THE HRAIS CHAMBER

Notes on Operationalizing Institutional Reconstructibility

JM García-Maceiras

President of the Spanish BPO Banking Association

Abstract

High-Risk AI Systems (HRAIS) do not primarily challenge institutions because they are opaque; they threaten them because opacity disrupts the chain through which responsibility is established under contestation.

A credit decision is denied; a fraud alert is triggered; a customer is excluded—in each case, the output functions as an institutional reason. Yet when queried by a court, a supervisor, or a counterparty, the institution may be unable to reconstruct, in attributable terms, how that reason was produced. This is not a deficiency of explainability: it is a failure of reconstructibility.

Where reconstructibility collapses, governance does not degrade internally but becomes displaced externally. In this sense, the absence of reconstructibility is not merely an informational deficit, but a probative failure**.** Institutions may continue to operate systems with acceptable performance, validation, and regulatory compliance. Yet if they cannot reconstruct the causal chain linking data, model configuration, human supervision, and decision outcome, those decisions become increasingly difficult to defend under adversarial scrutiny.

Moreover, reconstructibility does not fail abruptly; it degrades progressively under operational conditions. As systems evolve—through data drift, threshold recalibration, accumulated exceptions, vendor dependencies, partial modifications—the evidentiary chain may fragment without any visible decline in performance. This phenomenon, described here as *reconstructibility drift*, produces a distinct form of institutional fragility: systems that remain operationally valid while becoming institutionally difficult to defend.

This paper develops the HRAIS Chamber not as an advisory body, but as an accountability mechanism designed to preserve reconstructibility as a condition for retaining institutional authority over high-risk decisions.

The Chamber structures governance *ex ante* through an evidentiary architecture centered on the *Reconstruction File* (RF), supported by instruments such as the *Exception Ledger* (EL), the *Causal Chain Map* (CCM), and the *Decision Attribution Record* (DAR). As system complexity increases, additional mechanisms—such as reconstructibility monitoring and attribution stress testing—become necessary to anticipate degradation and preserve reconstructibility under adversarial conditions.

The objective is not to eliminate opacity; it is to ensure that opacity remains institutionally governable. Where reconstructibility fails, decisions do not stop, yet they cease to remain fully under institutional control.

Keywords: High-Risk AI Systems (HRAIS); institutional reconstructibility; AI governance; accountability; explainability; opacity risk; evidentiary architecture; banking supervision; decision attribution; adversarial scrutiny; governance under opacity; infrastructural constitutionalism.

THE SKETCH

Wrong Question—The issue is not whether institutions adopt principles, values, or ethical frameworks, nor whether systems can be explained in technical terms. The crucial topic is whether institutions can reconstruct, in attributable and evidentiary terms, the decisions produced by high-risk systems when those decisions are contested.

Current Drift—Responsibility collapses into narrative reconstruction and is progressively reassigned by external actors. Where reconstructibility is insufficient, liability expands and evidentiary asymmetries intensify.

Concrete Exposure (I)—The problem is no longer hypothetical. A credit denial is issued through a high-risk system integrating multiple data sources and a third-party scoring component. The customer challenges the decision. Under scrutiny, the institution can demonstrate model validation and regulatory compliance. It cannot reconstruct, in attributable terms, the causal chain linking data inputs, transformations, and threshold configuration to the specific outcome. The court determines whether the decision can be sustained as an act of the institution.

Concrete Exposure (II)—A fraud detection system flags a transaction and blocks customer access. The system performs within expected accuracy parameters. However, repeated exceptions—manual overrides justified *ex post*—have altered its operational logic without corresponding evidentiary capture. When the case escalates to supervisory review, the institution cannot demonstrate which configuration governed the contested decision. Performance remains intact; attribution does not. The system is not deemed incorrect, but indefensible.

Structural Gap—Power has moved upstream, toward architectures, thresholds, and system design; responsibility has not. The result is structural: decisions without identifiable authorship.

Nuclear Concept—Reconstructibility is the institutional capacity to recover, under adversarial scrutiny, a causal chain linking data and model lineage, transformation into domain-relevant variables, human supervision, corporate oversight, and externally intelligible justification. It is not technical interpretability, but institutional accountability. Reconstructibility defines not only whether decisions can be explained, but whether they can be proven as attributable acts of the institution.[1]

Levels—This capacity operates at two levels: system-level (architecture, thresholds, design) and decision-level (individual outputs under challenge)—both are required: neither is sufficient alone.

Threshold—Only judicial reconstructibility stabilizes governance. Courts do not resolve epistemic limits; they assign responsibility. Where reconstructibility fails, responsibility is not clarified—it is displaced. Reconstructibility is to AI governance what solvency is to banking. Judicial reconstructibility marks the boundary of institutional authority.[2]

Dynamic Constraint—Reconstructibility does not fail abruptly; it degrades progressively under operational conditions. Data evolves, thresholds are recalibrated, exceptions accumulate, system configurations shift over time. These changes may preserve performance while fragmenting the evidentiary chain. This phenomenon—*reconstructibility drift*—produces systems that remain operationally valid while becoming increasingly fragile under contestation.

Functions—The Chamber preserves reconstructibility as a continuous condition of operation. It structures evidence *ex ante*, treats opacity as a distinct risk, intervenes upstream in system design, monitors degradation over time, governs exceptions. A system that cannot be reconstructed cannot be governed. A system that cannot withstand contestation cannot be sustained.

Evidentiary Architecture—Governance is not complete until it produces evidence. The central element is the *Reconstruction File* (RF), the minimum evidentiary structure required for institutional accountability. No decision is institutionally valid in its absence. The RF must enable reconstruction of the causal chain under adversarial conditions, including model and data lineage, transformation pathways, human supervision, decision logic, and justification.

Additional Instruments—The RF operates with supporting structures that preserve continuity of attribution across the chain: the *Exception Ledger* (EL) records deviations

[1] The concept of *reconstructibility* developed here should not be confused with interpretability or explainability in machine learning literature. It refers instead to the institutional capacity to recover an attributable chain linking system operation, decision production, and responsibility allocation.

[2] See also JM García-Maceiras, *Tolerance for Opacity*, ADR Notebooks No. 3 (2026), discussing the conditions under which opacity may remain institutionally tolerable.

and overrides; the *Causal Chain Map* (CCM) represents the system's causal architecture; and the *Decision Attribution Record* (DAR) anchors responsibility across actors and layers. As system complexity increases, further mechanisms become necessary to sustain reconstructibility over time, including monitoring of reconstructibility degradation and *ex ante* stress testing of attribution under adversarial scenarios.

Authority—Responsibility requires power. The Chamber must be able to block deployment, require redesign, suspend systems, escalate exposure. Without such authority, reconstructibility remains declarative.

