SYSTEMIC OPACITY RISK

Macroprudential Dimension of AI Governance in Banking[1]

JM García-Maceiras

President of the Spanish BPO Banking Association

Abstract

The integration of artificial intelligence into banking has intensified reliance on decision systems whose internal logic may not be fully reconstructible. Existing analyses have addressed opacity primarily at the institutional level, focusing on governance, explainability, and model oversight.

This paper introduces *Systemic Opacity Risk (SOR)* as a macroprudential dimension of AI-driven finance. SOR is defined as the risk that the aggregation and correlation of institutional opacity—even where individually contained within tolerance thresholds—may impair the financial system's collective capacity to reconstruct decisions under systemic stress.

The framework remains conceptual. It does not propose quantification or immediate regulatory intervention. Its contribution lies in identifying a dimension of systemic vulnerability linked not to the propagation of financial losses, but to the preservation of system-level intelligibility under stress.

Keywords: *Systemic Opacity Risk; collective reconstructibility; macroprudential stability; correlated opacity; supervisory architecture; systemic vulnerability; reconstructibility.*

1. Opacity and Systemic Intelligibility

The integration of artificial intelligence into banking has altered the epistemic structure through which financial decisions are produced, understood, and justified. Algorithmic systems no longer operate at the periphery of banking activity. They increasingly occupy

[1] This expanded version develops the structural and macroprudential dimensions of *Systemic Opacity Risk* introduced in the earlier conceptual paper circulated in March 2026; it builds and extends the author's prior research on AI explainability, prudential architecture, and tolerance for opacity (García-Maceiras, 2026a, 2026b, 2026c).

positions within the core decision-making architecture, informing outcomes that carry legal, prudential, and economic consequences.

These systems do not function in the absence of transparency. They function within conditions in which transparency is partial, contingent, or deferred. In many cases, the internal logic of decision-making processes remains accessible only through reconstruction, often requiring technical intervention, interpretive models, or ex post analytical effort.

The institutional response to this condition has been consistent and, to a large extent, effective. Governance frameworks have evolved to ensure that opacity remains bounded: validation procedures, documentation standards, model oversight practices, and explainability techniques have been designed to ensure that institutions retain sufficient capacity to reconstruct and justify their decisions when required.

This response, however, is structured around a specific assumption: that opacity is fundamentally an institutional phenomenon. It assumes that the relevant unit of analysis is the individual entity and that the adequacy of reconstructibility can be assessed within those boundaries. Such an assumption may hold under conditions in which opacity remains locally contained. It loses stability when opacity is distributed and structurally aligned across the system.

In such contexts, the question shifts. It is no longer sufficient to ask whether institutions can reconstruct decisions individually; it becomes necessary to examine whether reconstructive capacity, when required simultaneously across institutions, remains sufficient at the level where it becomes systemically meaningful.

The possibility explored in this paper is that, under conditions of aggregation and alignment, reconstructibility may cease to be exclusively an institutional attribute and may instead operate as a constrained system-level resource. *Systemic Opacity Risk* designates that possibility.

2. Systemic Risk and Reconstructibility

Systemic risk has historically been defined through processes of transmission. The analytical focus has been on how shocks propagate across interconnected institutions and markets, generating cascades of financial loss that exceed the capacity of individual entities to contain them. This framework remains central. However, it presupposes an underlying condition that often remains implicit: that the system retains the capacity to interpret, diagnose, and respond to those shocks in a coordinated manner.

Financial systems are not only mechanisms of allocation; they are also systems of interpretation. They produce decisions that must be understood—not only in normal operation, but under conditions of stress, when those decisions become subject to scrutiny from supervisors, courts, and market participants. Under such conditions, intelligibility is not ancillary; it operates as a functional condition. Financial systems do

not merely process risk; they also preserve the conditions under which risk remains institutionally interpretable.

