BOUNDARY BLINDNESS
UNDER ARTIFICIAL INTELLIGENCE

Early Findings and a Call for Industry Validation

Alexander D. Barrett

SENIOR FELLOW

VERIDA CHARTER FOUNDATION

A NONPROFIT RESEARCH FOUNDATION ON CROSS-FIRM COORDINATION IN REGULATED MULTI-PARTY MARKETS

JULY 2026

Boundary Blindness Under Artificial Intelligence

Early Cross-Industry Findings on the Missing Decision-Evidence Layer

Alexander D. Barrett, Senior Fellow, Verida Charter Foundation, a nonprofit research foundation on cross-firm coordination in regulated multi-party markets, 501(c)(6), Working paper, July 2026, drawn from a larger thesis in progress, *The Aetiology of Boundary Blindness*.

This paper consolidates the artificial-intelligence dimension of a twelve-month research program (October 2025 to October 2026) and generalizes it beyond the insurance market in which the program began. It draws on seventy-one semi-structured interviews with senior practitioners and on twenty-five independently published sources across seven sectors and five jurisdictions. It presents findings, not a product, and makes no commercial claim. Where it proposes an open standard, it means a specification governed as a public good and freely implementable by any party, not a proprietary product.

1. Summary of findings

The central finding is one sentence. Artificial intelligence does not create the gap this research has been tracing across regulated markets; it converts that gap from a chronic condition the market tolerated for decades into an acute one it cannot tolerate.

The gap is this. When a decision crosses a boundary between two organizations, two software systems, or two regulatory regimes, the data may cross, but the evidence of how the decision was reached does not. Who reviewed what, under what authority, against which rules, with what conclusion, and whether that conclusion still holds, stays behind. The receiving party gets the outcome, not the basis for relying on it. The research calls the failure to see this one cause behind many separate symptoms *"boundary blindness."*

Five findings follow.

1. Artificial intelligence is an amplifier, not a cause. The condition predates it; what AI removes is the slack, the human memory, that made it survivable.
2. Five mechanisms convert the chronic gap into an acute one: explainability that fails at the boundary, agentic volume that outruns human audit, models that do not hold still, media that remove the review step, and composite human-and-machine decisions whose reliance goes unrecorded.
3. The AI-governance regimes now in force largely certify the wrong thing. They attest to the management system rather than the decision, and they stop at the firm wall.
4. The condition is cross-industry, appearing in customs, logistics, banking, trade finance, and anti-money-laundering, and compounding as a transaction crosses more sectors.
5. The architecture that would close the gap is being independently reinvented, sector by sector, by parties who have never encountered this research.

The paper develops each finding, sets out the architecture and governance they imply (a coordination layer that belongs to no one, not another platform), and closes with the questions that would confirm or falsify the thesis.

2. What this paper is, and how the findings were gathered

The program began as an inquiry into the London insurance market, where a multi-party subscription structure exposes cross-firm coordination failures with unusual clarity. Over eight months, it conducted 71 interviews (51 in the insurance sectors of the United Kingdom and the United States, and 20 across adjacent regulated domains). It analyzed twenty-five published sources from regulators, reinsurers, trade bodies, and other sectors. The method was convergence analysis: reading independently published work from different sectors as a single pattern and testing whether the same structural condition recurs.

It does. The insurance case is developed elsewhere and not repeated here. This paper takes the artificial-intelligence dimension, which the program had treated as one of several forcing functions. It states it on its own terms because AI is where the condition stops being tolerable and where the diagnosis travels most cleanly across industries.

Three terms recur. ***Decision evidence*** is the record of what was decided and why: who reviewed what, under what authority, applying which evidence and rules, reaching which conclusion, as of when. ***Boundary blindness*** is the failure to recognize that many separate symptoms share a single cause, and that decision evidence does not cross boundaries. A ***portable decision record*** is the artifact that would carry that evidence across a boundary with its integrity intact and independently checkable, without moving the underlying data.

3. Finding one: artificial intelligence is an amplifier, not a cause

The amplifier framing is defensible and corroborated by insiders in the most exposed industries. A corporate operations executive in Munich Re's study of AI risk observed that these risks *"are all existing problems, maybe just exacerbated by artificial intelligence."* A Group Data Protection Officer interviewed for this research reached the same conclusion: the difficulty AI introduces is not a new category of harm but the acceleration and dispersal of an old one.

