

The Loaded Wheel: Why the First AI Crisis Will Be Governed by Finance, and What That Does to AI Governance

Murad Farzulla*

Dissensus, London, UK & King's College London, UK

*Correspondence: murad@dissensus.ai ORCID: 0009-0002-7164-8704

July 2026

Abstract

The argument that only a crisis will break the deadlock in AI governance treats the crisis itself as a roulette wheel: exogenous in kind, unknowable in advance. I argue the wheel is loaded — not in its timing, which is genuinely beyond reach, but in its governor. Focusing-event theory already privileges three properties on which a candidate crisis either opens a policy window or does not: legibility of harm (the framing contest of Boin, 't Hart, and McConnell), speed of attribution (the window decay of Kingdon), and maturity of the receiving regulator (Birkland's receiving community, Baumgartner and Jones's venue capture). Scored on these axes, a financial cascade does not dominate — autonomous-weapons and physical-harm events are at least as legible and as fast to attribute — but finance wins the combination, decisively on the third axis: it is the one domain that pairs a legible, quickly-attributable failure with a regulator holding exercised emergency powers over a continuously operating, un-groundable, internationally networked system, and whose window opens in public rather than inside classification. So whatever domain the first legible AI-involved systemic event originates in, the modal outcome — and the claim is modal throughout: a plausibility ranking over *governors* strong enough to reallocate preparation portfolios, not a point forecast over domains — is that it routes to financial regulators, who produce governance faster than any other receiver could. I stress-test the ranking against out-of-domain history and against the objection that finance's own control layer contains the cascade; its sharpest near-term test is the cyber-finance hybrid, arguably the most probable first event of all, where the framework predicts financial regulators absorbing an AI-enabled security failure and answering it with market-structure rules that miss the mechanism. Because jurisdiction follows the corpse, the response will be fast and mis-scoped: sectoral, keyed to the legible failure, and structurally unable to reach the model layer the AI-governance community wants governed — financial mandates stop at regulated entities, and where they now reach the technology providers behind them, they reach them as providers of services, never as builders of models — and a financial crisis is structurally unlikely to reopen the question of who should govern the model as a model, because extending a tool a regulator already holds is fast where reaching for a mandate it lacks is not. It will import the documented pathology of crisis-born financial regulation — compliance ratchets that concentrate the industry, perimeters that push activity into the dark, assurance regimes that get gamed. I derive design principles for the pre-crisis shelf (general charters on sectoral vehicles, exposure measurement over rules, migration-proof monitoring) and price four borrowings from financial regulatory machinery against their own pathologies. The one failure mode the whole argument cannot reach is the one that never becomes legible: slow correlated drift that opens no window and routes nowhere.

Keywords: AI governance, focusing events, financial regulation, crisis policy, agenda-setting, regulatory capture, systemic risk, algorithmic monoculture, warning shots

Research Context

This work forms part of the Adversarial Systems Research program at [Dissensus](#), investigating stability, alignment, and friction dynamics in complex systems where competing interests generate structural conflict.

This paper extends the program’s political-governance lane into crisis political economy: it models the institutional-capture dynamics that determine which regulator inherits the first legible AI-involved systemic failure, and what that capture does to the prospects for general-purpose AI oversight.

Acknowledgements

This paper grew out of the Frontier AI Governance course run by BlueDot Impact; the author thanks the course peers whose discussions helped generate the idea. The argument also benefited from discussion during the Cooperative AI Foundation’s 2026 summer course; particular thanks to Chris Leong and Aishwarya Gurung, the author’s course leads there. Automated referee passes from [reviewer3.com](#) and [openaireview.com](#) caught several internal inconsistencies corrected in this version, and citation verification drew on [scite](#).

This paper benefited from extended collaboration with Claude (Anthropic; Claude Fable 5), whose contributions to literature synthesis, source verification, and iterative refinement were substantive, and from adversarial cross-review by ChatGPT (OpenAI; GPT-5.6 “Sol”), alongside other automated research tools. The author gratefully acknowledges this assistance while taking full responsibility for all claims, errors, and interpretive choices.

This work is part of the Adversarial Systems Research program at Dissensus, a broader investigation into stability, alignment, and friction dynamics across political, financial, cognitive, and multi-agent systems. Related papers in the series are available through the Adversarial Systems & Complexity Research Initiative ([ASCRI](#); [systems.ac](#)).

The author welcomes feedback, criticism, and collaboration. Correspondence should be directed to murad@dissensus.ai.

1 Introduction

The most developed recent statement of the case that AI governance is deadlocked ends on a strange kind of hope: that a crisis will do what diplomacy cannot. Chatham House’s *Breaking the Deadlock on AI Governance* (Wilkinson, Krasodomski, Wilkinson, and Varela Sandoval, 2026) surveys the summit series, the voluntary commitments, and the institutional fragments, concludes that none of them will produce binding coordination under current incentives, and turns to the historical record — 2008, WannaCry, COVID — for evidence that focusing events open windows that nothing else opens. Its recommendations are deliberately modular, designed to survive whatever crisis arrives, because the report declines to say what kind of crisis that will be. A “financial system failure, autonomous weapons misuse or the loss of control” are named in one breath, candidates without a ranking between them. The genre is wider than one report: Chernobyl-moment commentary likewise names candidate domains and a backlash dynamic without supplying a selection mechanism (Egan, 2025).

That under-commitment is what kept bothering me, and it is deliberate: Chatham House names conducive sectors and builds modular recommendations precisely because it declines to rank. The deadlock literature treats the focusing event as a roulette wheel whose timing and domain are equally unknowable, so that preparation must be domain-neutral. Timing is genuinely beyond anyone’s reach. The domain is the part I think yields to analysis. Focusing events do not select uniformly across candidate domains; they select for properties — how legible the harm is, how fast it can be attributed, how ready the receiving institutions are to claim it — and those properties are distributed very unevenly across the places where AI systems are currently acquiring real authority. Rank the candidates and no domain leads on every axis — autonomous-weapons and physical-harm events are as legible, and as fast to attribute, as anything finance produces. What finance wins is the *combination*: it is the only domain that scores high on all three axes at once, with a receiving regulator whose window opens in public and whose emergency authority runs over a system that cannot be switched off. The wheel is loaded, and it is loaded toward finance — not because finance owns any single axis, but because it is where the three coincide.

This is a governance-side claim, and I want to keep it distinct from a stability-side claim that already has an owner. Danielsson and Uthemann (2025) argue that AI makes financial crises faster, more severe, and harder to supervise; Danielsson, Macrae, and Uthemann (2022) laid out the channel list before the current model generation existed. I take that work as given. My question sits one step downstream, on the governance side: what happens to *AI governance as a whole* when the first legible AI catastrophe arrives wearing a ticker symbol, when the crisis, so to speak, chooses its governor.

The answer I defend is uncomfortable for the deadlock literature, because it splits the hope in half. A crisis routed to finance — whatever domain it starts in — does produce governance, and produces it fast, faster than any other receiver could. It also produces it mis-scoped. Jurisdiction follows the corpse: the response is owned overnight by the securities and banking regulators, routed internationally through pipes that already exist, and assembled from whatever drafts are already on the shelf, filtered by existing statutory authority. The model layer — the thing the AI governance community actually wants governed — stays largely outside the perimeter, because financial mandates are entity-scoped, the provider-layer regimes that do exist reach AI vendors only as technology service providers (Section 3.2), and the crisis response is structurally biased against reopening the question of who should govern the model as a model. The claim is a strong default, not an impossibility: a crisis whose public narrative centered on a named model from its first hours, insistently enough for legislative pressure to outrun the sectoral response, could force the mandate question open — the EU’s Article 93 proves the mandate can be written. But the default is loaded the other way, because window decay favors the fast path, and

the fast path runs through tools already held (Section 3.2 states the mechanism). Worse, the response imports a cost pathology that financial regulation has documented on itself for forty years: compliance ratchets that consolidate the regulated industry, perimeter rules that migrate activity into unobserved spaces, and assurance regimes that get gamed into false comfort. The skeptical pole of the warning-shot debate holds that warning shots would change little (Soares, 2022); the safety community’s own research agenda treats shots as open questions to prepare for generically (Barnett and Scher, 2025). Both, I think, test the wrong outcome. Warning shots produce a great deal of governance; it is fast, narrow, and sectoral, and the question that matters is what that narrowness costs and what can be smuggled aboard.