Opacity, in this context, does not introduce instability *per se*. Financial systems have long operated under conditions of complexity that exceed immediate comprehension. What is analytically relevant, therefore, is whether that complexity remains navigable when it must be collectively interrogated.[2]

The increasing reliance on algorithmic systems introduces a subtle shift. Complexity is no longer only structural or institutional; it is embedded in processes whose internal logic may not be fully accessible without reconstruction. At the institutional level, this condition is managed. At the systemic level, its consequences depend on distribution and alignment.

Systemic stability may therefore depend not only on whether the system can absorb shocks, but on whether it can sustain intelligibility when those shocks require collective interpretation.[3]

3. Systemic Opacity Risk: The Concept

Systemic Opacity Risk refers to the macroprudential risk that emerges when the aggregation and correlation of institutional opacity impair the financial system's collective reconstructibility under stress.[4]

The concept does not presuppose opacity as failure. Nor does it assume that institutions operate beyond acceptable governance thresholds. On the contrary, it identifies a potential vulnerability that arises when opacity is managed locally but aligned structurally.

Aggregation reflects scale. As algorithmic systems permeate decision-making processes, the demand for reconstructibility expands correspondingly. Reconstruction ceases to be episodic and becomes structurally embedded in the functioning of the system.

[2] Contemporary literature has highlighted that artificial intelligence may simultaneously enhance efficiency and generate new forms of systemic risk, including the amplification of procyclicality, the emergence of "unknown unknowns", and the optimization of behavior against system-level constraints (Danielsson et al., 2022).

[3] The macroprudential literature has long emphasized the role of correlated exposures and conditional dependence under stress (Adrian and Brunnermeier, 2016), as well as the system-level contribution of individual institutions to aggregate risk (Acharya et al., 2012). Network-based approaches further highlight the possibility of *robust-yet-fragile* configurations in which stability under normal conditions coexists with vulnerability under stress (Gai and Kapadia, 2010). The present framework extends these intuitions to the domain of decision-process opacity.

[4] This definition does not presuppose full algorithmic transparency or complete ex ante explainability. It refers instead to the practical capacity to reconstruct and justify decision processes under conditions of stress; it emphasizes reconstructibility as a functional condition rather than a formal requirement.

Correlation reflects alignment. It denotes the extent to which opacity is not independently distributed but reflects shared constraints across institutions. These constraints may arise from similarities in technological architecture, common dependencies, or convergent governance practices.

Collective reconstructibility introduces a system-level property that is not reducible to individual capacities. It refers to the capacity of the system to sustain multiple processes of reconstruction simultaneously without degradation in coherence or interpretive reliability.[5]

The defining feature of SOR is its conditional nature. It remains latent under normal conditions; it is analytically relevant when stress generates simultaneous demand for reconstruction across institutions whose opacity structures are sufficiently aligned. In such conditions, reconstructive capacity may be constrained not at the level of individual institutions, but at the level of the system.[6]

4. Structural Conditions of Systemic Opacity

Systemic Opacity Risk emerges not from opacity alone, but from the structural conditions under which opacity becomes collectively relevant. These conditions arise through the interaction of aggregation and correlation across institutional environments increasingly shaped by shared technological and governance architectures.

4.1 Aggregation

Aggregation refers to the cumulative systemic dependence on multiple institutional reconstructive capacities. In highly digitalized financial systems, reconstructive processes are increasingly distributed across institutions simultaneously reliant on complex algorithmic infrastructures. The resulting exposure does not derive from the opacity of any individual system, but from the systemic concentration of reconstructive demand within structurally comparable environments.

4.2 Correlation

Correlation refers to the alignment of opacity structures across institutions. Such alignment may arise through common modelling practices, dependence on similar data

[5] *Reconstructibility* should not be conflated with *accountability* or *explainability*. Rather, it designates a minimal technical and epistemic condition that may enable, but does not by itself constitute, broader frameworks of *accountability*. While accountability requires decisions to be intelligible, traceable, and justifiable ex post within institutional and legal settings, reconstructibility merely preserves the possibility of such assessments without guaranteeing their normative sufficiency.

