require demonstrable human oversight as a condition of access. This is currently operational, not projected. Organizations with Decorative human layers cannot access these markets regardless of their AI performance metrics. The Human Layer Score functions as a market access credential for the premium market. The credential acquired before the August 2026 enforcement deadline is acquired at lower cost than the credential acquired under remediation pressure after a formal compliance deficiency.

The second is liability exposure differentiation. A structural human layer, documented through the compliance trace Paper 3 requires, provides an affirmative defense that a decorative or nominal layer does not. The six-field minimum audit trail, the identity-anchored authorization records, and the delta logs distinguishing AI recommendation from human decision are simultaneously the evidence that satisfies a regulator and the documentation that limits an organization's liability in the event of a challenged outcome. For organizations with material AI exposure in regulated markets, this is a quantifiable reduction in tail risk, not a soft governance benefit.

The third is due diligence signal value. An investor or institutional counterparty evaluating an AI company operating in a Tier 2 or Tier 3 market currently has no standardized instrument for answering the question: is the human layer structural or decorative? The Human Layer Score is designed to answer this with a number, a component breakdown, and a tier compliance determination. A self-assessed score is a working document. An independently verified score is positioned to function as a diligence artifact, the AI governance equivalent of a SOC 2 Type II report or an ISO 27001 certificate, if and as institutional demand for AI governance verification develops along the trajectory documented in Sections 4.1 through 4.3. The AI governance market is growing at 28.8 percent annually, from $340 million in 2025 to a projected $1.21 billion by 2030. [43] That growth reflects institutional demand for verifiable governance signals. Whether the Human Layer Score becomes a standardized instrument within that market depends on adoption by auditors, regulators, and procurement functions, a process that is underway but not yet complete.

A legitimate question for any scoring framework is whether organizations can satisfy its formal requirements without corresponding operational integrity. The Human Layer framework reduces this risk through several design properties that make surface compliance progressively harder to sustain. Floor-based scoring prevents high performance on one component from masking weakness in another: an organization with Level 4 trust calibration and Level 1 override mechanisms is assessed at Level 1, not the average. Verification sampling inserts known-answer cases with intentionally incorrect AI output to test whether human oversight is genuinely engaged, not merely present at the interface. Identity anchoring requires pre-assigned accountability before output is produced, which distinguishes documented authorization from post-hoc blame distribution. The compliance trace's delta logging distinguishes active human judgment from passive approval by recording the difference between the AI recommendation and the human's final decision. No single design property prevents all forms of nominal compliance, and independent audit, which Paper 3 requires for Tier 2 and Tier 3 systems making regulatory submissions, is the enforcement mechanism that makes sustained simulation costly. The framework's architecture creates incentives for genuine implementation, but it does not make gaming impossible; it makes gaming structurally more expensive than building the governance function genuinely.

5.3 The Honest Limit of This Argument

The pre-enforcement positioning argument does not apply uniformly. For Tier 1 advisory systems operating in low-consequence, reversible output contexts, the enforcement cost is lower, the market access premium is lower, and the investment calculus for a structural human layer is governed primarily by operational performance considerations rather than by the capital signal mechanisms described here. Paper 3's tier-specific minimums reflect this: Tier 1 systems face a lower required score precisely because the market access and liability exposure dynamics operate at lower magnitude in those contexts.

The argument presented in this section is specific to Tier 2 and Tier 3 deployments in regulated, trust-dependent environments. That specificity is not a weakness. It reflects where the Human Layer matters most architecturally, where the evidence for augmentation over automation is most consistent, and where the capital-market consequences of getting it wrong are most material.

5.4 Limitations

Four limitations bound the claims made in this section and in the series as a whole.

First, the Human Layer Score has not yet been externally validated through independent application by qualified auditors at scale. Inter-rater reliability across auditors operating in different regulatory jurisdictions and industry contexts remains untested. The self-assessment instrument in Paper 3 is a diagnostic tool; independent audit is required for regulatory submissions, but the auditor qualification standards the framework defines have not yet been tested against a body of practice.

Second, the capital signal argument draws on analogies from prior regulatory transitions (SOX, GDPR, ISO 27001) whose conditions may not replicate precisely in AI governance. The pace of enforcement, the consistency of regulatory interpretation across jurisdictions, and the speed at which procurement practices incorporate governance standards are all variables that the analogy cannot fully account for.