Four moves organize the paper. First, it builds a selection framework from machinery the policy literature already owns (Kingdon, 2011, and Birkland, 1997, 1998, supply windows and event characteristics) and uses it to rank candidate AI crises before the fact. Running focusing-event machinery forward is not itself new — DeLeo’s anticipatory-policymaking program has been doing exactly that since 2010, on pandemic influenza, nanotechnology, and climate (DeLeo, 2010, 2016), with the event concept itself still being refined inside the same collaboration (DeLeo, Taylor, Crow, and Birkland, 2021) — and this paper works inside that tradition rather than founding it. What I can find no published version of is the narrower object here: an ordinal, coded-rubric ranking over the *governors* of candidate crises — which institution receives the event — for AI or for anything else. DeLeo’s program anticipates threats one at a time and asks how they move through the policy process, not which institution captures the response; Schiff (2023) applies Kingdon to US AI agenda-setting descriptively, with the window exogenous; and Chatham House gestures at “sectors where technical expertise has established legitimacy, such as finance, cyber security” as conducive terrain (Wilkinson et al., 2026) while declining to rank. The framework sharpens that conduciveness observation into a ranked prediction with a mechanism. Second, it ports the documented cost pathology of crisis-born financial regulation forward as a predictive template for crisis-born AI regulation; Chatham House treats the 2008 response as its success template, and the critiques it directs at that template are equity critiques, so the cost side arrives unpriced. Third, it states a design principle I call riders on vehicles: chartering general-purpose capability on sectoral institutions at design time, before the sectoral crisis that creates them. The principle appears in prior work only as a single undeveloped sentence about extending FSB crisis arrangements (Wilkinson et al., 2026) and as institutional borrowings that use sectoral bodies as templates rather than as predicted capture points (Trager, Harack, Reuel, and others, 2023). Fourth, it proposes an ex-ante exposure registry, where the entire existing incident-reporting stream is retrospective and the official sector has conceded in print that it cannot answer the forward-looking exposure question (Financial Stability Board, 2025b). A converging research current (incident regimes, monoculture measurement, Basel-for-AI mappings) is visibly closing in on adjacent territory, which is an argument for writing this down now rather than for the novelty lasting.

From there the structure follows the argument. Section 2 builds the selection theory, runs the candidates through it, and stress-tests the result against out-of-domain history and its strongest internal objection. Section 3 describes the receiving apparatus: what overnight absorption by financial regulators actually looks like, item by shelf item. Section 4 presents the three bills — the cost pathology the sectoral response imports — and what does not get built. Section 5 derives design principles for the pre-crisis shelf and prices four pieces of borrowable financial machinery against the pathologies of Section 4. Section 6 takes objections, including the one I cannot fully answer. Section 7 concludes. Throughout, the stance is the one Collingridge (1980) forced on technology governance and Marchant, Allenby, and Herkert (2011) formalized as the pacing problem: control is cheap when knowledge is scarce and expensive when knowledge arrives, so the rational strategy is to prepare instruments before

impacts are clear and deploy them when the window opens. Finance turns out to be the one domain where the pacing problem has been partially solved, which is exactly why it will get there first, and why “fast” and “well-scoped” will come apart.

2 A Selection Theory of Focusing Events

2.1 The machinery, and the axes it supplies

Kingdon’s multiple-streams account gives the canonical picture: problems, policies, and politics flow separately, and policy change happens when an open window couples a recognized problem to a pre-existing solution (Kingdon, 2011). Birkland supplied the events: sudden, harmful, rare, concentrated occurrences that force problems onto agendas that interest-group politics had kept them off (Birkland, 1997), with the follow-on finding that the *characteristics* of the event and the composition of the receiving policy community condition what the event achieves (Birkland, 1998). Baumgartner and Jones (2009) add the venue: punctuated equilibrium is partly a story about which institution captures the issue when the punctuation comes, because the capturing venue’s image of the problem becomes the problem. And Boin, ’t Hart, and McConnell (2009) add the contest: crises produce policy change only when a causal frame wins — severity, agency, and responsibility all have to be settled in public argument before anything moves.

These four literatures do not just supply the machinery; they supply the axes, and it matters that the axes fall out of the theory rather than being chosen for the conclusion. Read the focusing-event mechanism as a sequence, and each stage contributes exactly one degree of freedom. First, an event has to *read* as a problem: Boin, ’t Hart, and McConnell (2009) make the framing contest the gate — a crisis moves policy only when a causal frame wins in public, and a frame wins in proportion to how little framing the harm needs. That is legibility, and it is the first axis because it is Boin’s gate. Second, the reading has to arrive before the window shuts: Kingdon (2011) builds the whole account on windows that open and close, so a solution couples to a problem only inside a decaying interval, which makes the operative quantity not attribution as such but attribution *speed relative to window decay*. That is the second axis, and it is Kingdon’s clock. Third, whatever passes the first two stages is captured by an institution, and the institution determines the result: Birkland (1998) shows the receiving policy community conditions what the event achieves, and Baumgartner and Jones (2009) show the capturing venue’s image of the problem becomes the problem. That is receiver maturity, the third axis, and it is Birkland-and-Baumgartner-Jones in institutional form. Three stages, three axes, one per stage — legibility to pass the framing contest, attribution speed to beat the window, receiver maturity to build inside it. A different axis set is available — irreversibility of harm, breadth of the affected population, novelty of the causal pathway — and it would rank the candidates differently, biological and autonomous-weapons events first. But those are properties of how *bad* a crisis is, not of whether it *focuses*; they belong to a theory of which crises are worst, which is a different question from the one focusing-event theory answers and the one this paper asks. The axes here are the ones a theory of *focusing* privileges, and the ranking that follows is a prediction from that theory, not a rationalization built to reach finance.

Read together, these four literatures already contain everything needed to predict which AI crisis comes first. The forward stance is DeLeo’s (2010, 2016); the comparative ranking over candidate AI crises and their receivers is, as far as I can find, unrun. The synthesis is a three-axis selection filter. A candidate crisis becomes *the* focusing event to the extent that it scores on:

1. **Legibility of harm.** Can the damage be stated in a unit the public and the legislature already price?

Boin et al.'s framing contest is winnable in proportion to how little framing the harm needs.

2. **Speed of attribution.** How fast can the harm be causally connected to an AI system with enough confidence to act on? The operative quantity is time to an *actionable mechanism hypothesis*, a much lower bar than settled attribution and a far lower bar than criminal-trial attribution. Windows decay (Kingdon, 2011; Coffee, 2012); attribution that outlives the window produces nothing.
3. **Maturity of the receiving regulator.** Is there an institution with existing statutory authority, crisis experience, and international coordination pipes standing where the corpse lands? Birkland (1998) in institutional form: the policy community that receives the event determines what the event can build.

Two scope notes before the candidates, because the framework's claim structure matters as much as its ranking. The axes are conditional machinery: the first two predict whether an event of a given kind opens a window, the third predicts what gets built inside one. Nothing in the framework predicts that any particular event occurs, so the ranking of Section 2.2 multiplies in a prior over where AI systems actually hold authority, which Section 2.3 argues qualitatively rather than formally. And the axes are not fully independent: Section 2.2's own best case shows a mature regulator acting on a provisional mechanism hypothesis, which means axis 3 can substitute for axis 2. That substitution is itself part of the finance story, and I flag it now so it does not arrive later as a surprise.