[6] Recent developments in the regulatory and supervisory treatment of artificial intelligence—particularly those emphasizing standardization, centralization of oversight, and convergence in governance practices—may be interpreted as structurally consistent with the mechanisms identified in this framework. While such developments aim to enhance institutional robustness, they may simultaneously contribute to alignment in the conditions governing reconstructibility across institutions.

environments, shared technological infrastructures, or convergence in governance standards. Under these conditions, reconstructive constraints that remain manageable in isolation may acquire systemic significance when activated concurrently.

4.3 Homogeneity

Institutional homogeneity amplifies reconstructive synchronization. Financial institutions subject to comparable incentives, supervisory expectations, and technological benchmarks may converge toward similar architectures of explainability and governance. The resulting opacity profiles need not be identical to produce correlated reconstructive limitations under stress.[7]

4.4 Infrastructure Concentration

Dependence on shared technological infrastructures may intensify systemic opacity. Cloud providers, external model components, data intermediaries, and third-party governance tools may generate latent reconstructive dependencies extending beyond institutional boundaries. In such environments, opacity becomes partially embedded within common infrastructural layers rather than exclusively within institutional systems themselves.

4.5 Governance Convergence

Supervisory harmonisation and market standardisation may unintentionally reinforce reconstructive alignment. As institutions adopt similar validation procedures, explainability frameworks, and documentation practices, the diversity of reconstructive approaches may diminish. Convergence may therefore strengthen institutional consistency while simultaneously increasing the possibility of correlated reconstructive constraints. Under sufficiently convergent conditions, opacity may cease to be institution-specific while remaining formally distributed across institutions.

The interaction of *aggregation* and *correlation* defines the structural basis of *Systemic Opacity Risk*. Opacity becomes systemically relevant not because institutions lose reconstructive capacity individually, but because multiple reconstructive limitations may emerge simultaneously within interconnected environments.

5. Reconstructive Stress

Systemic Opacity Risk remains latent unless activated under conditions of concentrated reconstructive demand. In ordinary institutional conditions, opacity may remain manageable within local governance structures. Systemic relevance emerges when reconstructive processes must be sustained concurrently across multiple institutions under stress.

[7] The role of homogeneity and correlated structures as amplifiers of systemic vulnerability is well established in macroprudential analysis. The present framework extends this intuition to the domain of decision-process opacity.

Activation does not require technological malfunction or supervisory failure. It may arise from conditions that intensify simultaneous demands for explanation, justification, traceability, or institutional accountability. Supervisory investigations coordinated market disruptions, litigation pressures, or systemic governance reviews may generate reconstructive loads exceeding the system's collective capacity to process opacity coherently.

The relevant constraint is not institutional opacity in isolation, but reconstructive throughput under systemic conditions. Opacity becomes systemically relevant not when explanation disappears, but when reconstruction loses simultaneity—that is, when multiple processes cannot be sustained concurrently. Financial systems may sustain dispersed reconstructive demand without difficulty while remaining vulnerable to synchronized justificatory pressure.

Collective reconstructibility therefore depends not only on institutional preparedness, but also on the system's capacity to sustain concurrent reconstructive operations without degradation of supervisory coherence or justificatory continuity. Reconstructibility, while institutionally implemented, is not institutionally exhausted once reconstructive demand becomes concurrent.

The activation condition is structural rather than event-specific. *Systemic Opacity Risk* concerns the possibility that reconstructive demand, when sufficiently synchronized, may exceed the coordination capacity of institutional and supervisory environments organized around partially opaque systems.

6. Distributed Reconstructibility

At the level of individual institutions, opacity is bounded through governance. *Tolerance for Opacity* (TfO) defines the limits within which reconstructibility remains viable.[8] These limits are calibrated locally; they reflect institutional capacity, legal requirements, and operational considerations. However, they do not incorporate system-wide interactions.