Under Contestation—When a decision is challenged, performance is secondary. The question is not whether the system worked, but whether responsibility can be established. Where reconstructibility fails, the institution cannot produce evidence sufficient to sustain its position. Attribution diffuses, burden shifts, and liability expands. Where it holds, the chain is visible, responsibility is attributable, and defense becomes possible.

Consequence—Opacity degrades reconstructibility, what conditions authority. Where reconstructibility fails, governance is not weakened—it is replaced. Not internally, but externally: by courts and supervisors, by crisis dynamics. The Chamber is the institutional response. What cannot be reconstructed will be reassigned. What cannot be attributed will not be controlled. What cannot be defended will not endure. Institutions that cannot reconstruct their decisions will not retain authority over them.

THE SHAPE

Opening

High-Risk AI Systems (HRAIS) do not present themselves to institutions as abstract governance problems. They appear as decisions. A credit application is denied; a transaction is blocked; a customer is flagged. In each case, the outcome is treated—internally and externally—as an institutional act. It carries consequence, and it demands justification.

The difficulty emerges later. When these decisions are challenged—by a client, by a supervisor or a court—the institution is required to do something more demanding than explaining a model. It must establish, in attributable terms, how that specific outcome came to be, and under whose responsibility it stands.

In an increasing number of cases, this reconstruction is not available. Not because the system is incorrect or lacks performance or validation, but because the chain linking design, data, transformation, and decision has not been structured in a way that can be recovered under scrutiny. This is the point at which governance ceases to be an internal matter. The question is no longer whether the system works, but whether the institution can still stand behind it.

The governance of artificial intelligence in banking has been framed, to date, in terms that remain largely continuous with existing control paradigms: principles, model validation, compliance structures, oversight committees. These instruments are necessary, but unsatisfactory.

High-Risk AI Systems (HRAIS) alter the conditions under which decisions are produced. Decisions are no longer discrete acts attributable to identifiable agents, but outputs of systems whose operative logic is distributed across data pipelines, model architectures, and organizational processes. The production of reasons shifts from human deliberation to system configuration. Power moves upstream; responsibility does not.

Institutions remain accountable for decisions whose causal chains they may not be able to reconstruct under scrutiny. Where reconstructibility degrades, accountability does not disappear; it is reassigned—externally and without deference. Courts, supervisors, and counterparties do not resolve epistemic limitations; they allocate responsibility.

The resulting problem is not adequately captured by existing governance categories. It is not reducible to model performance, nor to compliance with regulatory requirements, nor to ethical alignment in the abstract. It concerns the institution's capacity to sustain a defensible chain of responsibility linking system design, operation, and outcome.[3]

This paper takes that capacity—reconstructibility—as its organizing concept. Building on prior work that frames explainability as a distinct banking risk and situates AI governance within a judicial nexus of causality and liability, the paper develops the HRAIS Chamber as an institutional form designed to preserve reconstructibility under conditions of opacity. The Chamber is not conceived as an advisory body, but as an accountability mechanism with defined competences, authority, and evidentiary requirements.

The analysis proceeds from problem to design. It first positions the Chamber within the existing governance structure of the institution, identifying the gap it addresses. It then defines its composition as a configuration of responsibility nodes, specifies its competences and decision process, and develops the evidentiary architecture required to sustain its function. The paper further examines its integration with the risk framework, outlines a staged implementation, and identifies limits and failure modes.

The objective is not to provide a procedural manual, nor to replicate existing control functions. It is to define a minimal institutional architecture capable of preserving responsibility where the conditions of decision-making have fundamentally changed. Reconstructibility is not a technical feature; it is a condition of governance.

The HRAIS Chamber is presented here as an experimental institutional form. Its purpose is not to offer a finalized governance solution, but to define a structure capable of being

[3] See Frank Pasquale, *The Black Box Society* (2015); Mireille Hildebrandt, *Law for Computer Scientists and Other Folk* (2020).

tested under real conditions of deployment, operation, and contestation. Its validity is therefore not assumed; it must be established in practice.

Under Pressure

The issue becomes visible when decisions are contested. A credit decision is escalated following a client complaint. The system integrates internal data, external scoring, and multiple transformations across business lines. Model validation has been completed. Documentation is in place.

At the moment of review, however, the institution cannot establish, with adequate clarity, how the relevant inputs were combined and weighted under the configuration active at the time of decision. Different functions provide partial accounts. None provides a complete, attributable chain. The discussion does not revolve around model accuracy; it twists to whether the institution can defend the decision as its own.

In another case, a fraud system blocks customer activity. The system performs within expected thresholds but has accumulated a pattern of manual overrides and incremental adjustments. These changes have not been consistently captured in a structured evidentiary form.

When the case reaches supervisory attention, the question is simple: *which system produced this decision?* The institution cannot answer unambiguously. At that point, performance becomes secondary. The issue is no longer the correction of the system, but whether the institution retains control over what it has produced.

At the Table

The relevance of reconstructibility is not theoretical. It emerges in decision forums. At the point where a system is to be approved, challenged, or suspended, the discussion tends to converge on a limited set of questions: *Can we explain this system? Is it compliant? Is it performing within tolerance?* These questions are imperative but incomplete.

A different question determines the outcome under pressure: *Can we defend a decision produced by this system, as our own, under external scrutiny?* Where the answer is uncertain, approval becomes a transfer of risk rather than an exercise of control. The function of the Chamber is to force that question to be answered before the system is relied upon.

Institutional Positioning

The Chamber is not an additional layer within existing governance structures. It is a structural insertion at the point where decision-making power has already migrated: upstream, into system design, model configuration, and deployment thresholds. Its positioning cannot be derived from existing committees. It must be defined from the logic of responsibility.

In traditional banking governance, responsibility is anchored in identifiable decisions, taken within established committees and documented through conventional control frameworks. HRAIS disrupts this equilibrium. Decisions are no longer discrete acts but outputs of systems whose operative logic is distributed across data, models, vendors, internal processes. Power is exercised before the decision appears. Responsibility, however, remains assessed after the fact.

The result is a misalignment between where decisions are shaped and where responsibility is attributed. The Chamber is positioned precisely at that fault line. It does not replicate the functions of the Risk Committee, the Model Risk function, or Legal and Compliance. Each of these retains its domain. The Chamber integrates them at the point where reconstructibility must be preserved as a condition of accountability.

Model Risk evaluates performance, robustness, and validation of models; it does not ensure that decisions derived from those models can be reconstructed under adversarial scrutiny. Legal assesses compliance and exposure once decisions are challenged; it does not structure the evidentiary chain *ex ante*. Risk Committees oversee aggregated exposures and capital implications; they do not intervene in the causal architecture of individual decision systems.

The Chamber operates where these functions intersect but cannot substitute for one another. Its role is not to assess whether a model is valid, nor whether a process is compliant, but whether the institution can sustain a defensible chain of responsibility from system design to individual outcome. This is a distinct question, and it requires a clear-cut institutional locus.