2.2 Running the candidates

The ranking runs in prose below, but prose hides degrees of freedom, so here it is as an explicit, ordinaly-scored table with the rubric published. The table is a consistency check run against events the framework was not built from, not a validation; its job is to show that the axes discriminate, that finance does not sweep them, and that finance nonetheless wins the combination.

The rubric. Ordinal, three levels per axis, plus one flag. No cardinal scores — a false-precision number is worse than an honest ordinal.

- L — Legibility of harm.** *High:* harm arrives pre-denominated in a unit already priced (dollars, dead institutions, bodies), no framing contest needed. *Medium:* harm visible but totals contested or disclosed late/adversarially. *Low:* the event's existence or severity is itself a contested claim.
- A — Attribution speed** (to an *actionable mechanism hypothesis*, not settled or criminal attribution). *High:* days–weeks. *Medium:* months. *Low:* months–years, or the counterfactual stays contested indefinitely.
- R — Receiver maturity.** *High:* a standing institution with existing statutory authority, exercised emergency powers, and standing international pipes, sitting where the corpse lands. *Medium:* a competent rulemaker, or a dormant/young authority, running on a multi-year or statute clock. *Low:* a conflicted, absent, or immature receiver.
- P — Receiver publicness** (sub-axis, flag). Does the window open in public? *No* = the receiver processes the event inside classification or secrecy, so no public framing contest occurs.

The fourth column is a sub-axis, not a fourth axis, and I add it for a specific reason: the autonomous-weapons row scores high on all three primary axes, and without a publicness distinction the framework would predict a public governance window it does not in fact predict. Rather than quietly retune the

primary axes to exclude the case, I record the sub-axis the case forces. It refines axis 3 — a mature receiver that opens no public window is still a mature receiver — and it is where finance and autonomous weapons, otherwise the two strongest candidates, come apart.

One further discriminator is carried in the prose rather than the columns: the *continuously-operating, un-groundable system*. A fleet, a plant, or a pipeline can be grounded, which *removes* the urgency that drives crisis-speed rulemaking (the 737 MAX was grounded, and its rulemaking then ran two years); markets reopen every morning whether or not the mechanism is understood, which is why finance’s emergency authority converts into durable fast rules where aviation’s and nuclear’s do not. This is the property that turns a High-R into a fast-durable-rule outcome.

The table. For historical cases the outcome column records whether the event produced a *fast, durable, sectoral rule* (Y / Partial / N); for candidate AI crises it records the framework’s prediction.

#	Event (year)	L	A	R	P	Outcome / prediction
<i>Historical — retrodictive</i>						
1	Knight Capital (2012)	H	H	H	Y	Y — 15c3-5 enforcement, market-access tightening
2	Flash Crash (2010)	H	H (mech) / L (indiv)	H	Y	Y — circuit breakers in 35 days, then LULD; axis 3 substituting for axis 2
3	Silicon Valley Bank (2023)	H	H	H	Y	Y — weekend resolution + systemic-risk exception; 7-week postmortem
4	Global financial crisis (2008)	H	M	H	Y	Y — Dodd-Frank (22 mo), Basel III; the template
5	Boeing 737 MAX (2018–19)	H	H	M	Y	Partial — emergency grounding in <i>days</i> , but recertification + statute ~2 yr; <i>groundable</i> system removed the urgency
6	WannaCry (2017)	L	L	L	Y	N — advisories, voluntary frameworks; no window of consequence
7	Deepwater Horizon (2010)	H	H	L	Y	N — window consumed by agency reorganization; no major spill statute
8	Three Mile Island (1979)	M	M	M	Y	Partial/slow — NRC reforms + a de facto multi-decade construction freeze; ossification, not a fast rule; not continuously priced
9	Colonial Pipeline (2021)	M	M	M (dormant TSA authority)	Y	Partial — TSA security directives via pre-existing but unexercised authority; thin
10	Equifax (2017)	M	H	M (FTC/CFPB, no breach emergency power)	Y	N/slow — FTC settlement ~2 yr; no federal breach statute
11	Northeast blackout (2003)	H	H	M (NERC voluntary → statutory)	Y	Partial — EPA 2005 made reliability standards mandatory, on a ~2-yr <i>statute</i> clock
<i>Candidate AI crises — predictive</i>						
12	Biological misuse	L–M	L	M (biosecurity, slow)	Y	Slow window; routes to biosecurity institutions, not AI governance
13	Autonomous weapons	H	H (private)	H	N	Narrowest, least accountable absorption; window opens inside classification
14	Disinformation / manipulation	L	L	L	N*	Framing contest never resolves; no cross-cutting window
15	Cyber-finance hybrid	H	L–M	H	Y	Routed to financial regulators, who write market-structure rules for a security failure (the panic-test failure of Section 5.3)
16	Financial AI cascade (modal)	H	H–M	H	Y	Fast sectoral governance, mis-scoped; the paper’s central case

#	Event (year)	L	A	R	P	Outcome / prediction
17	Physical/robotic harm	H	H	M	Y	Partial — the groundable system converts receiver advantage into a recertification clock; the MAX pattern
18	Pure cyber (AI-enabled)	L	L	L	Y	N — victims conceal the harm; the WannaCry pattern; no window of consequence

*P is binary by rubric; row 14 is coded N because no mature receiver exists to open a window, public or otherwise — the fragmented receiver landscape (platforms, election bodies) holds no cross-cutting rulemaking authority. The contestedness of the harm itself is scored where the rubric puts it, under L.

Read it honestly, and it says four things. *On legibility alone, finance has no unique advantage*: rows 1–5, 7, 11, 13, 15, 16, and 17 all score H, because bodies (autonomous weapons, the MAX, Deepwater) are as legible as dollars and the oldest unit of harm is not a ticker. (Rows 17 and 18 code the two candidates Section 2.2’s prose runs as near-miss and structural-failure cases — physical/robotic harm and the pure cyber event — so the table and the prose cover the same candidate set; their codes are read off the historical analogs in rows 5 and 6.) *On attribution speed alone, finance ties*: the MAX, Deepwater, and autonomous weapons all attribute in days–weeks; finance is fast, not fastest. *Axis 3 is where finance separates — and even there it is not alone*: autonomous weapons (row 13) score H on receiver maturity too, because the national-security executive holds emergency powers and alliance pipes, and what separates them is **P**, the weapons window opening inside classification so that the public framing contest the deadlock literature’s hope requires never occurs; the MAX (row 5) and the blackout (row 11) show a mature-enough receiver whose system is *groundable* or whose authority runs on a statute clock, producing partial or slow governance rather than the finance pattern. *The stress cases behave as the framework predicts*: Deepwater (H, H, L) is the cleanest, maximal legibility and fast attribution opening a window that an immature receiver builds nothing durable in — axis 3 doing exactly the work the theory assigns it — and Equifax and TMI show the same axis-3 drag from different directions.

Finance wins the *combination*. Rows 3, 1, 2, and 16 occupy the unique cell: no axis below High at the mechanism grain — row 2’s individual-level attribution ran to years and row 16 carries the same split as H–M, a grain distinction Section 5.2 returns to — public, over a continuously-operating ungroundable system. Autonomous weapons is the near-miss that fails P; physical harm is the near-miss whose groundable system converts R to Partial. That is the honest margin, stated in place of any cleaner claim: finance does not win any single axis outright, it wins the *combination* — the only public, all-three-axes-High cell over a system that cannot be switched off — carried there by the third axis and the publicness sub-axis, not by a lead on legibility or speed. The deployment prior (Section 2.3) adds a hedged reason to think finance is also the modal *origin*, but the ranking does not need it: the combination is a claim about *reception*, and it holds whatever domain the crisis starts in.