A structural asymmetry emerges between institutional adequacy and systemic sufficiency. Institutions may operate within well-calibrated thresholds while collectively generating conditions in which reconstructive capacity is constrained. This asymmetry reflects a familiar macroprudential dynamic. Individually rational behaviour may generate collective vulnerability.

In the context of SOR, the mechanism is not financial exposure, but epistemic alignment. The system may retain sufficient knowledge at the institutional level, yet encounter limits when required to mobilize that knowledge collectively. Reconstructibility may remain locally sufficient while becoming collectively discontinuous.

[8] The concept of *Tolerance for Opacity* (TfO) refers to the institution-specific threshold beyond which reconstructibility would become materially fragile. It is developed in the author's prior work on institutional governance of algorithmic systems.

At the institutional level, *Tolerance for Opacity* (TfO)—understood as the maximum degree of non-traceability compatible with effective reconstructibility—functions as a prudential boundary condition. It is grounded in governance capacity, documentation discipline, validation architecture, and legal defensibility. Properly calibrated, TfO preserves local reconstructive resilience.

Yet the macroprudential question does not turn on whether these thresholds exist, nor on whether they are well-calibrated in the narrow sense. It turns on what happens when thresholds are *independently* calibrated in a system whose operational substrate is increasingly aligned.

A tension emerges because reconstructibility, while institutionally implemented, is not institutionally exhausted. In other words, the capacity to reconstruct is not purely local once the demand for reconstruction is collective. The system does not confront institutions one at a time when stress occurs; it confronts them at scale, and frequently under simultaneity.

This is the conceptual space within which SOR acquires meaning. It is not a claim that institutions neglect governance; it is not a claim that explainability is absent: it is the proposition that a set of individually tolerable opacity configurations may prove collectively constraining when stress transforms reconstructibility from a bounded institutional requirement into a system-level throughput problem.

The analogy with other macroprudential dynamics is not metaphorical but structural. In liquidity risk, the system is fragile when a collectively rational preference for liquidity produces collectively illiquid outcomes; in correlated exposures, when individually prudent allocations converge. In SOR, the system attains fragility when individually adequate reconstructive architectures converge into a configuration that reduces the diversity, redundancy, and independence of reconstructive pathways.

The strongest form of this micro–macro tension is not expressed in the failure of a single institution to reconstruct; it is expressed in the *simultaneous partialness* of many reconstructions—each defensible in isolation, but collectively insufficient to sustain supervisory coherence, legal clarity, or credible system-level narratives under stress.

This is why the relevant unit of analysis cannot remain the institution alone. The institution calibrates TfO to preserve local accountability. The system requires, in addition, that accountability remains jointly operable under concurrency. SOR marks the conceptual gap between these two conditions. The system may therefore remain operationally coherent while becoming reconstructively fragmented.

7. Configurations of Reconstructive Stress

Systemic opacity is unlikely to emerge through a single explanatory rupture. Its more plausible form is the gradual synchronization of many partial reconstructive constraints.

SOR is not a risk that continuously expresses itself—it is conditional. Its activation depends on the emergence of stress conditions that demand reconstruction *at scale*. In ordinary operational contexts, opacity remains distributed and manageable: institutions reconstruct selectively, sequentially, and within internal time horizons shaped by governance protocols and resource constraints.

Under stress, these conditions invert. Reconstruction becomes externally demanded rather than internally scheduled. It becomes time-compressed rather than process-aligned, multi-institutional rather than isolated.

The critical feature of stress, in this framework, is not financial loss itself but the *structure of demand*. A systemic event—whether market-driven, supervisory, or legal—can create a requirement that decisions be rendered intelligible in parallel across institutions. That requirement is not simply an informational request; it is a stabilizing function. In highly scrutinized environments, the capacity to provide coherent reconstruction is a component of legitimacy. And legitimacy, in finance, is never purely reputational: it interacts with confidence, funding conditions, litigation dynamics, supervisory escalation, the velocity of corrective action.