For this reason, the Chamber must be structurally linked to the highest level at which responsibility ultimately resides. Its outputs must be intelligible and actionable at Board level, even if its operation is embedded within the first and second lines of defense.

This dual positioning is essential. If placed too low, the Chamber becomes operational and loses authority; if placed only at Board level, it becomes abstract and loses traction. It must instead function as a translating mechanism: converting technical and procedural complexity into attributable responsibility.

Its authority must therefore be recognized across the three lines of defense: where systems are designed and deployed, cannot bypass it; where risk and compliance are structured, must incorporate its outputs into their frameworks; through internal audit, must be able to assess its functioning as part of the institution's control environment.

Composition

The composition of the HRAIS Chamber cannot be defined as a list of functions; it must be defined as a configuration of responsibility nodes. HRAIS fragments the production of decisions across multiple domains: technical design, data structuring, business objectives, legal framing, risk control. No single function holds the full chain. The Chamber must therefore assemble, in a single locus, the points at which responsibility is

effectively exercised. Its composition is not representative, but structural. At minimum, the Chamber must integrate four irreducible nodes.

A/ Technical node, where system design, model configuration, and data transformations are determined. This node carries responsibility for the internal logic of the system: how inputs are processed, how outputs are generated, and how thresholds are set. Without its presence, reconstructibility collapses at the point where decisions are actually framed.

B/ Risk node, where exposure is measured, classified, and escalated. This node translates opacity into institutional risk, linking reconstructibility failures to supervisory and capital implications. It anchors the Chamber within the prudential logic of the institution.

C/ Legal node, where decisions acquire external meaning. This node determines whether the institution can sustain its positions under challenge, and whether the causal chain produced internally can be defended in adversarial settings. It is here that reconstructibility becomes evidence.

D/ Business node, where decisions produce economic effects and strategic direction. This node ensures that responsibility remains anchored in the domain where consequences materialize. Without it, the Chamber risks becoming detached from the very decisions it is meant to govern.

These nodes are not interchangeable. Each corresponds to a well-defined dimension of responsibility that cannot be absorbed by the others. Additional nodes may be incorporated—compliance, internal audit, data governance—but their inclusion must follow inexorableness, not convention. The Chamber is not a forum for broad representation, but a mechanism for assembling the minimum set of perspectives required to preserve reconstructibility.

The individuals occupying these nodes must be senior enough to commit their respective domains. Delegation without authority undermines the Chamber at its core. Participation is not advisory; it is binding. For the same reason, the Chamber cannot operate through loose attendance or informal consultation. Its composition must be stable, identifiable, and accountable over time. Responsibility cannot be reconstructed if the actors themselves are not traceable.

This does not imply rigidity. The Chamber may expand or contract depending on the system under review, but its core nodes must remain present. Flexibility applies to context, not to responsibility. Finally, the composition of the Chamber must mirror the structure it seeks to govern. HRAIS distribute agency across systems and functions. The Chamber reassembles that distribution into a visible and accountable configuration.

Competences

The competences of the HRAIS Chamber are defined by a single constraint: responsibility cannot be assumed without the power to intervene where decisions are shaped. The Chamber does not evaluate models in the abstract, nor does it certify compliance in isolation. Its remit is narrower and more demanding: to determine whether a High-Risk AI System can be deployed, operated, and defended with a reconstructible chain of responsibility.

From that route, the Chamber sets out its competences:

1/ The Chamber determines the *reconstructibility status* of a system. This is not a technical rating but an institutional judgment: whether the causal chain linking design, data, transformation, supervision, and outcome can be recovered under adversarial scrutiny. Where reconstructibility is insufficient, deployment cannot proceed.

2/ The Chamber *authorizes deployment*. Authorization is conditional, not declarative. It attaches to a defined configuration—model, data, thresholds, governance conditions—and lapses if that configuration materially changes. There is no standing approval independent of system state.

3/ The Chamber *requires and validates the Reconstruction File* (RF). No high-risk system enters production without an RF capable of sustaining judicial reconstructibility. The Chamber may require completion, restructuring, or rejection of the RF. Absence or insufficiency of the RF precludes deployment.

4/ The Chamber *classifies and escalates opacity.* Opacity is treated as an unambiguous exposure. Where opacity reaches defined thresholds, the Chamber imposes constraints: restricted use, additional controls, redesign, or suspension. Where residual exposure exceeds tolerance, escalation to the Risk Committee or Board is mandatory.

5/ The Chamber *governs exceptions*. Deviations from standard system behavior require prior authorization, explicit reasoning, and recorded counterfactuals. Repeated exceptions trigger review of the system itself.

6/ The Chamber *imposes design conditions ex ante*. It may require changes to features, thresholds, human-in-the-loop controls, or documentation structures where these are necessary to preserve reconstructibility. Intervention occurs before harm materializes.

7/ The Chamber *suspends or withdraws authorization.* Where reconstructibility degrades, where RF integrity is compromised, or where operational practice diverges from approved conditions, the Chamber may halt deployment.

8/ The Chamber *structures the evidentiary posture of the institution*. It determines whether the RF and associated instruments—the *Exception Ledger* (EL), the *Causal Chain Map* (CCM), and the *Decision Attribution Record* (DAR)—are sufficient to sustain the institution's position under challenge.

These competences are not advisory—they are binding and bounded. The Chamber does not replace Model Risk, Legal, or Compliance. It does not set business strategy. It does not manage day-to-day operations. It intervenes only where reconstructibility and responsibility are at stake.

Two principles govern the exercise of these competences: (a) *configuration specificity*: decisions attach to defined system states and documented conditions; change the system, and the decision must be revisited; and (b) *adversarial sufficiency*: the relevant standard is not internal comfort but external challenge—what cannot withstand scrutiny cannot be authorized.

Functions

The functions organize the exercise of Chamber`s competences across the lifecycle of high-risk systems. The objective is continuity: reconstructibility must be preserved before deployment, during operation, and under challenge. The Chamber does not operate episodically; it maintains a persistent relation with the systems it governs.

I/ *Ex ante*, the Chamber acts at the point where systems take shape. It reviews design choices, data dependencies, feature construction, threshold configuration insofar as they affect reconstructibility. The question is not whether the system performs, but whether its future decisions will be attributable. At this stage, the Chamber imposes conditions: requirements on the *Reconstruction File*, constraints on opacity, design adjustments necessary to sustain a defensible causal chain. Authorization, where granted, is conditional and configuration-specific.

II/ *During operation*, the Chamber maintains oversight of the system as an evolving object. High-risk systems do not remain static; data drifts, models are recalibrated, thresholds are adjusted, and usage expands. Each of these changes may degrade reconstructibility. The Chamber monitors this degradation through the evidentiary architecture it requires: the *RF*, the *Exception Ledger*, the *Causal Chain Map*, and the *Decision Attribution Record*.[4] Where deviations accumulate or conditions are breached, the Chamber intervenes. Authorization is not a one-time event, but a maintained state. The Chamber ensures that the chain of responsibility remains intact as the system evolves. Where preservation fails, operation cannot continue under the same terms.