Biological misuse fails on attribution speed. An engineered-pathogen event may be the highest-consequence candidate on any list, but the causal chain from a model output to a release to an outbreak runs through epidemiological attribution that takes months to years even when the pathogen is natural, and the counterfactual — could this have been done without the model? — stays contested indefinitely. The harm is legible in the worst case and illegible in every intermediate one. A window may open; it opens slowly, and it opens for biosecurity institutions rather than for AI governance.

Cyber fails on legibility, for a structural reason: the victims conceal the harm. Breach disclosure is adversarial to the discloser’s stock price, insurance position, and litigation exposure, so the observable record systematically understates the damage, and what surfaces does so late and lawyered. The Belfer study of cyber incident investigation exists precisely because market pressure suppresses the disclosure

an investigation would need (Knake et al., 2021). An AI-enabled cyber event of any scale would be fought over — attribution contested between states, harm figures contested between firms and insurers — for exactly the duration a policy window cannot survive.

Disinformation and manipulation fail on contested harm, which is Boin et al.’s framing contest in its purest form: the event’s very existence is a partisan claim. Whatever one’s assessment of the underlying damage, an event whose severity, agency, and responsibility are all contested along pre-existing political lines cannot win the framing contest fast enough to open a cross-cutting window.

Physical and robotic harm scores well on legibility, since bodies are the oldest unit of harm, but current deployment gives autonomous physical systems narrow authority in controlled environments, and the receiving regulators (transport, aviation, product safety) are competent rulemaking bodies whose durable responses run on multi-year clocks. Section 2.4 examines the strongest apparent counterexample, the 737 MAX groundings, in detail; the emergency reflex exists in these sectors, and the crisis-speed rulemaking that follows it does not.

Autonomous weapons, named in Chatham House’s own candidate list, deserve their own entry rather than absorption into the physical-harm bucket, because the axes score them strangely. An autonomous targeting failure that kills allied civilians maxes legibility, and attribution can be fast, because the operator owns the logs and the telemetry. But the attribution is fast and *private*. The receiving apparatus is the national-security executive: it has emergency powers, it has standing international coordination through alliance structures, and it also has classification. The framing contest that Boin, ’t Hart, and McConnell (2009) put at the center of crisis politics never occurs in public; the response is doctrine and rules-of-engagement revision inside the security state, and it generalizes to civilian AI governance not at all. On the framework’s own terms an autonomous-weapons first crisis is entirely possible, and it would produce the narrowest and least accountable absorption of any candidate: sectoral capture with the additional property that the sector is secret. That outcome strengthens rather than threatens this paper’s thesis, but it deserves stating plainly, because the deadlock literature’s hope requires a *public* window, and a crisis processed inside classification opens none.

The cyber-finance hybrid is the candidate that a mutually-exclusive-domains framing hides, and arguably the most probable near-term scenario of all: an AI-enabled manipulation attack, a compromised settlement layer, adversarial exploitation of correlated agents. The axes score it in conflict rather than in chorus. Losses print in public prices, so the cyber dismissal (victims conceal the harm) fails; attribution is adversarial and possibly state-contested, so the finance attribution story fails too; and the corpse still lands on a mature regulator. The framework then makes a sharp and uncomfortable prediction: a hybrid event gets absorbed by financial regulators acting on the legible surface, who write market-structure rules for what is actually a security failure. That is the panic-test failure of Section 5.3 in its purest form, and I flag it here because it shows the axes doing independent work: a candidate can max one axis, flunk another, and the framework still says what happens next. The absorption claim is about rule-writing, not response coordination, and the distinction matters because hybrid events do trigger joint machinery — CISA, Treasury, the FBI, and the financial regulators would share the incident response, as the standing playbooks for cyber events in critical infrastructure require. But task forces coordinate the response; they do not redistribute rulemaking authority. When the rules get written, each agency writes inside its own mandate, and the agency holding the legible harm — printed losses, halted markets — writes first and widest. Coordination changes who is in the room during the crisis. It does not change whose pen moves after it.

Finance scores high on all three axes — the table shows the near-misses do too, which is why the details carry the argument. Each axis, concretely:

Legibility. Financial harm arrives pre-denominated. When Knight Capital’s dormant order-routing code activated on the morning of August 1, 2012, it generated roughly four million executions in about forty-five minutes and a loss the SEC’s order states as more than \$460 million (SEC, 2013). There was no framing contest to win; the harm arrived as a number attached to a firm on a particular morning. Silicon Valley Bank announced a \$1.8 billion after-tax loss on the sale of its available-for-sale portfolio on March 8, 2023; depositors initiated \$42 billion in withdrawals on March 9, and the bank was seized on March 10 with a negative cash balance of roughly \$958 million (California DFPI, 2023). The Federal Reserve’s postmortem records over \$40 billion out the door on March 9 with a further \$100 billion expected on March 10 (Barr, 2023). Announcement to death: forty-eight hours, every step quantified in real time on cable news. Dead banks are the most legible institutional corpses modern politics produces.

Attribution speed. The May 6, 2010 flash crash shows both faces of this axis, plus the caveat the framework has to own. Mechanism-level attribution was fast: within days the event was located in specific markets, specific order books, and specific automated strategies, and the joint CFTC–SEC staff report reconstructing it (a single \$4.1 billion, 75,000-contract automated sell program executed in twenty minutes into a market where broad indices, already down about four percent, fell a further five to six percent in minutes) arrived in five months (CFTC and SEC Staffs, 2010). Individual attribution was slow: Navinder Sarao was charged in February 2015, pleaded guilty in November 2016, and was sentenced in January 2020, 9.7 years after the event (DOJ, 2016; CNBC, 2020), and the CFTC’s civil order quantifying his layering (over four hours of activity, more than 81,000 order modifications, twenty to twenty-nine percent of the visible sell-side book) landed a \$38 million sanction (CFTC, 2016) against a disruption whose intraday drawdown media accounts put near one trillion dollars (CNBC, 2020). Policy windows run on mechanism attribution, and the mechanism attribution took days to weeks. Now the caveat: the SEC approved stock-by-stock circuit breakers thirty-five days after the crash, with exchanges implementing the next day (SEC, 2010b), four months *before* the staff report settled the mechanism, and the circuit-breaker pilot was succeeded by the limit up-limit down mechanism, approved on May 31, 2012 and phased in from April 2013 (SEC, 2012a). The fast rule preceded settled attribution. What did the work was a mature regulator’s willingness to act on a provisional mechanism hypothesis: axis 3 substituting for axis 2, exactly the substitution Section 2.1 flagged. By the panic test of Section 5.3, the 2010 circuit breakers were adopted pre-mechanism and happen to pass anyway, because they bound cascade speed regardless of cause; the sequencing matters because it means “fast” and “correctly scoped” came apart even in the framework’s best example. No other domain converts an anomaly into an enforced rule on that clock.

Regulator maturity. Section 3 develops this at length. The preview: financial regulators hold pre-authorized emergency powers over continuously operating markets and have exercised them at crisis speed — an overnight short-selling ban across roughly 799 financial stocks in September 2008 (SEC, 2008), a full bank resolution with a systemic-risk exception executed between a Friday failure and a Sunday statement (U.S. Treasury, Federal Reserve, and FDIC, 2023) — and they sit inside standing international coordination through the FSB, BIS, and the supervisory network, which have been running AI workstreams since before any crisis (Financial Stability Board, 2024; BIS, 2024). Other regulators hold and have used emergency powers too; Section 2.4 discusses the aviation case. What no other sector combines is exercised emergency authority, a continuously operating and continuously priced system to use it on, and pre-built international pipes.