The activation condition can be expressed in a simple conceptual form: opacity becomes systemically relevant when reconstructive demand is *synchronized*. Outside synchronization, opacity remains complex but tractable. Under synchronization, opacity operates as a constraint because the system's reconstructive capacity is not infinite, and—more importantly—it is not perfectly independent across institutions.

This independence is often implicitly assumed. Institutional governance frameworks may treat reconstructibility as a firm-level capability: *Can the institution reconstruct?* SOR invites a complementary question: *Can the system reconstruct when many institutions are simultaneously required to do so?*

The distinction matters because systemic stress rarely takes the form of orderly sequencing; it is shaped by concurrency: multiple supervisory requests, market interrogations, internal escalations. In such conditions, the limiting factor is not merely the existence of reconstructive procedures but their throughput under load.

It is at this point that *collective reconstructibility* acquires its intended meaning. It is not a philosophical add-on, but the operational expression of a macroprudential condition: the ability of the system to sustain intelligibility when such a fundamental requirement becomes simultaneously demanded.

The following configurations do not constitute predictive scenarios. They illustrate structural conditions under which opacity may acquire systemic relevance through synchronized reconstructive demand.

7.1 Supervisory Concurrency

Multiple institutions simultaneously become subject to intensified supervisory review following a period of market stress. Although each institution retains formally compliant governance structures, supervisory authorities encounter partially aligned reconstructive constraints across institutions operating with comparable AI-dependent decision architectures.

7.2 Shared Infrastructure Dependency

Several institutions depend on common external infrastructures involved in model deployment, data processing, or explainability support. Reconstructive limitations emerging within shared infrastructural layers generate concurrent justificatory constraints extending across otherwise independent institutions. Under such conditions, reconstructive limitations may propagate through dependency layers rather than through institutional balance-sheet exposure.

7.3 Coordinated Reconstructive Pressure

System-wide demands for explanation emerge simultaneously through litigation, regulatory scrutiny, public accountability pressures, or market disruption. Institutional reconstructive processes remain locally functional but collectively strained by synchronized justificatory intensity.

These configurations are not defined by technological collapse, but by concentrated reconstructive load within structurally correlated environments. Their analytical relevance lies in illustrating how individually tolerable opacity may acquire collective significance under systemic conditions.

8. Prudential Implications

The introduction of SOR is not a call for a new regulatory category. It is an analytical extension of macroprudential observation. The concept is intended to clarify a latent condition: that systemic resilience may depend on collective reconstructibility under stress.

This introduces an important nuance. Supervisory frameworks for AI governance are primarily firm-based. They operate through expectations around validation, explainability, documentation, governance committees, accountability lines—these remain essential. SOR does not compete with them. It complements them by identifying an exposure that cannot be fully captured through institutional compliance alone.

SOR suggests that macroprudential resilience may depend on the distribution and alignment of opacity across institutions. This raises a question of *visibility*. *Do supervisory authorities possess the informational vantage necessary to observe not only what happens within firms, but what aligns across them?*

Such visibility is not necessarily invasive. It does not require access to proprietary code. It requires mapping of shared infrastructures, common dependencies, and governance convergence patterns. In the grammar of SOR, it requires an ability to observe both aggregation and correlation.

A relevant implication concerns *reconstructive stress*. Stress testing has historically focused on capital, liquidity, and market shocks. SOR suggests that there may exist an additional stress domain: the system's capacity to sustain concurrent reconstructive demands. This does not imply formal stress-testing requirements at this stage, but it suggests that macroprudential frameworks may need to recognize that system-wide intelligibility can be strained in ways that are not reducible to loss propagation.