Because the gap is structural and predates the technology, AI functions as an accelerant that removes the slack. A market that could absorb boundary blindness when humans made decisions at human speed cannot absorb it when agents make decisions at machine speed. The human memory that quietly compensated is leaving the workforce. Agentic AI is, in that precise sense, the deadline.

The deadline is concrete. The Financial Conduct Authority's June 2026 horizon scan describes autonomous systems executing thousands of decisions in seconds, with the human positioned *"only before and after the loop."* The records that would later settle a dispute over what an agent did, as one industry catalog puts it, *"are kept in systems the operator controls, and therefore systems the operator can curate after the fact."* When the actor moves faster than any human can intervene, and the only record is stored in a store that the operator can rewrite, the verification gap ceases to be an inefficiency and becomes an evidential impossibility.

4. Finding two: five mechanisms convert the chronic gap into an acute one

Each mechanism was evident in the evidence; together, they describe a single escalation. For each, the research identifies what the record would need to contain to answer it and specifies those requirements in concrete fields.

Explainability fails at the boundary. The EU AI Act requires deployers of high-risk systems, a category covering much of underwriting, claims, lending, and benefit determination, to explain not just what a system decided but how, on what inputs, under what oversight; Article 86 extends to affected persons a right to a meaningful explanation. Within one firm, this is hard but achievable. Across a boundary, it is not: a system cannot explain a reasoning chain whose upstream links never arrived in structured form. The duty to explain falls on the party least able to discharge it, the one holding the output but not the provenance.

Agentic volume outruns human audit. When an agent decides faster and more often than any supervisor can review, the response cannot be an inspection of each decision. It must be a signed record of the agent's rationale that a supervisor, and in time a regulator, can sample against. This is also where silent drift hides: when reviewers quietly stop overriding an agent, or the share of decisions marked "not reviewed" climbs, the shift is invisible unless the disposition mix itself is monitored. Pure information-gathering without the capacity to act, as one policy analysis notes, is monitoring, not governance.

The models do not hold still. A decision recorded today against a given model state is not the artifact that the same model would produce next week. A Group Data Protection Officer put it operationally: *"Am I going to get the same answer next week, based on the artificial-intelligence learning," at rates of change by the minute, by the day, by the hour."* The record must therefore capture the model's version, configuration, and state at the moment of decision, including the inference conditions, the reasoning budget, the tools it could call, and the depth of sub-agent delegation that shaped its realized capability.

The medium removes the review artifact. Document-based AI at least leaves something for a human to review. Voice AI removes the review step and the artifact that came with it: the loop where accountability lived is gone, and by the time anyone could step in, the action has already happened. The seal must then be captured at the moment of decision, not at the moment of review, in any medium whose design has removed the review stage.

The composite human-and-machine decision. Two parts. First, reliance: an inexperienced handler may lean on a system's output for the bulk of an analysis, whereas an expert leans on it for only a fraction, and a record noting only that AI was "used" cannot provide evidence that the human review was actually applied. The fields a graduate-scheme practitioner designed unprompted, the proportion relied upon, the tool and version, and the supervisor's approval are the answer. Second, recursive validation: where AI produced part of a decision, the validation of that part currently sits with neither the system nor the reviewer, and a downstream party cannot build on a foundation whose machine-produced portion is itself unvalidated.

A distinction runs through all five: a human can be in the loop and still not be judging. If the person approving a recommendation lacks the context, authority, evidence, and time to challenge it, the institution has not preserved judgment; it has relocated liability. Telling the two apart means recording not only what the human did, but whether the human was positioned to do it.

5. Finding three: current artificial-intelligence governance certifies the wrong thing

The five mechanisms make a portable decision record necessary. A separate finding shows that the records current regimes produce are the wrong kind, and it is the strongest single point here.

The EU AI Act, the most developed instrument in force, delegates conformity assessment for most high-risk systems to providers' own self-assessment, with no routine public inspection; such auditing tests the internal system rather than the decision. The empirical record on self-certification is poor: self-certified consumer-product recalls in the European Union have occurred nine to twenty times as often as third-party-verified ones. The PIP breast-implant case is the cautionary instance, a valid certificate issued against the manufacturer's quality-management system, without inspecting the implants, behind which industrial-grade silicone reached tens of thousands of women. The failure was not the absence of a seal. A seal existed, was valid, and was worthless because it certified the paperwork rather than the artifact and could not be checked against the thing it vouched for.