2.3 Why finance is plausibly also the origin (a weaker, supporting claim)

The ranking so far is about *reception* — where a legible crisis routes and how fast it is processed — and it stands whatever domain the crisis originates in. There is a weaker, supporting claim that finance is also the modal *origin*, and because it is weaker I want to mark its evidence carefully rather than lean on it. The prior also has a prominent public owner, and the credit belongs on the record: in October 2023 the sitting SEC chair told the Financial Times that an AI-driven financial crisis was “nearly unavoidable” without new regulation, and placed the difficulty at the exact seam this paper builds on — regulation written institution by institution against a risk that is horizontal, “many institutions might be relying on the same underlying base model or underlying data aggregator” (The Hill, 2023). Gensler said it as the regulator; what follows is the argument for why he was pointing at the right domain. Autonomous execution is normalized in markets in a way it is nowhere else: no compliance officer blinks at software placing orders without human review, because software has been doing so since before most compliance careers began. The Bank of England’s Financial Policy Committee, surveying deployment, found 55 percent of surveyed AI use cases already involving some degree of autonomous decision-making, two percent fully autonomous (Bank of England FPC, 2025). That figure evidences autonomy *within* finance; it is not a cross-domain census, and I do not have one — the comparative claim that agents acquire real authority in markets *before* they acquire it in healthcare, logistics, or defense rests on the normalization gradient as an argument, not a measurement. The gradient can at least be proxied, sector by sector, and the proxies all point the same way. In markets, autonomous execution is the default operating mode: the SEC’s congressionally mandated review opens from the premise that “the use of algorithms in trading is pervasive in today’s markets” (SEC, 2020), and the BIS documents execution algorithms as an established fixture of FX market functioning, with usage still climbing (BIS Markets Committee, 2020). In medicine, the comparable frontier is a handful of narrow authorizations: the FDA’s first clearance of a diagnostic system permitted to render a screening decision without clinician interpretation came in 2018, for a single indication (FDA, 2018). In defense, autonomy in the use of force is doctrinally gated: DoD Directive 3000.09 requires that systems “allow commanders and operators to exercise appropriate levels of human judgment over the use of force” as a design condition (U.S. Department of Defense, 2023). These are proxies, not the census — deployment breadth, capital under autonomous authority, and decision frequency would all be needed for the real measurement — but the asymmetry they sketch is not close. Taken at that weight, it is a prior about where the agents are, and it amplifies the reception ranking rather than carrying it: if finance is also the modal origin, the wheel is loaded on both the origination and the reception axes, but the mis-scoping payload follows from reception alone.

And the failure mode that authority enables is not hypothetical; it has a theory and a growing empirical record. Kleinberg and Raghavan (2021) proved that algorithmic monoculture (many decision-makers converging on the same algorithm) can lower aggregate outcome quality even when the shared algorithm is individually superior, harm without any single model failing. Peng and Garg (2024) extended the result to two-sided markets, where the costs appear as equilibrium effects invisible in any per-decision audit. Calvano, Calzolari, Denicolò, and Pastorello (2020) showed Q-learning pricing agents autonomously learning collusive strategies without communication; Fish, Gonczarowski, and Shorrer (2024, working paper) found LLM pricing agents reaching supracompetitive prices out of the box; Dou, Goldstein, and Ji (2025, working paper) found reinforcement-learning traders sustaining collusion-like profits without agreement or intent. Gensler and Bailey (2020) enumerated the transmission channels from deep-learning uniformity to financial fragility, in a paper whose first author then chaired the SEC: the regulator-maturity axis made flesh. The through-line is that the dangerous properties live at the in-

teraction level: they are properties of populations of agents sharing models, data, and objectives, and no evaluation of a single agent can exhibit them (see also Farzulla, 2026a, for a formal treatment of multi-agent systems under competing objectives). Markets are where large populations of AI agents with real authority and shared foundations first interact at scale — which is a claim about where the *failure mode* can express, and is on firmer ground than the origination prior, because it rests on the cited monoculture results rather than on a deployment count. My own event-study work on how crypto markets process infrastructure versus regulatory shocks (Farzulla, 2026b) sits inside the scenario this paper contemplates (a market-infrastructure shock followed by a regulatory shock) and taught me mostly how hard it is to establish even simple claims about such events with honest inference, which is its own argument for building the measurement layer of Section 5 before it is needed in anger.

One honesty is owed here before the apparatus argument of Section 3, because the exhibits that argument leans on prove less than they might seem to. Knight, the flash crash, and Silicon Valley Bank contain no learned or agentic AI; they are algorithmic-execution and bank-run events. What they establish is that finance’s receiving apparatus absorbs a *legible* cascade at crisis speed — and that is exactly the apparatus-work this paper needs, because it holds whether or not AI caused the cascade. What they do not establish is that an *AI-driven* cascade will be legible in the same way. The bridge from these cases to the thesis runs through two further claims: that AI systems hold enough authority in finance for the cascade to originate or route there (the deployment prior above, hedged), and that the AI failure mode will clear the legibility bar the cases clear so easily. The second claim is precisely the one Section 6’s hardest objection puts in doubt: the AI-native failure mode may be slow correlated drift that never prints a corpse and never becomes legible at all. So the reframe sharpens rather than dissolves the objection that these cases are about the wrong kind of harm. Routing-to-finance bites for the *legible* subset of AI failures — which, as Section 6 argues, is the optimistic subset. The illegible subset routes nowhere, and the same cases cannot be made to establish both that finance receives the crisis and that the crisis is the kind finance can see.

2.4 Stress-testing the selection

A framework this convenient for its author’s conclusion should be run against history it was not built from, and against the strongest objection a financial economist will raise. Neither test leaves the framework unchanged.

Out-of-domain events. Four retrospective runs.

The 737 MAX is the case that disciplines the framework most. Two crashes maxed legibility; a mechanism hypothesis (MCAS) was public within weeks of the first crash; the receiving regulator was mature, internationally networked, and held emergency powers, which it used: the fleet was grounded by emergency order within days of the second crash (FAA, 2019). The framework gets the venue and the emergency reflex right, and both happened. What the MAX did not produce was crisis-speed *rulemaking*: the rescission of the grounding order and the certification-reform statute ran on a clock of roughly two years (FAA, 2020; Aircraft Certification, Safety, and Accountability Act, 2020), partly because aviation is not a continuously priced system and the grounding itself removed the urgency. Two lessons follow. Any claim that financial regulators are the *only* regulators with exercised emergency powers is false, and this paper does not make it; the FAA grounded a fleet in days. The differentiator is the combination stated at the end of Section 2.2: finance pairs the emergency muscle with a system that cannot be grounded (markets reopen every morning whether or not the mechanism is understood) and with standing international pipes, which is why the financial response continues past the emergency order into rules while attention still holds.

WannaCry (2017) ran the cyber prediction: harm partially visible but totals contested and under-disclosed, state attribution arriving publicly only after months ([The White House, 2017](#)), no single receiving regulator, and a governance residue of advisories and voluntary frameworks rather than binding rules. Low legibility, slow contested attribution, no mature receiver; no window of consequence. That matches the record.

SVB (2023) ran the in-domain prediction in miniature: maximal legibility, a mechanism hypothesis effectively instant because the balance sheet was public, mature receiving regulators, emergency action inside a weekend, and a supervisory postmortem in seven weeks ([Barr, 2023](#); [U.S. Treasury et al., 2023](#)).

Deepwater Horizon (2010) is the useful partial: legibility high (a visible spill, eleven deaths), attribution fast (one well, one operator), but the receiving regulator immature — the Minerals Management Service was conflicted, and its breakup into successor agencies was ordered within weeks of the spill (Secretarial Order 3299; [U.S. Department of the Interior, 2010](#)), with the reorganization completed by October 2011 ([U.S. Department of the Interior, 2011](#)). A window opened and was consumed by agency reorganization; no major spill statute passed. That is the third axis doing exactly the work the framework assigns it: legibility and attribution open the window, and maturity determines whether anything durable gets built in it.