III/ *Ex post*, the Chamber operates under contestation. When a decision is challenged—by a customer, by a supervisor or a court—the Chamber becomes

[4] See *Appendix* for *Illustrative Supplementary Instruments*.

the locus where the institution's position is reconstructed. It does not generate explanations ad hoc; it relies on the evidentiary structure produced *ex ante* and maintained during operation. Defense that cannot reconstruct the causal chain exposes the institution beyond the individual case.

These three functions—*ex ante* conditioning, ongoing preservation, and *ex post* reconstruction—are not separable phases but a continuous cycle. Weakness in any phase propagates to the others. A system authorized without reconstructibility will fail under operation. A system operated without preservation will fail under challenge. Its functions follow a simple rule: governance must anticipate the moment of contestation.

Decision Process

The decision process of the Chamber is designed to produce attributable outcomes under conditions of disagreement. It is not a forum for consensus, but a mechanism for decision.

Every determination of the Chamber attaches to a defined system configuration and to an explicit evidentiary basis. Decisions are not generic approvals; they are configuration-specific authorizations, conditions, or prohibitions, grounded in the *Reconstruction File* and its associated instruments. Alter the configuration, and the decision must be revisited. The process is structured around three moments: submission, determination, and record.

Submission initiates the process. A system, or a material change to an existing system, is brought before the Chamber with a complete evidentiary package. At minimum, this includes a *Reconstruction File* capable of supporting judicial reconstructibility, together with the current state of the *Exception Ledger*, the *Causal Chain Map*, and the *Decision Attribution Record.* Incomplete submission is not deferred; it is not admitted. The process does not begin without evidence.

Determination is the act of decision. The Chamber assesses whether reconstructibility is sufficient under adversarial standards and whether the conditions for operation are met. Outcomes are limited and explicit: authorization (with conditions), refusal, or suspension. Conditionality is integral. Where uncertainties remain within tolerance, the Chamber may authorize subject to defined constraints—on use, monitoring, or redesign—with specified review points.

The Chamber does not decide by simple aggregation of views. Its composition reflects distinct responsibility nodes; disagreement is expected. Where positions diverge, the decision must render that divergence visible and attributable. Minority positions are recorded where they bear on reconstructibility or risk exposure—attribution is not diluted by consensus.

Quorum is defined by the presence of all core nodes of responsibility. Absence of any core node invalidates the decision. Delegation is admissible only where authority is preserved. Attendance without authority does not constitute quorum.

Decision rules follow a principle of responsibility anchoring. Where reconstructibility is contested between nodes, the stricter position prevails unless overruled through formal escalation. In particular, where the legal or risk node determines that the evidentiary chain is insufficient for adversarial defense, authorization cannot be granted at the Chamber level.

Escalation is not failure; it is part of the design. Where the Chamber cannot reconcile positions within its mandate, the matter is elevated to the appropriate governance level—typically the Risk Committee or the Board—together with the full evidentiary record and the explicit statement of disagreement. Escalation transfers the locus of responsibility; it does not erase it.

Record completes the process. Every determination is formalized as a decision record linked to the system configuration and to the evidentiary set on which it relies. The record specifies the outcome, the conditions attached, the rationale, and the attribution of responsibility across nodes. It also defines review triggers: events or thresholds that require the decision to be revisited. Decisions are time-bound by design. Authorization lapses upon material change, breach of conditions, or degradation of reconstructibility as evidenced in operation: there is no perpetual approval.

Two constraints govern the entire process: (a) *evidentiary primacy*: no decision is taken in the absence of a *Reconstruction File* capable of sustaining adversarial scrutiny; process cannot compensate for lack of evidence; and (b) *attributable dissent*: disagreement is not resolved by dilution but by assignment; where the Chamber decides, it does so with explicit attribution; where it cannot, it escalates with the disagreement intact. A decision that cannot be attributed cannot be defended; a process that suppresses disagreement will reproduce it under contestation.

Integration with Risk Framework

The Chamber does not operate outside the institution's risk framework—it reconfigures a gap within it. Existing frameworks are structured around identifiable risk types—credit, market, operational, model risk—each with established methodologies, metrics, and governance channels. HRAIS intersect with all of them, but are not reducible to any.

Model Risk addresses validation, performance, and robustness of models. It does not ensure that the decisions produced by those models remain reconstructible under adversarial scrutiny. Operational Risk captures failures in processes and controls, but does not account for the structural opacity embedded in system design. Legal and Compliance manage exposure once it materializes; they do not determine whether the institution can sustain its position *ex ante*—not a gap of coverage, but of articulation.

The Chamber introduces reconstructibility as a condition that cuts across existing risk categories while remaining irreducible to them. It does not create a parallel framework; it imposes a constraint within the current one. Opacity, in this context, is treated as a distinct source of exposure. Not because it produces losses directly, but because it degrades the institution's capacity to attribute responsibility. This degradation has

second-order effects: uncertainty in litigation, amplification of supervisory intervention, and, ultimately, impact on capital. Opacity is an independent vector that interacts with multiple risk types.

The Chamber translates this vector into institutional terms. Its determinations—on reconstructibility status, on conditions of deployment, on suspension—must be reflected in the risk framework as binding inputs. Systems authorized with conditions generate corresponding risk constraints; systems suspended or escalated alter the institution's exposure profile.

This translation operates in both directions. From the Chamber to the risk framework, decisions become constraints, limits, or escalations; from the risk framework to the Chamber, thresholds of tolerance—defined at institutional level—inform the admissibility of residual opacity. The connection is therefore not hierarchical but reciprocal.

In practical terms, this implies that the outputs of the Chamber must be integrated into existing governance channels: Risk Committees must receive and act upon escalations grounded in reconstructibility; capital and provisioning discussions must incorporate the uncertainty introduced by opacity; internal reporting must reflect the status of high-risk systems not only in performance terms but in evidentiary terms.

The Chamber does not replace existing risk functions: it renders them complete. Where this integration fails, opacity remains unpriced, responsibility remains diffuse, and governance becomes reactive; where it holds, reconstructibility becomes part of the institution's risk language.

Implementation Roadmap

The implementation of the Chamber is not a project layered onto existing structures, but a reconfiguration that must be staged to preserve continuity of operations while establishing a new locus of accountability. The objective is not speed, but irreversibility. The roadmap therefore proceeds in phases, each of which establishes conditions that the next cannot bypass.

Phase I — Recognition and Mandate

The first step is institutional recognition. The Chamber must be explicitly defined within the governance framework, with a mandate anchored at the level where responsibility ultimately resides. Without this, subsequent steps produce form without authority. At this stage, scope is determined. Not all AI systems fall within the Chamber's remit; only those classified as high-risk under institutional criteria. This classification must be aligned with existing regulatory definitions but cannot be limited to them. Internal thresholds may be stricter. The core principle is established: no high-risk system operates without reconstructibility. This phase does not require full operational capability. It requires clarity of intent and formal anchoring.