Third, the sector applicability of the tier classification and scoring requirements is established through regulatory framework analysis and operational precedent rather than longitudinal empirical study. The framework's prescriptive authority in specific industry contexts would be strengthened by deployment case studies and implementation analysis that this series does not yet include.

Fourth, the organizational automation bias concept, while formally delimited in Section 3, has not been operationally measured. Its four channels are derived from the convergence of documented capital allocation behavior, executive compensation structure, and corporate disclosure, rather than from a controlled study of organizational decision-making. It functions as an explanatory framework, not a measured construct.

These limitations are normal for a first-publication conceptual framework. They define the next phase of work, not a deficiency in the current one.

For boards with material AI deployment in regulated or high-consequence environments, this paper carries a specific governance implication. AI oversight can no longer be treated as a compliance function delegated entirely to legal or technical teams. Where AI systems operate in the contexts this framework addresses, including regulated finance, healthcare, critical infrastructure, and other Tier 2 and Tier 3 environments, governance architecture is simultaneously a capital allocation question, a liability management question, and a market access question. The Human Layer Score provides a board-level metric that standard financial reporting does not: whether the organization's AI systems carry structural human oversight or decorative oversight, and whether that distinction is independently verifiable. Boards that surface this question before the enforcement window closes are positioned differently from those that encounter it first as a remediation task.

---

6. What the Series Has Built

This series set out to make a case that most AI governance frameworks leave unmade: that the human layer in an AI system is not a philosophical commitment or a regulatory checkbox, but a structural component with measurable design properties, enforceable requirements, and capital-market consequences.

Papers 1 through 3 built the foundation. The economic case: augmentation outperforms automation, documented across 1,500 organizations and confirmed in post-GPT field experiments, with the performance differential concentrated precisely in the regulated, trust-dependent environments where human judgment is irreplaceable. The architecture: five components forming a dependency graph, each independently required by major regulatory frameworks, none optional without compromising the guarantees of the others. The measurement instrument: a scoring methodology that prevents averaging away weaknesses, enforces tier-specific minimums, and produces the compliance trace that serves simultaneously as operational telemetry, audit evidence, and liability defense.

This paper answered the question those three raised without resolving. If the evidence for augmentation is this consistent, why does capital continue to flow toward replacement? The answer is organizational automation bias: a systematic institutional tendency to overweight automation benefits that are immediately visible and directly attributable, while underweighting governance costs that are deferred, fragmented across reporting periods and accountability domains, and critically not represented in the standard cost accounting that governs executive performance evaluation. It is distinct from short-termism and principal-agent problems because the mechanism is not merely discounting the future. It is the structural invisibility of governance degradation as a cost category, which suppresses the organizational learning that would otherwise correct it.

Three capital incentive structures sustain that invisibility: venture capital pricing that rewards headcount-free automation narratives, vendor pricing architectures that are structurally aligned with minimizing human involvement in workflows, and executive performance measurement calibrated to timelines shorter than the governance failures that accumulate when human oversight is removed. Three forces are now collapsing it: regulatory enforcement that is actualized rather than theoretical, insurance and liability pricing that is beginning to incorporate AI governance as an underwriting variable, and institutional market access that is already functioning as a gate rather than an aspiration.

The empirical precedent from SOX, GDPR, and ISO 27001 governance transitions suggests that organizations operating in regulated markets may benefit from building structural human layers before enforcement converts governance architecture from differentiator to prerequisite; though whether the Human Layer Score itself becomes a standardized institutional signal depends on adoption processes that are underway but not yet complete.

The architectural decisions made in this period will determine whether organizations enter regulated markets with structural oversight already in place, or under remediation pressure after enforcement has concentrated the market around those that built it first.

References

[1] Sheridan, T.B. and Verplank, W.L., "Human and Computer Control of Undersea Teleoperators," MIT Man-Machine Systems Laboratory, 1978.

[2] Parasuraman, R., Sheridan, T.B., and Wickens, C.D., "A Model for Types and Levels of Human Interaction with Automation," IEEE Transactions on Systems, Man, and Cybernetics, Part A, 30(3), 286-297, 2000.

[3] European Union, Regulation (EU) 2024/1689, "Artificial Intelligence Act," Articles 14, 50. Entered into force August 1, 2024. Full applicability for high-risk systems August 2, 2026.

[4] Angelova, V., Dobbie, W., and Yang, C., "Algorithmic Recommendations and Human Discretion," NBER Working Paper No. 31747.