The market's reflexive answer is a shared management-system standard, and ISO/IEC 42001, the AI-management-system standard, is increasingly treated as the common reference. It is a useful baseline. But a management-system standard certifies that an organization runs a governance process; it does not make a single decision portable or verifiable by the counterparty that must rely on it. Adherence tells a relying party that the issuer has a conformant process, not what was decided, under whose authority, against which rules, in the case at hand. The governance frameworks proliferating across 2026, the firm-internal inventories, the self-attested conformity declarations, and the policy audits are all of this kind: they certify a system rather than a decision and cannot travel to a relying party in independently verifiable form.

A second pattern compounds the first. A large and useful literature now governs AI at the level of the single firm, with decision records, model and data lineage, freshness-triggered re-approval, and audit trails. Each is an advance; each stops at the firm wall. The most revealing case is an executive guide that turns the firm-internal posture into strategy, urging carriers to capture every decision and then "own the data" as a competitive moat. That is boundary blindness prescribed rather than diagnosed. The market is already building the decision record; what it is building is a wall. Firm-internal governance is necessary and not sufficient, and the layer it leaves unbuilt, the portable record that crosses the boundary under an open standard, is the layer this research names.

6. Finding four: the gap is cross-industry

The condition is not an insurance peculiarity; it appears wherever regulated data crosses an organizational boundary, and the AI amplifier applies in each case.

In **customs and trade**, marine cargo insurance and customs share the same underlying data. Yet an insurer's verification of documentation does not reach the customs authority, which verifies it again. The EU Deforestation Regulation sharpens this: the same sourcing evidence must satisfy both insurance and trade, and each verifies independently because no portable container carries the acceptance decision between them.

In **logistics**, a provider's sensor data and chain-of-custody records establish that a shipment was damaged; the insurer reviews the claim, the reinsurer reviews the insurer; and each begins from zero because the prior verification does not travel in a structured form. Parametric cover triggered by a sensor reading should flow straight from the logistics layer to the insurer, but no structured format

carries the provider's decision across. A supply-chain standards chairman has observed that the industry spent years trying to solve this with distributed ledgers, only to find the real gap was never the ledger but the decision metadata that should travel with the goods.

In **banking and trade finance**, financing a shipment depends on insurance, but the bank extending credit sees only that a policy exists, not whether the insurer has verified the risk. United States legislation of 2025 draws a clean line between infrastructure that lets a system function and intermediation that exercises discretion over transactions; a portable decision record sits on the infrastructure side, carrying the assurance without controlling the decision.

In **anti-money-laundering**, the duplication is continuous. Every institution in a correspondent chain runs its own due diligence, sanctions screening, and transaction monitoring afresh, because the prior institution's verification does not travel in a form it can rely on. The Financial Action Task Force's travel rule moves originator and beneficiary data along the chain, but not the compliance decision behind it: a correspondent receives the payment, not the basis for clearance. The cost shows up in wholesale de-risking, where banks exit entire categories of counterparties rather than rely on controls they cannot verify. AI sharpens this: monitoring and customer-risk models generate and increasingly triage the alerts, but when an alert is escalated to a financial intelligence unit, the model, the analyst's disposition, and the reliance placed on the machine do not travel with the outcome.

Two features make the picture worse than the sum of its parts. First, the existing standards, ACORD in insurance, the World Customs Organization, GS1 in product data, ISO 20022 in payments, the STIX and TAXII threat formats, standardize the data content but not the acceptance decision about it. The rails carry the boxcar; none carries the seal. Second, the symptoms do not scale linearly: a single cross-border shipment may cross insurance, customs, logistics, the bank that finances it, and the anti-money-laundering screening on its payment, and each boundary multiplies the duplication into a web of bilateral verification gaps. The Digital Operational Resilience Act, which requires financial entities to manage risk across their full cross-sector supply chain, depends on the structured cross-boundary metadata that the current infrastructure does not provide.