Phase II — Minimum Viable Chamber

The *Chamber* is constituted in its minimal form, integrating the core nodes of responsibility: technical, risk, legal, and business. Composition is stable, authority is explicit, and the decision process is defined, even if not yet optimized. A limited set of systems is selected for initial review. These are not edge cases but representative high-risk systems already in operation. The purpose is not validation, but exposure: to test reconstructibility under real conditions. The *Reconstruction File* is introduced as a requirement. It will be incomplete at this stage. That is expected. The objective is to surface gaps in evidentiary structure and attribution. Early decisions will be imperfect. What matters is that they are taken, recorded, and attributed. This phase establishes the Chamber as a functioning decision body.

Phase III — Evidentiary Consolidation

The focus shifts from decision to structure. The RF is standardized across systems. The *Exception Ledger*, *Causal Chain Map*, and *Decision Attribution Record* are implemented in consistent form. At this stage, reconstructibility becomes measurable. Not in quantitative terms, but in institutional terms: the Chamber can determine, with increasing consistency, whether a system can sustain adversarial scrutiny. Integration with the risk framework begins to take effect. Chamber determinations are reflected in risk reporting, escalation channels, and governance discussions. The Chamber moves from reactive to structured.

Phase IV — Integration and Scaling

The Chamber is fully integrated into the lifecycle of high-risk systems. No new system enters production without prior Chamber determination. Existing systems are progressively brought within scope. Decision processes stabilize. Conditions attached to authorizations become more precise. Escalations to Risk Committee or Board follow established patterns. At this stage, the Chamber's outputs are part of the institution's operating rhythm. Reconstructibility becomes embedded in design practices. System owners anticipate Chamber requirements. Governance shifts upstream.

Phase V — Maturity

The Chamber operates as an established institutional form. Its authority is uncontested internally. Its outputs are routinely incorporated into risk, legal, and business processes. Most importantly, the institution's evidentiary posture is transformed. Under challenge, decisions can be reconstructed without improvisation. Responsibility remains anchored within the institution. At this stage, the absence of the Chamber becomes difficult to figure out. The progression across these phases is not strictly linear. Feedback loops are inherent. Failures in later phases may require revisiting earlier assumptions. What matters is not sequence, but accumulation: each phase must leave behind a structure that cannot be reversed without explicit decision. Implementation fails when it is treated as compliance. It succeeds when it alters how decisions are made, recorded, and defended.

Internal Tension

The preservation of reconstructibility is not neutral within the institution. Business incentives favor speed, scale, competitive differentiation. Technical functions favor performance and optimization. Legal and risk functions require defensibility under adversarial scrutiny—these incentives do not align naturally.[5]

The *Chamber* operates precisely at this point of tension. Where reconstructibility imposes constraints on system design or deployment, it will be perceived as a limitation. Where it is bypassed, the constraint does not disappear—it reappears under contestation, where it is no longer controllable.

Limits

The Chamber does not eliminate opacity, nor does it guarantee correctness of outcomes. Its function is more exacting: to preserve reconstructibility as a condition for attributable responsibility. Its limits follow from that scope:

(i) *Epistemic*—Certain systems, by design, compress information in ways that resist full reconstruction. The Chamber cannot reverse that compression. It can only determine whether the residual opacity remains within defensible bounds. Where it does not, the only coherent response is restriction or non-deployment. There is no procedural remedy for structural opacity.

(ii) *Organizational*—The Chamber relies on the effective presence of responsibility nodes. Where authority is diluted—through insufficient seniority, fragmented mandates, or informal delegation—the Chamber's determinations lose force. Participation without the capacity to commit a domain converts decisions into recommendations. In that condition, the Chamber exists in form but not in substance.

(iii) *Evidentiary*—The *Reconstruction File* and its associated instruments can degrade in practice. Documentation may become retrospective, incomplete, or disconnected from actual system behavior. Where the evidentiary structure no longer reflects the system it is meant to represent, reconstructibility becomes fictitious. The Chamber cannot rely on evidence it cannot trust; in such cases, authorization must be revisited.

(iv) *Processual*—The decision process is designed to surface and attribute disagreement. If disagreement is suppressed—through pressure for consensus, time constraints, or hierarchical override without record—the process loses its adversarial sufficiency. Apparent alignment at decision time will reappear as conflict under contestation, where it is harder to manage and no longer internal.

[5] Diane Vaughan's analysis of normalization processes in high-risk organizations remains particularly relevant to the gradual erosion of governance integrity through accumulated operational deviations. See Diane Vaughan, *The Challenger Launch Decision* (1996).

(v) *Integration*—The Chamber does not operate in isolation. If its determinations are not incorporated into the risk framework, legal strategy, and operational practice, its outputs remain local and do not alter institutional exposure. In such cases, reconstructibility may exist in principle but not in effect. Governance becomes fragmented.

(vi) *Scale*—As the number of high-risk systems increases, the Chamber may be subject to volume pressure. If throughput becomes the dominant constraint, decisions risk becoming standardized or superficial. Scaling the Chamber requires preserving the integrity of its process while adapting its capacity. Without this, it either becomes a bottleneck or loses depth.

(vii) *Strategic Misalignment*—Where business incentives favor speed, complexity, or competitive opacity, the Chamber may be perceived as a constraint rather than a condition of governance. If unresolved, this tension leads to circumvention—formal compliance with informal bypass. The result is erosion of authority without explicit decision.

Failure Modes

Becoming Nominal—Failure occurs when the Chamber is present in governance charts but absent in decision reality. This takes recognizable forms. The Chamber is consulted after deployment rather than before it. *The Reconstruction File* is treated as documentation rather than as a condition of validity. Exceptions become routine and unexamined. Escalations are avoided to preserve speed. Decisions are recorded without attribution. In each case, reconstructibility is assumed rather than established.

Overextension—If the Chamber attempts to replace existing functions—model validation, legal analysis, operational control—it diffuses its mandate and loses precision. Its strength lies in its constraint, not in breadth.

Formalism—The evidentiary architecture may be implemented as a set of artifacts without corresponding change in decision practices. In that case, the institution accumulates documentation without increasing its capacity to defend decisions.

These failure modes are not external risks; they arise from within the institution. They cannot be eliminated by design alone; they require continuous recognition. The presence of limits does not weaken the Chamber. It defines the boundary within which it is effective. Where those limits are acknowledged and managed, reconstructibility can be preserved under real conditions. Where they are ignored, governance will revert to external imposition. It delays it only to the extent that responsibility remains visible and attributable within the institution. Its adequacy will not be determined by design, but by performance under scrutiny.

Conclusion

High-Risk AI Systems do not eliminate responsibility. They displace it to a level where it is harder to observe and, if not deliberately structured, difficult to recover. Institutions can continue to operate such systems with acceptable performance, regulatory compliance, and internal confidence. These conditions are necessary for operation, but not sufficient for control.