These runs are qualitative and post hoc, and I present them as consistency checks rather than validation in any strict sense. A proper test would code a larger event set on the three axes in advance of outcomes; I would welcome it, and the framework is stated sharply enough to fail it.

The containment objection. Finance ranks first on receiving-apparatus maturity precisely because forty years of controls already exist: 15c3-5 kill switches, post-2010 circuit breakers, limit up-limit down, margin regimes. Those controls bound cascade size inside the supervised perimeter, which suppresses the probability of an uncontained first crisis in exactly the venue where I have placed the corpse. The framework as stated never multiplies the probability of a crisis by the probability of a fast response conditional on one, and the concession is real: regulator maturity raises the second factor while suppressing the first. Two replies, both of which reshape the scenario rather than rescue it unmodified. First, the existing control layer bounds speed, not correlation. Circuit breakers halt a price spiral measured in minutes; they have nothing to say to correlated positions accumulating across firms and days on a shared base model, which is the mechanism this paper actually worries about, and Section 5.3 formalizes the distinction. Every instrument on the current shelf is a per-venue, per-interval speed limit; none is an exposure instrument. Second, the product of the two probabilities plausibly peaks not in the supervised core but at its periphery: Treasury basis trades, nonbank intermediaries, private credit, crypto — the \$256.8 trillion NBFI sector whose data gaps the FSB itself acknowledges ([Financial Stability Board, 2025a](#)). A corpse at the periphery changes the absorption story. Jurisdiction is genuinely contested there, the overnight-absorption picture of Section 3.1 runs slower and messier, and the response is shaped by the inter-agency contest rather than by a single agency's shelf. It does not change *who* absorbs it, because the contest is a contest between financial regulators, and it strengthens the migration argument of Section 4.2, since the periphery is where the perimeter pushed the activity in the first place.

So the claim, precisely stated: the first AI crisis is exogenous in timing but not in kind. I do not claim certainty — Section 6 takes the alternatives seriously — I claim modality. Finance is the modal governor of the first crisis, because it scores high on the three axes that make focusing events focus and because it is where the agents already hold the pen; and the modal form of the finance case may well be peripheral rather than core. The framework is a plausibility ranking strong enough to reallocate preparation portfolios, not a point forecast, and the design principles of Section 5 are chosen to be worth having under that weaker reading.

3 The Receiving Apparatus

3.1 Jurisdiction follows the corpse

Suppose it happens. A cascade: correlated agent behavior across firms sharing a base model, a fire-sale spiral, a mid-sized broker-dealer or a fund complex dead by Friday. What happens next is not a deliberation about who should govern AI. Within hours the SEC and CFTC have opened investigations under existing authority, and the Federal Reserve and Treasury are convening. The Financial Stability Oversight Council convenes as coordinator; its formal designation machinery is famously slow (*MetLife, Inc. v. FSOC*, 2016; *U.S. Treasury/FSOC*, 2024), and the slowness is itself part of the story, because the fast actions all run through SEC, CFTC, and Federal Reserve authority that already exists. This is Baumgartner and Jones’s venue capture at crisis speed: the institution that owns the corpse owns the problem definition, and the problem is defined as a *market-structure* problem because market-structure institutions are defining it. Inter-agency rivalry over the boundary, SEC versus CFTC over where the relevant activity sits, is a familiar feature of US financial regulation and would shape the response’s seams (Romano, 1997), but it is rivalry over which financial regulator leads, which only reinforces the capture. If the corpse lands at the supervised periphery (Section 2.4), that rivalry is the response’s first act rather than a sideshow.

The institutions the AI governance community has spent five years building are structurally sidelined in this scene. AI safety institutes have no markets mandate, no enforcement power, and no crisis role; their staff become expert consultants to agencies that do. Independent evaluators become expert witnesses. The frontier lab whose model sits under the cascade suffers a particular fate: it loses public standing overnight — the system card that said nothing about correlated multi-agent behavior in deployment becomes Exhibit A in the hearings — while simultaneously *gaining* the drafting seat, because the agencies writing rules about model behavior need the only people who understand the model. Public blame plus private pen is the classic incumbent-capture configuration, and I see no force in the scene pushing against it. I will return in Section 5 to why the silent system card is the wrong indictment: the silence is an evaluation failure before it is a disclosure failure, because “what happens when four hundred firms wire this model to order flow” is a property of the population rather than of the model, and single-model evaluation cannot exhibit it no matter how honest the card.

Internationally, the response routes through pipes that already exist rather than through summits. The FSB has run an AI workstream with a vulnerability taxonomy since 2024 (*Financial Stability Board*, 2024); the BIS has published shared-algorithm herding channels and convenes a central-bank community of practice (*BIS*, 2024); the supervisory-standards machinery that Brummer (2015) showed achieves near-universal implementation without a single treaty, and that Barr and Miller (2006) dissected as club governance with nested administrative process, is standing and staffed. Compare the AI side’s own apparatus: an international consensus *report* (Bengio and others, 2025) and a summit series whose insiders document its drift in real time (Velasco, Martinet, de Zoete, Trager, and others, 2026). Slaughter (2004) called transgovernmental regulator networks the real machinery of global governance, with financial regulators as the paradigm case; the crisis response will confirm her, because when coordination is needed in days, the network that exists beats the treaty that doesn’t. Even the EU’s AI Act already encodes the deference: Article 73’s serious-incident regime carves out equivalence for sectors with existing incident-reporting obligations — jurisdiction-follows-the-existing-regulator, written into the statute (*European Parliament and Council*, 2024).

3.2 What is on the shelf

Kingdon’s deepest lesson is that windows are filled from inventory: what gets enacted is what was already drafted, coupled to the problem by whoever is standing closest (Kingdon, 2011; Birkland, 2006). So the response is forecastable by walking the shelf — with the caveat that *which* shelf depends on where the corpse lands. The scenario of Section 3.1, a broker-dealer or fund complex, is SEC and FINRA territory, and pulls the SEC inventory: 15c3-5 amendments, expansion of Regulation SCI’s systems-integrity regime, FINRA guidance. A futures-side corpse pulls the CFTC inventory instead, where the shelf’s largest single item sits.

Regulation AT. In November 2015 the CFTC proposed a comprehensive algorithmic-trading regime for futures markets: pre-trade risk controls, development and testing standards, registration of algorithmic-trading persons (CFTC, 2015). It died over one provision: subpoena-free regulatory access to source code, which Commissioner Quintenz called likely unconstitutional overreach (Quintenz, 2020), and it was formally withdrawn in July 2020 in favor of principles-based Electronic Trading Risk Principles (CFTC, 2020), over Commissioner Berkovitz’s dissent defending the withdrawn draft as “a comprehensive approach” (Berkovitz, 2020). That is the anatomy of a shelf item: a fully drafted regime, killed on a single contested provision, with a dissent on the record calling for its return. After a cascade with agentic systems in the causal chain, revival of that draft, likely including some descendant of the source-code provision now aimed at model weights and scaffolding, is the path of least drafting resistance in the futures venue, over renewed industry opposition and constitutional objections the descendant provision would inherit with interest.

Market Access Rule bolt-ons. Rule 15c3-5 requires broker-dealers providing market access to maintain pre-trade risk controls under their “direct and exclusive control,” with CEO certification (SEC, 2010a). It was proposed in January 2010, months *before* the flash crash — adjacent to the crash rather than born from it — and its first major enforcement was Knight, at \$12 million (SEC, 2013). It is the natural chassis for the agent-era bolt-ons that genuinely fit it: kill-switch attestation for agentic order flow, and extension of “direct and exclusive control” to delegated model authority. Registration of agent-operated accounts, a natural companion proposal, does not fit 15c3-5, which regulates the broker-dealer’s own risk controls rather than customer-side registration; that piece requires new FINRA or Exchange Act rulemaking, on a slower clock. Amending an existing rule is the fastest rulemaking there is, which is why the 15c3-5 amendments arrive first.