A related implication concerns supervisory coherence itself. If collective reconstructibility is constrained under stress, supervisory assessment may face a problem of comparability. Even when institutions provide explanations, those explanations may not be commensurable across entities if reconstructive approaches are aligned in limitations rather than in outputs. The system may become explainable in fragments but not collectively interpretable.

These implications remain observational. They do not recommend immediate intervention. They clarify where a macroprudential lens might extend without expanding prudential ambition: it would simply recognize that reconstructibility has system-level conditions, as also reflected in recent supervisory developments. The prudential relevance of opacity may ultimately depend less on what institutions know than on what financial systems remain capable of reconstructing under pressure.

9. Position within Existing Frameworks

SOR is deliberately delimited. It does not seek to replace existing risk categories or supervisory frameworks. Its function is to isolate a dimension that is structurally adjacent to them but not reducible to them. Institutional AI governance frameworks address questions of accountability, explainability, and oversight. Model risk frameworks address calibration, performance, and validation; data aggregation frameworks address reporting integrity and traceability; operational resilience frameworks address continuity of critical functions.

SOR sits orthogonally to these domains. It concerns the macro-level conditions under which reconstructibility remains jointly available when demanded collectively. This is why SOR can exist even when institutional governance functions correctly: the risk is not misgovernance but alignment. In this sense, SOR highlights a potential analytical gap: the difference between *institutional reconstructibility* and *collective reconstructibility*. The former can be secured through firm-based governance; the latter depends on systemic configuration.

This positioning matters because macroprudential analysis has historically been attentive to correlated structures even when individual firms remain prudent. SOR extends that attentiveness from exposures to epistemic conditions: from correlated

balance sheets to correlated opacity. The aim is not to add a new requirement. It is to add a new question: *what does systemic resilience mean when system-wide decisions cannot be collectively reconstructed at the pace of stress?*

10. Observational Dimensions

The conceptual nature of SOR implies restraint. Yet conceptual restraint does not entail observational emptiness. SOR can be associated with a space of observation that remains non-quantitative while still being analytically meaningful. The objective is not to measure opacity. It is to observe conditions of alignment that may increase the probability that reconstructive constraints are synchronized. Four observational domains can be articulated without converting the framework into a metric:

(i) *Infrastructure dependency mapping*—If critical layers of AI deployment are concentrated in shared providers—cloud layers, external components, data platforms—then reconstructive constraints may be co-located. This does not mean that providers are unsafe; it means that reconstructibility may depend on shared technical layers that become simultaneously relevant under stress.

(ii) *Architecture convergence visibility*—Homogeneity in modelling approaches—whether driven by talent markets, tooling ecosystems, or benchmarking norms—can create similarity in reconstructive pathways. Convergence can enhance performance and governance consistency, but it may also reduce the diversity of interpretive approaches under stress.

(iii) *Governance standardization patterns*—Explainability and documentation practices, when standardized, can improve consistency. Yet they can also define common boundaries of reconstructibility. In a system, variance is not always noise; it can be redundancy.

(iv) *Data dependency alignment*—Common data sources and shared feature engineering paradigms can produce similar opacity profiles even when models differ. Under stress, reconstructive demand may then converge on similar interpretive bottlenecks.

This observational toolkit is deliberately non-prescriptive. It is a way of expressing the core macroprudential intuition: what matters is not the opacity of one institution, but the structural alignment of opacity across the system.

11. Macroprudential Observation

The framework developed in this paper remains conceptual and does not propose immediate supervisory instruments or quantification methodologies. Its relevance lies in identifying a potential analytical dimension of systemic resilience associated with reconstructive capacity under conditions of stress.

The concept of Systemic Opacity Risk may support future analysis of how opacity is distributed across financial systems characterized by technological convergence,

infrastructural concentration, and increasingly standardized governance environments. In particular, the framework suggests the importance of observing not only institutional explainability, but also the structural alignment of reconstructive dependencies across institutions.