Control depends on whether the institution can stand behind the decisions those systems produce when they are challenged. This cannot be improvised at the moment of contestation. It must be built into the system, its governance, and its evidentiary structure from the outset.[6]

The HRAIS Chamber is proposed as a mechanism to enforce that condition. It does not resolve the limits of opacity. In its absence, institutions may continue to act. Whether they remain accountable for those actions is a different question.

Madrid, June 2nd, 2026

[6] This paper forms part of a broader line of work on infrastructural constitutionalism and institutional accountability under conditions of systemic opacity. See JM García-Maceiras, *Systemic Opacity Risk*, ADR Notebooks No. 4 (2026).

APPENDIX I

Instrumental Reconstructibility Architecture

This Appendix defines the evidentiary instruments through which reconstructibility is operationalized. These instruments do not document governance; they constitute the minimum architecture required to preserve attributable decision-making under conditions of opacity, complexity, and adversarial scrutiny.

—CORE EVIDENTIARY ARCHITECTURE—

The following instruments are constitutive of reconstructibility. In their absence, governance remains formally articulated but evidentially incomplete.

Reconstruction File (RF)

Function—The RF is the primary evidentiary structure through which the institution preserves the reconstructibility of high-risk decisions. It constitutes the minimum condition for institutional accountability. Scope—The RF must enable reconstruction of the causal chain linking: data inputs, model selection and configuration, transformation processes, threshold calibration, human supervision, decision pathway, institutional justification. Constraint—No high-risk decision is institutionally valid in the absence of an RF capable of sustaining adversarial scrutiny.

Illustrative Example—A customer challenges a credit denial issued six months earlier. The institution is able to retrieve the full *Reconstruction File* associated with the decision, including active data sources, model configuration, threshold settings, supervisory controls, and the institutional rationale governing the decision pathway at the time. The RF allows the institution to reconstruct not only the outcome, but the chain of responsibility supporting it.

Exception Ledger (EL)

Function—Records all deviations from standard system behavior, including manual overrides, exceptional interventions, and *ad hoc* adjustments. Core Elements—description of deviation; justification provided at the time; counterfactual reasoning (what would have occurred under standard operation); approving authority; timestamp and system state. Purpose—To prevent the silent erosion of reconstructibility through accumulated exceptions.

Illustrative Example—A fraud detection system repeatedly triggers manual overrides for high-net-worth clients during periods of market volatility. The *Exception Ledger* records each override, the justification provided, the approving authority, and the counterfactual outcome under standard system operation. Over time, the accumulation of overrides reveals a structural mismatch between system thresholds and operational practice.

Causal Chain Map (CCM)

Function—Represents the causal architecture of the system, enabling reconstruction of the pathways through which decisions are produced. Scope—nodes (data, models, transformations, decision points); relationships and dependencies; points of intervention; points of responsibility. Purpose—To render the system's complexity structurally visible and reconstructible under scrutiny.

Illustrative Example—During supervisory review of an SME pricing engine, the institution produces a *Causal Chain Map* identifying how transaction data, liquidity indicators, and external macroeconomic signals flow through successive transformation layers into final pricing outputs. The CCM also identifies where human intervention remains possible and which functions retain responsibility at each stage.

Decision Attribution Record (DAR)

Function—Anchors institutional responsibility across the decision chain. Scope—identification of contributing systems; identification of responsible functions; human intervention (if any); allocation of accountability across nodes. Purpose—To prevent attributional opacity in distributed decision environments.

Illustrative Example—A customer disputes the closure of an account flagged for suspicious activity. The Decision Attribution Record identifies the systems contributing to the alert, the operational team responsible for escalation, the compliance function validating the action, and the executive authority responsible for maintaining the underlying policy framework.

—EXTENDED EVIDENTIARY STRUCTURE—

These instruments extend reconstructibility under conditions of increased system complexity, vendorization, or adversarial exposure.

Configuration Snapshot (CS)

Function— Captures the exact system configuration in force at the moment a decision is produced. Includes—model version; parameter settings; threshold configuration; active data sources; external dependencies. Purpose—To prevent retrospective reconstruction detached from operational reality.

Illustrative Example—A lending decision issued on 12 March 2026 is later contested in court. The institution retrieves the *Configuration Snapshot* associated with the case, establishing the exact model version, threshold calibration, external scoring dependencies, and override conditions active at the moment the decision was produced.

Counterfactual Justification Sheet (CJS)

Function—Documents why a given configuration was selected over reasonably available alternatives. Includes—alternatives considered; trade-offs (performance vs explainability

vs stability); risks consciously accepted; rationale at time of decision. Purpose—To preserve institutional responsibility for choice, not only execution.

Illustrative Example—Prior to deployment of a fraud detection model, the institution evaluates two configurations: one maximizing detection sensitivity and another minimizing customer disruption. The *Counterfactual Justification Sheet* records why the institution accepted increased false positives in exchange for lower fraud exposure, together with the associated governance rationale.

External Dependency Attribution Addendum (EDAA)

Function— Ensures that reliance on third-party components does not dilute institutional responsibility. Includes—identification of external components; limits of transparency; internal ownership of dependency; justification for continued use. Purpose—To prevent attributional diffusion through vendorization.

Illustrative Example—A credit scoring workflow incorporates a third-party behavioral scoring component whose internal methodology remains partially opaque. The EDAA identifies the dependency, documents the limits of transparency, assigns internal ownership of the risk, and records the institutional justification for continued reliance on the external component.

Institutional Reasoning Statement (IRS)

Function—Distinguishes system output from institutional reasoning. Scope—articulation of policy-level rationale behind system use; translation of outputs into decision-relevant institutional logic. Purpose—To preserve external intelligibility of institutional decisions.

Illustrative Example—In response to a complaint regarding automated lending restrictions, the institution does not disclose model internals or feature weights. Instead, the IRS articulates the institutional reasoning underlying the system's use: prevention of unsustainable indebtedness under conditions of income volatility and elevated portfolio stress.

—DYNAMIC AND ADVERSARIAL INSTRUMENTS—

These instruments address the temporal and adversarial dimensions of reconstructibility: degradation over time and exposure under challenge.

Reconstructibility Drift Monitor (RDM)

Function— Detects the progressive degradation of reconstructibility under operational conditions. Triggers—accumulation of exceptions; loss of alignment across responsibility nodes; increasing difficulty of legal or supervisory translation; fragmentation of evidentiary chain; Key Insight—Reconstructibility may degrade while system performance remains stable. Purpose—To identify evidentiary fragility before it materializes under contestation.

Illustrative Example—Over several months, a customer onboarding system accumulates increasing numbers of manual interventions and undocumented threshold adjustments. System performance metrics remain stable. However, the RDM identifies growing difficulty in reconstructing the relationship between operational practice and approved governance conditions, triggering Chamber review.