[5] Romeo, G. and Conti, D., "Exploring Automation Bias in Human-AI Collaboration: A Review and Implications for Explainable AI," AI and Society, Springer Nature, 2025. Systematic review of 35 studies (19,774 participants).

[6] Google DeepMind, retinal disease AI screening system deployed in NHS UK. Three-tier triage protocol.

[7] McLaughlin, B. and Spiess, J., "Complementary Algorithms," Stanford Graduate School of Business, 2025.

[8] Kahn, L., Probasco, E.S., and Kinoshita, R., "AI Safety and Automation Bias," Center for Security and Emerging Technology, Georgetown University, November 2024.

[9] International Organization for Standardization, ISO/IEC 42001:2023, "Information Technology: Artificial Intelligence: Management System."

[10] Frontiers in Big Data, "On the Purpose of Meaningful Human Control of AI," December 2022.

[11] National Institute of Standards and Technology, "AI Risk Management Framework (AI RMF 1.0)," NIST AI 100-1, January 2023.

[12] Infocomm Media Development Authority (IMDA), Singapore, "Model AI Governance Framework for Agentic AI," 2025.

[13] Kupfer, C. et al., "Automation Bias in Public Administration: An Interdisciplinary Perspective from Law and Psychology," Government Information Quarterly, 2024.

[14] Carnat, I., "Human, All Too Human: Accounting for Automation Bias in Generative Large Language Models," SSRN, March 2025.

[16] Poitevin, H. and Suda, N., "Why Your Headcount Strategy Matters More Than AI Downsizing," Gartner, January 23, 2026.

[17] Elish, M.C., "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction," Engaging Science, Technology, and Society, 5, 40-60, 2019.

[18] Daugherty, Paul R. and Wilson, H. James, "Human + Machine: Reimagining Work in the Age of AI," Harvard Business Review Press, 2018 (updated and expanded edition, 2024; ISBN 978-1647827205).

[19] Dell'Acqua, F., Ayoubi, C., Lifshitz, H., Sadun, R., Mollick, E., et al., "The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise," Harvard Business School Working Paper No. 25-043, March 2025.

[20] Noureddine, A., "The Human Layer: Why the Most Critical Infrastructure in AI Isn't Artificial," Paper 1 in The Human Layer series, February 2026. DOI: 10.5281/zenodo.19119699.

[21] Noureddine, A., "The Human Layer Architecture: A Specification for Human-AI System Design," Paper 2 in The Human Layer series, March 2026. DOI: 10.5281/zenodo.19120077.

[22] Organisation for Economic Co-operation and Development, "OECD AI Principles," adopted May 2019, amended 2024.

[23] Institute of Electrical and Electronics Engineers, IEEE 7000-2021, "Standard Model Process for Addressing Ethical Concerns During System Design," September 2021.

[25] Noureddine, A., "The Human Layer Audit: Measuring Accountability in AI Systems," Paper 3 in The Human Layer series, March 2026. DOI: 10.5281/zenodo.19453026.

[26] De Neve, J-E., Hancock, J.T., and Niederhoffer, K., "Why Companies That Choose AI Augmentation Over Automation May Win in the Long Run," Harvard Business Review, April 15, 2026.

[27] Organisation for Economic Co-operation and Development, "Venture Capital Investments in Artificial Intelligence Through 2025," February 17, 2026.

[28] ResumeBuilder.com, Survey of 866 U.S. Business Leaders on AI Spending and Compensation, March 2026.

[29] Andreessen Horowitz, "AI Is Driving a Shift Towards Outcome-Based Pricing," Enterprise Newsletter, December 2024.

[31] Bessemer Venture Partners, "The AI Pricing and Monetization Playbook," February 2026.

[32] Intercom, Fin AI product pricing model. Per-resolution pricing at $0.99 per AI-resolved support ticket, as referenced in Bessemer Venture Partners [31].

[34] Pertama Partners, "AI Project Failure Statistics 2026," February 21, 2026.

[35] SQ Magazine, "AI Compliance Cost Statistics 2026," April 1, 2026.

[36] Chhaochharia, V., Grullon, G., Grinstein, Y., and Michaely, R., "Product Market Competition and Internal Governance: Evidence from the Sarbanes-Oxley Act," Management Science, 63(5), 1405-1424, 2017.

[37] Digital Applied, "AI Compliance for Small Business 2026: Bias Audit Guide," March 15, 2026.

[38] Frey, L., "Privacy Regulation and Firm Performance: Estimating the GDPR Effect Globally," Economic Inquiry, 2024.