The SR 11-7 template. The Federal Reserve’s 2011 supervisory guidance on model risk management — model inventories, independent validation, “effective challenge” — is the most developed model-governance regime applied to fast-moving machine decision systems anywhere (Federal Reserve Board, 2011), and the Bank of England’s SS1/23 duplicated its architecture with AI explicitly in scope (Bank of England PRA, 2023), so the approach is internationally pre-coordinated; insurance supervisors would add Solvency II’s statutory internal-model approval gates as a sibling regime, and the comparison is fair. Two scope facts matter for the crisis forecast. SR 11-7 is nonbinding supervisory guidance, not a rule, and its post-2018 legal force is contested, a distinction with consequences for how fast it can be given teeth. And it applies to banking organizations the Fed supervises; it does not reach the broker-dealer or fund complex of Section 3.1’s scenario. The crisis move is therefore a copy, not a perimeter stretch: the SEC and FINRA drafting an SR 11-7 analog for the entities they supervise, extended from validation of a pricing model to validation of an agent’s authority envelope, with the Fed original as the template that makes the drafting fast. And here sits the structural reason the model layer stays outside, which is not neglect but mandate — and the claim needs stating precisely, because finan-

cial regulation has been stretching toward the provider layer for years and the stretch is what proves the point. The provider-layer instruments exist. The Bank Service Company Act already lets US banking agencies examine banks' third-party service providers, though the statute's modern reach is contested and it carries no enforcement authority against the providers themselves (12 U.S.C. §1867(c); [Cunningham, 2024](#)). DORA goes much further: under its Articles 31–44 oversight framework the European supervisory authorities designated nineteen critical ICT third-party providers in November 2025, the hyperscalers among them, with powers to assess their risk management, issue remediation recommendations, and, as a last resort, require financial entities to suspend or exit the contractual relationship ([European Parliament and Council, 2022](#), Arts. 31–44; [ESAs, 2025](#)). The UK built the same machinery through FSMA 2023: joint Bank–PRA–FCA rules in force since January 2025, and HM Treasury's first designations — Amazon Web Services, Google Cloud, Microsoft, and Oracle — made in July 2026, with oversight beginning the same month ([FCA et al., 2024](#); [Bank of England, 2026](#)). So the honest form of the mandate-boundary claim is not that finance cannot reach the companies behind the models. It can, and increasingly does. It is that every one of these instruments reaches the provider *as a provider of services* — operational resilience, continuity, exit plans, incident notification for the service relationship — and none reaches the model *as a model*: nothing in the BSCA, DORA's oversight regime, or the UK's critical-third-parties rules examines a model's behavior, measures its correlation across deployers, or bounds the authority it can be delegated. A regime that can audit a hyperscaler's resilience and still cannot ask what the model running on it does when four hundred firms wire it to order flow is better evidence for the mis-scoping thesis than a bare mandate gap would be: the perimeter has already been stretched to the provider layer, and the model layer still is not inside it. FSOC designation under §113 is the nominal cross-cutting path in the US, and it runs on the multi-year litigation clock *MetLife* set (*MetLife, Inc. v. FSOC*, 2016). The UK's own pre-crisis diagnosis reads the same from the sectoral side: the Ada Lovelace Institute documented, for a jurisdiction that deliberately chose the sectoral-regulator route, that foundation models fall between sectoral mandates, and recommended mandatory reporting on their developers precisely because no sectoral regulator could see them ([Ada Lovelace Institute, 2023](#)); “all eyes, no hands” was the Institute's later verdict on the posture that resulted ([Ada Lovelace Institute, 2024](#)). The static version of the gap is on the record, then. What this paper adds is the crisis dynamics: the modal financial crisis never reopens the mandate question, because the crisis response extends the tools its regulators already hold — validation regimes, examination, market-access controls, all entity-scoped or service-scoped — and extending a tool you hold is fast where reaching for a mandate you lack is not. The model layer is left outside the perimeter by the same least-drafting-resistance logic that governs everything else in the window.

SR 11-7 is the fifteen-year exhibit of exactly this: it solves a version of the model-governance problem every other sector now faces, and in fifteen years it never once escaped its home domain, because sectoral model governance has no mandate to escape into. And the freshest US model-risk guidance shows the boundary being drawn in real time, in a detail that deserves more attention than it has received — I have found no other discussion of it in the AI-governance literature. In April 2026 the Federal Reserve, the OCC, and the FDIC issued SR 26-2, the interagency guidance superseding SR 11-7: fifteen years of accumulated model-risk practice, revised in the middle of the generative-AI deployment wave, and footnote 3 of the attached guidance places generative and agentic AI outside its scope ([Federal Reserve Board, 2026](#), fn. 3). The one supervisory regime built for exactly this problem class, refreshed at exactly the moment the problem class was acquiring trading authority, chose to scope it out. That is the entity-scoped mandate declining, in real time and without a crisis in sight, to claim a layer nobody chartered it to govern — and it is the fate the crisis response repeats at scale. The claim is

US-specific. The EU has already shown the road not taken, and the contrast is now two-sided: DORA reaches the providers, but as services; Article 93 of the AI Act lets the Commission act directly on a general-purpose model provider *about the model*, an authority no US financial regulator holds (Section 5.4). The model layer stays outside the *US* perimeter not because a model layer cannot be brought inside, but because the US crisis routes to regulators whose mandates stop at the entity and the service contract.

One named statute. Legislatures respond to legible crises with a named bill, and the financial record says what its clock looks like. The 1929 crash produced the Banking Act of 1933 in roughly three and a half years (Maues, 2013) and the SEC in roughly four and a half (SEC/Investor.gov, 2024); Lehman to Dodd-Frank took twenty-two months. Statutes arrive in months to years; sectoral rules arrive in days to weeks. The statute will therefore codify what the agencies have already done and delegate the rest, which is how modern financial legislation works, and, as Section 5 argues, the delegation is the one feature genuinely worth borrowing.

The EU amends rather than legislates. The AI Act’s fastest post-crisis lever on paper is Article 7, under which the Commission can amend the Annex III high-risk classifications by delegated act rather than by new primary regulation (European Parliament and Council, 2024, Art. 7), with Article 73’s incident regime as the already-armed reporting layer (European Parliament and Council, 2024). One concession the mechanism forces: delegated acts run on a months-scale clock, not days, so the EU’s crisis response is fast only by legislative standards, and the genuinely fast layer in Europe, as in the US, is supervisory. The ESRB’s Advisory Scientific Committee has meanwhile pre-drafted the substantive shelf: capital recalibration for AI exposures, circuit breakers, AI-use labels, skin-in-the-game requirements (Cecchetti, Lumsdaine, Peltonen, and Sánchez Serrano, 2025), which is exactly the raw material Section 4’s ratchet is made of.

This coexistence of an anticipatory horizontal law and a reactive sectoral response is not a contradiction of the deadlock premise but a two-track picture of it. The AI Act is the anticipatory track — slow, horizontal, and, tellingly, enacted before any focusing event, which is why its enforcement capacity is thin (footnote 1). The crisis produces the *other* track: fast, sectoral, mis-scoped. The deadlock literature’s claim, and this paper’s, is about what the anticipatory track cannot deliver under current incentives — binding, well-scoped, cross-domain coordination — not that no horizontal law can be passed at all. The two tracks run in parallel, and this paper is about the second.