Attribution Stress Test (AST)

Function—Simulates adversarial conditions ex ante to test whether reconstructibility can be sustained. Method—simulated litigation scenario; supervisory inspection; case-level reconstruction exercise. Outcome—identification of gaps in attribution; detection of evidentiary discontinuities. Purpose—To ensure that reconstructibility is not assumed, but tested.

Illustrative Example—Before deployment of a dynamic pricing engine, the Chamber conducts an *Attribution Stress Test* simulating both a judicial challenge and a supervisory inspection. The exercise reveals that pricing outputs cannot be translated into contestable institutional reasoning at individual customer level, resulting in suspension of deployment approval.

Post-Contest Reconstruction Review (PCRR)

Function—Institutionalizes learning following litigation, supervisory action, or contested outcomes. Scope—identification of reconstructibility failures; mapping of evidentiary breakdown; redesign requirements. Purpose—To convert individual breakdowns into systemic improvement.

Illustrative Example—Following an adverse judicial outcome concerning an automated credit denial, the Chamber conducts a PCRR and identifies that historical system configurations were not preserved consistently across model updates. The review results in mandatory implementation of *Configuration Snapshots* and redesign of evidentiary retention practices.

—PRUDENTIAL AND STRUCTURAL INSTRUMENTS—

These instruments introduce explicit control over opacity and reconstructibility as institutional risk variables.

Reconstructibility Threshold Declaration (RTD)

Function—Defines the level of reconstructibility below which system operation becomes institutionally indefensible. Scope—acceptable residual opacity; conditions triggering restriction or withdrawal; linkage to governance escalation. Purpose—To formalize reconstructibility as a boundary condition of authority.

Illustrative Example—For a real-time transaction monitoring system, the institution defines a reconstructibility threshold beyond which unexplained escalation patterns

become institutionally unacceptable. Once the threshold is breached, the system may continue operating only in advisory mode pending redesign.

Design Constraint Memorandum (DCM)

Function—Translates reconstructibility requirements into binding design constraints. Examples—exclusion of non-translatable features; requirement of traceable transformation layers; constraints on model complexity. Purpose—To enforce reconstructibility upstream, rather than compensate downstream.

Illustrative Example—Before authorization of a behavioral pricing model, the Chamber issues a DCM prohibiting the use of composite latent variables that cannot be translated into institutionally defensible reasoning. Deployment approval is conditioned on redesign of the affected feature layer.

Evidentiary Lineage Protocol (ELP)

Function—Preserves evidentiary continuity across the lifecycle of the decision. Scope—data origin and transformation; model lineage; operational context; human intervention; decision outcome. Purpose—To preserve evidentiary continuity rather than fragmented documentation.

Illustrative Example—During reconstruction of a contested account closure, the institution is able to trace the full evidentiary lineage of the decision: origin of customer data, intermediate transformations, model updates, human escalation points, and final execution pathway. The continuity of the evidentiary chain allows the institution to demonstrate not only what decision was taken, but how institutional responsibility persisted throughout the process.

—SYSTEMIC COHERENCE—

These instruments do not operate independently; their function is systemic.

Integrated Effect

Reconstructibility is preserved only where: the causal chain is visible (CCM); decisions are attributable (DAR); deviations are traceable (EL); configurations are anchored (CS); evolution is monitored (RDM); attribution can be tested (AST). Progressive Degradation—Reconstructibility rarely collapses at a single point. It degrades through accumulation of exceptions, loss of configuration traceability, fragmentation of attribution, and dependence on opaque components

These instruments are not cumulative requirements. They define graduated layers of reconstructibility. At minimal complexity, core architecture suffices. As opacity increases, additional instruments become necessary. Their absence does not prevent decisions from being made; it prevents those decisions from being defended.

Appendix II

Adversarial Cases & Decision Practice

This Appendix provides illustrative (fictional) materials reflecting how the Chamber operates under conditions of decision, disagreement, and contestation. The main body of the paper defines the Chamber as an institutional mechanism designed to preserve reconstructibility. The materials below are not extensions of that definition, but instances of its application. Their purpose is not to exhaust possible scenarios, but to expose the points at which reconstructibility is tested at authorization, at refusal and under external challenge.

The inclusion of these materials raises a natural extension: a full treatment of decision practice, including patterns of escalation, institutional learning, and cross-case analysis. That extension is not developed here. The present paper maintains a narrower objective: to define the minimum institutional architecture required to preserve reconstructibility as a condition of accountability. A comprehensive treatment of decision practice would exceed that scope and alter the nature of the work—from structural definition to operational doctrine. The materials that follow should therefore be read as boundary cases, not as a complete framework.

*

A = Decision Memorandum / Conditional Authorization

HRAIS Chamber — Decision Memorandum

System: Retail Credit Decisioning Engine (R-CDE v4.3)
Date: June 2, 2046
Submission Type: Deployment Authorization (material upgrade)
Decision: Conditional Authorization
Validity: Configuration-specific — subject to conditions and review triggers

Context

The R-CDE v4.3 system integrates internal behavioral data, external income verification feeds, and a third-party risk scoring module. The proposed upgrade introduces: (a) revised threshold calibration for default probability; (b) new composite features derived from transaction patterns; and (c) expanded reliance on external scoring component (vendor module). The system is intended for full deployment across unsecured retail credit products.

Evidentiary Status

The submission includes: *Reconstruction File* (RF), complete but uneven in depth; *Causal Chain Map* (CCM), adequate at system level, partial at feature transformation layer; *Decision Attribution Record* (DAR), clear at functional level, diffuse at vendor dependency layer; *Exception Ledger* (EL), limited historical depth due to prior version constraints.

Key gap identified: (i) incomplete traceability of composite feature construction into decision-relevant variables interpretable in legal terms; and (ii) limited attribution clarity regarding third-party scoring contribution under specific decision pathways.

Points of Deliberation

Performance vs Attribution—The Model Risk confirms improved predictive performance and stability under stress scenarios. Legal node indicates that, under contestation, the institution may not be able to articulate how certain composite features contribute to individual outcomes in a defensible manner.

Vendor Integration vs Institutional Ownership—The third-party scoring module introduces incremental predictive value. However, its internal logic remains non-transparent, and current documentation does not fully establish attributable institutional responsibility for reliance on its outputs.

Business Urgency vs Governance Integrity—Business node supports immediate deployment due to competitive pressure and portfolio impact. Risk and Legal nodes indicate that current evidentiary structure may not sustain adversarial scrutiny in all cases.