[39] Johnson, G., Shriver, S., and Goldberg, S., referenced in: Goldberg, S. et al., NBER Working Paper No. 30705, "Economic Research on Privacy Regulation," 2022.

[40] Martin, K. et al., 2019, cited in NBER Working Paper No. 30705.

[41] Hail, L. et al., "Capital-Market Effects of Securities Regulation: Prior Conditions, Implementation, and Enforcement Revisited," NBER Working Paper No. 16737.

[42] Tsohou, A. et al., "Information Security and Value Creation: The Performance Implications of ISO/IEC 27001," Computers and Security, 2022.

[43] Mordor Intelligence, "EU AI Act Compliance: Strategic Advantage or Growth Barrier," October 2025.

[44] Allen, R.C., "Engels' Pause: Technical Change, Capital Accumulation, and Inequality in the British Industrial Revolution," Explorations in Economic History, Vol. 46, Issue 4, 2009.

[45] World Economic Forum, "Future of Jobs Report 2025."

[46] CBS News, "More Companies Are Pointing to AI as They Lay Off Employees," May 2026.

[47] Challenger, Gray and Christmas, "Challenger Report: March Cuts Rise 25% From February, AI Leads Reasons," April 2, 2026.

[48] Bloomberg Intelligence, "C-Suite AI Survey 2025," December 2025.

[49] Built In, "Did AI Take Your Job? The Truth About AI Washing," March 2, 2026.

[50] SullivanCotter, "2025 Executive Compensation Insights," February 2026.

[51] CNBC, "AI Job Cuts: Amazon, Microsoft and More Cite AI for 2025 Layoffs," December 21, 2025.

[52] Salesforce, Inc., Form 8-K (Q4 FY2026 Earnings Release), SEC filing. Available at SEC EDGAR.

[53] Salesforce, Inc., Q2 FY2026 Earnings Conference Call, September 3, 2025.

[54] Zendesk, Chris Donato (Chief Revenue Officer and President of Global Sales), interview with Frontier Enterprise, "Scaling Outcome-Based Pricing with AI," September 12, 2025.

[55] Zendesk, "AI Dynamic Pricing Plan" product documentation, 2026.

[56] Workday, Inc., Form 10-Q, filed with the SEC, 2025.

[57] Mello, M.M. et al., "The AI Arms Race In Health Insurance Utilization Review: Promises Of Efficiency And Risks Of Supercharged Flaws," Health Affairs, January 2026.

[58] Oberlander, J. et al., "Medicare Advantage Becoming a Disadvantage with Use of Artificial Intelligence in Prior Authorization Review," npj Digital Medicine (Nature Portfolio), February 2026.

[59] National Association of Insurance Commissioners (NAIC), Big Data and Artificial Intelligence (H) Working Group, "Health Artificial Intelligence (AI)/Machine Learning (ML) Survey Report," May 2025.

[60] RBC Wealth Management and Futurum Group, as cited in CBS News and Technocracy News, April 2026.

[61] Bank of America Corporation, Fourth Quarter and Full Year 2025 Financial Results Earnings Call, January 2026.

[62] Meta Platforms, Inc., Q4 2025 Earnings Release and Investor Call, January 2026.

[63] Gartner, Inc., "AI Projects in I&O Stall Ahead of Meaningful ROI Returns," Press Release, April 7, 2026.

[64] RAND Corporation, enterprise AI initiative meta-analysis, 2025, as reported in Pertama Partners, "AI Project Failure Statistics 2026," February 2026.

[65] Dell Technologies, Annual Report and Earnings Materials, Fiscal Year 2026. Atlassian Corporation, workforce reduction announcement, March 11, 2026.

[66] Block, Inc., workforce reduction, February 2026. CEO Jack Dorsey statement.

[67] Pinterest, Inc., workforce reduction announcement, January 2026.

[68] Cisco Systems, Inc., workforce reduction, 2024. CFO Scott Herren on record.

Ahmad Noureddine is Co-Founder & CEO at Timer, the organizational memory layer. 25+ years building systems that put humans at the center of technology.

This is Paper 4 in The Human Layer series, published at ahmad.pt/research.

Paper 1: The Human Layer: Why the Most Critical Infrastructure in AI Isn't Artificial Paper 2: The Human Layer Architecture: A Specification for Human-AI System Design Paper 3: The Human Layer Audit: Measuring Accountability in AI Systems