The analytical perspective introduced here may also contribute to broader macroprudential discussions concerning operational synchronization, institutional homogeneity, and systemic coordination capacity in highly algorithmic financial environments.

The framework does not assume that opacity should be eliminated, nor that explainability can be rendered complete. It instead proposes that collective reconstructibility may constitute an increasingly relevant condition of systemic resilience as financial systems become increasingly dependent on partially opaque decision architectures.

12. Conclusion—Systemic Intelligibility Under Stress

This paper has introduced *Systemic Opacity Risk* as a macroprudential dimension of AI integration in banking. SOR designates the risk that aggregation and correlation of institutional opacity, even when individually tolerable, may impair the financial system's collective reconstructibility under conditions of systemic stress.

The argument does not rely on technological malfunction, nor does it presuppose governance failure; it identifies instead a structural configuration: opacity acquires systemic relevance when reconstructive constraints are aligned and reconstructive demand is synchronized.

Systemic stability has long been associated with capital, liquidity, and interconnectedness. In increasingly algorithmic environments, stability may also depend on a less visible condition: whether the system can sustain intelligibility when it is most exposed to scrutiny.

Opacity, when locally managed, remains a governance challenge. Opacity, when systemically aligned, may operate as a macroprudential variable. The difference between the two is not intensity but configuration.

The *Five Beacons Model* defines responsibility; *Tolerance for Opacity* sets institutional limits; *Systemic Opacity Risk* marks the point at which individually tolerable opacity acquires collective relevance.

The financial system has long asked what happens when it cannot absorb losses. SOR introduces a complementary question: *what happens when, under stress, it cannot reconstruct what it has done*.

The framework therefore raises a broader prudential question: whether future systemic resilience may depend not only on the capacity to absorb shocks, but also on the capacity to sustain intelligibility under synchronized scrutiny.

[V.02] La Coruña, May 3rd, 2026

Bibliography / References

Acharya, V.V., Engle, R., Richardson, M., 2012. *Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks*. American Economic Review 102(3), 59–64.

Adrian, T., Brunnermeier, M.K., 2016. *CoVaR*. American Economic Review 106(7), 1705-1722.

Basel Committee on Banking Supervision (BCBS). *Principles for Effective Risk Data Aggregation and Risk Reporting (BCBS 239).* Basel, 2013.

Danielsson, J., Macrae, R., Uthemann, A., 2022. *Artificial intelligence and systemic risk*. Journal of Banking & Finance 140, 106290. https://doi.org/10.1016/j.jbankfin.2021.106290

European Banking Authority (EBA). *Discussion Paper on Machine Learning for IRB Models.* EBA/DP/2021/02.

Federal Reserve Board. *Supervisory Guidance on Model Risk Management (SR 11-7).* 2011.

Financial Stability Board (FSB), 2017. *Artificial intelligence and machine learning in financial services.*

Financial Stability Board (FSB), 2024. *The Financial Stability Implications of Artificial Intelligence*. FSB Report, 14 November.

Gai, P., Kapadia, S., 2010. *Contagion in Financial Networks*. Proceedings of the Royal Society A 466(2120), 2401–2423.

García-Maceiras, J.M., 2026a. *The Banking Risk of AI Explanation. ADR Notebooks No. 1. Zyphrum Alchemists.* ISBN 9789403845760

García-Maceiras, J.M., 2026b. *The Five Beacons Model: A Prudential Architecture for AI Explainability and Legal Liability in Banking.* DOI 10.5281/zenodo.18647317

García-Maceiras, J.M., 2026c. *Tolerance for Opacity: A Threshold Framework for AI-Driven Banking.* DOI 10.5281/zenodo.19144304.

Lipton, Zachary C. "The Mythos of Model Interpretability." *Queue* 16, no. 3 (2018).

Rudin, Cynthia. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1 (2019): 206–215.

10.5281/zenodo.19148608