Determination

The Chamber determines that reconstructibility is sufficient for controlled deployment, but not yet robust for unrestricted use. Authorization is therefore granted conditionally, subject to the following binding requirements (*Conditions of Authorization*):

> *(a) Feature Constraint*—Composite features lacking legal translatability into contestable reasoning must be removed, or accompanied by an explicit transformation layer within the RF enabling reconstruction into domain-relevant variables
>
> *(b) Vendor Attribution Addendum*—An *External Dependency Attribution Addendum* (EDAA) must be incorporated into the RF, specifying scope and limits of third-party module, internal ownership of dependency, and justification for continued reliance under residual opacity.
>
> *(c) Configuration Snapshot Requirement*—A *Configuration Snapshot* (CS) must be generated and stored for every high-impact decision, ensuring traceability of model version, thresholds, active data sources and override conditions.
>
> *(d) Exception Capture Reinforcement*—The *Exception Ledger* (EL) must be extended to capture all manual overrides, associated justification, counterfactual reasoning, and approving authority. Accumulation thresholds to be defined as review triggers.
>
> *(e) Legal Translation Layer*—*The Reconstruction File* must include a structured articulation of how system outputs map to institutional reasoning in credit decisions, suitable for external communication.

Review Triggers

This authorization will be automatically revisited upon occurrence of any of the following: (a) increase in exception rate beyond defined threshold; (b) material modification of vendor component; (c) evidence of inability to reconstruct individual decision under internal challenge; and/or (d) supervisory inquiry or formal complaint requiring case-level reconstruction.

Attribution of Decision

The decision reflects aligned but differentiated positions: (I) *Technical Node* supports deployment under conditions; (II) *Business Node* supports immediate deployment; accepts conditional constraints; (III) *Risk Node* supports conditional authorization; flags residual opacity exposure; (IV) *Legal Node* supports authorization only under explicit constraints; reserves escalation right if evidentiary gaps persist. No dissent requiring escalation has been recorded. Divergences are reflected in conditions imposed.

Final Statement

The system is authorized as configured only to the extent that reconstructibility is preserved in operation. Performance gains do not substitute for attributable responsibility. Failure to meet the above conditions will result in suspension of authorization.

B = Refusal Memorandum / Denial of Deployment

HRAIS Chamber — Refusal Memorandum

System: SME Dynamic Pricing Engine (DPE v2.1)
Date: June 2, 2046
Decision: Authorization Refused

Context

The system dynamically adjusts credit pricing for SME clients based on real-time transaction data, behavioral clustering, and external macroeconomic signals. The model demonstrates strong backtesting performance and revenue uplift potential.

Core Deficiency

The Chamber finds that the system does not meet the minimum threshold of reconstructibility; specifically: (i) pricing outputs are driven by latent clustering structures not mappable to domain-relevant variables; (ii) no reliable transformation exists from model states to institutionally articulable reasoning; (iii) the *Reconstruction File* (RF) documents system behavior, but does not enable attribution of individual outcomes

Deliberation

Technical Node confirms model robustness and innovation value; *Business Node*: emphasizes strategic importance and competitive urgency; *Risk Node*: identifies exposure to non-transparent pricing dynamics; *Legal Node*: concludes that individual pricing decisions cannot be defended under challenge.

Determination

The Chamber determines that the system produces outputs that function as institutional decisions without providing a reconstructible basis for attributing them. Under these conditions, deployment would constitute a transfer of decision authority to the system without a corresponding retention of responsibility by the institution.

Decision

Authorization is refused.

Conditions for Re-submission

Re-submission will be considered only if: (a) pricing drivers can be expressed in contestable, domain-relevant variables; (b) a complete causal chain from input to pricing decision can be reconstructed, and (c) institutional reasoning can be articulated independently of model internals.

Statement

Innovation does not mitigate accountability requirements—a system that cannot be reconstructed cannot be authorized.

C = Case Outcome → Judicial Challenge → PCRR

Case: Credit Denial — Judicial Outcome

A retail customer challenges a credit denial issued under a prior version of the decisioning system (R-CDE v3.8). The institution presents: (i) model validation documentation; (ii) compliance with regulatory requirements; and (iii) general explanation of decision factors.

However, during proceedings, it cannot establish the exact configuration active at decision time; it cannot reconstruct the specific transformation path from input data to outcome; it cannot attribute responsibility across internal and external components.

Judicial Finding

The court does not assess the correctness of the model. It finds that the institution failed to demonstrate, in attributable terms, the basis on which the contested decision was produced. As a result, the burden of proof shifts, the decision is ruled indefensible, and liability extends beyond the individual case into systemic governance failure.

Post-Contest Reconstruction Review (PCRR)

HRAIS Chamber — PCRR Extract

Case Reference: R-CDE v3.8 Credit Denial
Trigger: Adverse judicial outcome

Failure Identification

The Chamber identifies several points of reconstructibility failure: (a) configuration loss: no reliable *Configuration Snapshot* (CS) existed for the decision instance; (b) causal discontinuity: the transformation layer between raw inputs and decision variables was not preserved in a reconstructible form; and (c) attribution diffusion: responsibility for third-party scoring input was not institutionally anchored.

Assessment

The failure did not arise from model inaccuracy or regulatory non-compliance, but from the absence of an evidentiary structure capable of sustaining attribution under challenge.

Institutional Consequence

The system remained operationally valid; it became institutionally indefensible.

Corrective Actions

The Chamber rules: (a) mandatory *Configuration Snapshot* (CS) for all decisions; (b) restructuring of RF to include full transformation traceability; (c) introduction of EDAA for all external dependencies; and (d) deployment of *Reconstructibility Drift Monitor* (RDM).

Closing Statement

The case does not demonstrate a collapse of the system, but a failure to retain control over its outcomes.

Selected References

This paper does not attempt a comprehensive review of the literature on AI explainability, fairness, or ethics; its focus is the institutional conditions under which accountability can remain operational under increasing opacity.

Bank for International Settlements, *Supervisory Challenges of Artificial Intelligence in Banking* (BIS Papers and supervisory publications).

Cohen, Julie E., *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford: Oxford University Press, 2019).

European Banking Authority, *Report on the Use of Machine Learning in Credit Institutions* (2021).

European Union, *Artificial Intelligence Act* (Regulation laying down harmonised rules on artificial intelligence, adopted 2024).

Hildebrandt, Mireille, *Law for Computer Scientists and Other Folk* (Oxford: Oxford University Press, 2020).

Pasquale, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015).

Perrow, Charles, *Normal Accidents: Living with High-Risk Technologies* (Princeton: Princeton University Press, 1984).

National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (2023).

Vaughan, Diane, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA* (Chicago: University of Chicago Press, 1996).

Internal References

García-Maceiras, JM, *The Banking Risk of AI Explanation* (ADR Notebooks No. 1, 2026). / García-Maceiras, JM, *The Five Beacons Model* (ADR Notebooks No. 2, 2026). / García-Maceiras, JM, *Tolerance for Opacity* (ADR Notebooks No. 3, 2026). / García-Maceiras, JM, *Systemic Opacity Risk* (ADR Notebooks No. 4, 2026). / García-Maceiras, JM, *The Upstream Migration of Power* (ADR Notebooks No. 5, 2026).

10.5281/zenodo.19787461