7. Finding five: the architecture is being independently reinvented

If the diagnosis were idiosyncratic, the absence of a portable decision record could be read as a sign the market does not want one. The opposite is true: the architecture is being reinvented in adjacent regulated domains by parties who have never encountered this research, and AI is where the reinvention is densest, because that is where the need is most acute.

The instances assemble into one specification. An agentic-infrastructure proposal pairs a "proof of origin, process, and state" record with a persistent agent identity, storing only cryptographic fingerprints so the data stays with each participant. An asset-verification standard attests to an item's condition through a hash-anchored record whose evidence remains off-ledger, gated by a freshness check that reverts a transaction once the attestation lapses. A content-authenticity specification, in production across camera hardware and editing software, does the same for media. None generalizes beyond its vertical, and none names another. Yet, each builds the same primitive: a record that is distributed, signed, separable from the underlying data, and verifiable across a boundary.

Read with the five mechanisms, these efforts converge on a single recurring specification. A portable decision record adequate to AI carries, at a minimum:

- the identity of the decision-maker and the authority under which they acted, with the means to establish whether that authority was live at the moment of decision and to answer the question later, after it has lapsed or been revoked;
- the model's identity and state at that moment: version, configuration, and the inference conditions (reasoning budget, tool access, sub-agent depth) that shaped its realized capability;
- the human's disposition (accepted, modified, overridden, not reviewed), whether the human was positioned to judge, and how substantive the review was;
- the proportion of the decision relied upon from the machine, so genuine judgment is distinguishable from relocated liability;
- a hash of the explanation and evidence, so the content stays with its owner while a commitment to it travels and can be checked on demand;
- the risk tier and any prohibited-use class, as coordinates rather than conclusions, so the relying party applies its own rules;
- the conditions under which the record may be consumed (party, platform, industry, jurisdiction, purpose, time, sensitivity), evaluated so that an unmet condition fails closed;
- A cryptographic seal binding all of the above to the decision, verifiable by any party without trusting the issuer, with explicit treatment of freshness, supersession, and revocation.

What this paper proposes is that this specification be expressed as an open, permissively licensed standard, governed as a public good rather than owned by any vendor. That governance matters for the reason the standards-landscape finding does: markets adopt shared infrastructure when they trust that it belongs to no one. The point is not which standard prevails, or even that one yet exists. It is that several independent parties, working from different sectors and unaware of one another, have already arrived at the same fields, the strongest available evidence that the underlying deficit is real and singular.

8. A proposed architecture: a coordination layer, not a platform

A diagnosis implies the shape of a remedy, and this one implies a shape easily misread. The remedy is not another platform, a data lake, a shared ledger, a workflow system, or anything a party is asked to rip out and replace. The market has tried versions of all of these, and the gap has survived them because the gap is not a missing platform but a missing layer: the portable record that travels when a decision crosses a boundary. What the evidence implies is a coordination layer that sits beside existing systems and adds the one thing the data rails do not carry: the seal on the decision. The properties below are requirements any solution must meet, not a product specification.

It produces a record, not a copy: when a material decision is made, the layer emits a signed record and nothing else, while the underlying data stays where it is and only a hash of it travels. There is no central pool and no requirement that two firms share infrastructure for one to rely on the other. It is self-verifying: a receiving party checks the record offline, against a published key, without trusting the issuer or connecting to its systems, the property that separates a portable record from a log, and the one the independent efforts above have already built. It carries coordinates, not conclusions: the record states where the decision sits, its risk tier, authority, and jurisdiction, so the relying party

applies its own rules rather than inheriting the issuer's. It honors selective disclosure: different recipients see different subsets of a single record, preserving the controlled asymmetry the market already relies on. It is captured at the moment of decision, carrying the model's state, the authority in force, and the human's disposition at that instant, and treating freshness, supersession, and revocation as first-class. And it displaces nothing: ACORD, ISO 20022, and the customs and product-data standards carry the data as they do today, while the coordination layer rides alongside, adding the seal. The boxcar already rolls; what is supplied is the part that lets the party at the far end trust the contents were examined, by whom, and under what authority, without opening the box.

9. Governing the layer: an open standard that belongs to no one

The architecture has a governance problem built in, and governance is inseparable from the design. A portable record only works if the parties on both sides of a boundary trust the record format itself, and they will not extend that trust to a format owned by a competitor or by a vendor that can change the rules or charge for access. Neutrality is therefore a requirement, not a nicety. The principles below are offered as a proposal.

A neutral body should steward the format, not a vendor, governed as a public good by an organization whose purpose is the standard's integrity rather than the sale of any implementation. It should be openly licensed and freely implementable, so implementations compete on merit while the format stays common. Conformance should be open and testable by anyone against published tests, rather than certified behind a paywall, because the purpose of conformance is interoperability, not gatekeeping. And the standard and its implementations should be kept distinct: the open format is the public good, while the systems that produce and consume records under it may well be commercial. This is the same line regulators are beginning to draw between infrastructure that lets a system function and intermediation that exercises discretion within it. Infrastructure earns adoption precisely by not being anyone's product, and a neutral nonprofit standards body is the natural steward for a format that must be trusted by parties who do not trust each other. That is the governance model the evidence points toward.

10. Implications, open questions, and what would falsify this

The implication for organizations is that firm-internal governance, however mature, addresses the easier half of the problem. The moment an AI-influenced decision crosses to another firm, platform, or jurisdiction, the governance that produced it does not travel, and the relying party is left with an outcome it cannot interrogate. The implication for regulators is that a regime built on self-certified, firm-internal paperwork will not produce evidence a counterparty or supervisor can rely on at machine speed, and that the standards layer needed to make the obligations operational does not yet exist. The deferral, in 2026, of the EU AI Act's high-risk obligations on the stated ground that the harmonized standards are not ready is a concession of exactly this.

The implication for the architecture, set out in Sections 8 and 9, is the narrowest and most testable: the missing layer is a portable, tamper-evident decision record carried by a coordination layer that belongs to no one, not a platform behind a proprietary wall, and artificial intelligence is what turns its absence from an inefficiency into a liability.

Several questions remain open, and the research is explicit about them. The diagnosis is converging, not concluded. The interview base is weighted toward United Kingdom and United States insurance, and the cross-industry observations in customs, logistics, banking, and anti-money-laundering

require dedicated validation in their own right rather than by analogy. No complete security model, privacy and selective-disclosure scheme, or tested interoperability layer is offered here. The thesis would be weakened if the recurring specification reflected a single intellectual fashion rather than a structural cause, if firms proved able to discharge cross-boundary explainability without portable records, or if the market closed the gap durably behind proprietary walls. The opposite would strengthen it: further independent reinventions of the same primitive, and a demonstration that a portable decision record can be produced by one party and consumed by another, across a real boundary, in a way both can verify.

That empirical step, producing and consuming a portable decision record across real environments in more than one industry, is the natural next phase. The finding that motivates it is the one this paper opened with: artificial intelligence did not create boundary blindness, but it has removed the conditions under which the market could afford to ignore it.

Sources referenced include the EU Artificial Intelligence Act (Articles 13, 14, 25, 27, 50, 86) and the 2026 Digital Omnibus deferral; the Financial Conduct Authority Emerging Technology Horizon Scan (June 2026); the Australian Prudential Regulation Authority Letter to Industry on Artificial Intelligence (April 2026); Munich Re, "Mind the Gap" and the Tech Trend Radar 2026; the Coalition for Content Provenance and Authenticity (C2PA) Content Credentials specification; ISO/IEC 42001:2023 and the NIST AI Risk Management Framework; the EU Deforestation Regulation; the Digital Operational Resilience Act; FATF Recommendation 16; and sixty-one practitioner interviews conducted October 2025 to May 2026 under role-level anonymity. Full citations are held in the underlying thesis, "The Aetiology of Boundary Blindness."

About the author: Alexander D. Barrett is an independent researcher and a Senior Fellow at the Verida Charter Foundation, a nonprofit focused on cross-firm coordination in regulated multi-party markets. The findings here come from a twelve-month program of senior-practitioner interviews and cross-sector analysis into one question: what breaks when AI-influenced decisions have to be relied on across firms, not only inside one.

Disclosure. This research is independent and self-funded; no party has funded or directed it, and the findings are published in full regardless of whether they favor any particular approach. To test the concepts described here, the author has built prototype implementations, which are his own work. Any standard arising from this research is intended to remain open and freely implementable by any party, and a future implementation may be operated commercially without affecting that openness. The paper proposes a direction; it does not sell a product.