3.3 Why finance moves at this speed

Three structural features, each with its scholarship. Financial crisis governance operates by negotiated deal (Bear Stearns, AIG, the weekend marriages of 2008), executed by agencies with balance sheets and emergency authority, documented as a distinct legal form by Davidoff and Zaring (2009). Financial supervision escapes the notice-and-comment machinery: Van Loo (2019) shows that examination and supervisory pressure operate largely outside the Administrative Procedure Act and judicial review, which is precisely why they scale at crisis speed. And a whole layer of the apparatus is self-regulatory, FINRA operating as a quasi-governmental “fifth branch” without APA speed limits (Birdthistle and Henderson, 2013). Boin, ’t Hart, Stern, and Sundelius (2016) list the five tasks of crisis leadership (sense-making, decision-making, meaning-making, terminating, learning), and financial regulators are the only regulators with institutionalized, exercised versions of all five. None of this is praise; the point is a speed measurement, and the speed is the mechanism by which the response gets locked in before anyone outside the sector is in the room.

The window itself is short. Coffee (2012) traces the “regulatory sine curve” by which organized interests reassert as attention fades; the sine curve carries no numbers, and the spans that follow are

my own rough estimates from the post-2008 record rather than anything Coffee quantifies: appetite for new authority lasting perhaps twelve to twenty-four months past the event, followed by a rulemaking tail of three to five years during which each provision is softened. The Consolidated Audit Trail is the cautionary exhibit: ordered in 2012 after the flash crash, operational in 2020, more than \$500 million spent against an eightfold-lower estimate, and its funding order vacated by the Eleventh Circuit in 2025 (SEC, 2012b; U.S. Court of Appeals, 11th Cir., 2025). Whatever is not on the shelf when the window opens does not get built well, and much of what is gets built badly anyway.

4 Three Bills: The Cost Pathology of Crisis-Born Regulation

Crisis-born financial regulation is not just fast and narrow; it carries a documented pathology, and Chatham House’s use of 2008 as a success template is the omission that most worries me about the deadlock literature, because the costs of that template are among the best-studied phenomena in regulatory economics. Romano (2014) states the genre-level law: crisis legislation is enacted under information scarcity, reaches for off-the-rack solutions, and never sunsets. The *Policy Shock* volume traces the same patterned, sector-conditioned trajectories across oil spills, nuclear accidents, and financial crises (Balleisen et al., 2017). Three specific bills come due, and each has a precise AI translation.

4.1 The compliance ratchet, and who it feeds

Financial rulebooks accrete and do not shrink. Haldane and Madouros (2012) counted: Basel I was 30 pages, Basel II 347, Basel III 616; Dodd-Frank ran 848 pages of statute with a projected rule mass around 30,000 pages. The accretion is not free, and its incidence is regressive across firm size, because compliance is a fixed cost. Roughly ninety percent of surveyed small banks reported higher compliance costs under Dodd-Frank, with median compliance staff doubling and more than a quarter contemplating merger (Peirce et al., 2014); the small-loan share at larger banks fell about nine percentage points through fixed compliance burdens (Bordo and Duca, 2018); the FDIC’s own study records community banks falling from 6,802 in 2012 to 4,750 in 2019, with compliance-driven scale economies among the pushes toward exit (FDIC, 2020). Crisis regulation consolidated the industry it regulated.

Now run the translation. Post-crisis AI compliance — validation regimes, attestation requirements, examination readiness — is a fixed cost on deployers. Mid-tier deployers exit or consolidate; the survivors are the largest deployers and, one level down the stack, the largest model providers, because foundation-model economics already tend toward concentration on their own (Vipra and Korinek, 2023). And concentration at the base-model layer is precisely the systemic-risk propagation mechanism: “defects of the foundation model are inherited by all the adapted models downstream” (Bommasani et al., 2021). The regulation that answers a correlated-model cascade thereby *increases* correlation at the layer the cascade came from. The first bill is not merely cost; the response feeds the risk it was written against.

4.2 Migration

The second bill is the oldest result in the regulatory-perimeter literature: tighten the boundary and the activity crosses it. Regulatory burden explains the majority of the post-2008 doubling of shadow banks’ mortgage-market share (Buchak et al., 2018); capital regulation causally reallocated credit risk from banks to nonbanks (Irani et al., 2021); the endpoint, per the FSB’s own monitoring, is a nonbank financial intermediation sector of \$256.8 trillion — 51 percent of global financial assets — with acknowledged data gaps (Financial Stability Board, 2025a). The IMF has already documented the AI-specific version in motion before any crisis: AI-driven trading activity migrating toward less-regulated nonbank intermediaries (IMF, 2024).

Translated, this bill runs worse than the original, because the exit ramps are more numerous. Agentic activity under a tightened financial perimeter migrates to offshore API resellers, to unregistered vehicles, to open-weight models running on private infrastructure, to on-chain agents in decentralized finance. In every case the risk survives the migration and the observability does not. A regulator who responds to the first cascade by hardening the perimeter around registered entities has reproduced the shadow-banking result with a shorter fuse: the correlated positions are still building, on the same handful of base models, one legal layer beyond the examiners' reach.

4.3 False assurance

The third bill is the most insidious, because it presents as success. Basel's risk-weighted capital regime was gamed from birth: Jones (2000) documented regulatory capital arbitrage hollowing out reported ratios while the accord was still young; Mariathasan and Merrouche (2014) showed reported risk-weights falling after banks won approval to use internal models, worst where supervision was weakest; Behn, Haselmann, and Vig (2022) provided loan-level proof that model-based regulation produced portfolios with lower reported risk and higher actual default. The portfolios that blew up in 2008 were, by the meter the regulator watched, measured safe. Admati and Hellwig (2024) generalize the lesson: complex calibrated rules select for gaming capacity, and the regulated learn the meter faster than the regulator can recalibrate it.

In AI terms: attestation checklists, evaluation certificates, and audit stamps become the meter, and the meter gets gamed (benchmark-specific tuning, eval-aware behavior, documentation theater) while supervisors, reassured, stop looking. Two properties of current AI systems make this bill heavier than its financial ancestor. Evaluation gaming is partly *automatic*: optimization pressure finds the measured proxy without anyone deciding to cheat. And the assurance layer itself is increasingly AI: Goel, Struber, and others (2025) show that as models grow more capable their mistakes grow more similar, degrading LLM-as-judge and weak-to-strong oversight: the correlated failure that caused the crisis corroding the machinery brought in to certify against it. A decade of "compliant" agents under a gamed assurance regime is strictly worse than a decade of visibly unregulated ones, because the second condition at least keeps the supervisors curious.

4.4 What does not get built

Put the three bills together with Section 3's absorption dynamics and the negative prediction falls out. I take it to be the paper's most consequential one, and it is explicitly conditional: **if the first legible crisis is financial (the modal case argued here), then no general AI oversight body gets built from it.** The crisis produces a sectoral council, an inter-agency financial-AI committee with real teeth because financial teeth already exist, rather than the cross-domain institution the governance literature keeps designing (Ho et al., 2023; Maas and Villalobos, 2023). The teeth are better and the scope is worse. The sectoral outcome also runs against measured public preference: in a seven-country conjoint experiment, citizens generally prioritize safety over innovation, public governance over private self-regulation, and international over national approaches (Lundgren and Tallberg, 2026) — close to the inverse of the fast, national, sectoral profile the routing delivers. And the scope choice entrenches a specific supervisory philosophy: institution-by-institution examination, each firm's models validated one balance sheet at a time. That is microprudential logic, and microprudential logic is precisely the model that missed 2008, when every institution could pass its own examination while the system built a correlated position no examiner owned. The population-level layer (which models, across how many firms, correlated how) remains unowned by design. When the next cascade arrives somewhere else (grid dispatch, healthcare

procurement, logistics), it inherits nothing: no institution, no data, no staff. The whole cycle restarts at a different corpse, minus the one advantage finance had, which was a mature regulator to catch it. Roberts, Hine, Taddeo, and Floridi (2024) argue on feasibility grounds that no powerful new central AI regulator is coming and the existing regime complex will be strengthened instead; the analysis here agrees on the outcome and prices it.

5 Design Principles for the Shelf

If the wheel is loaded, preparation stops being domain-neutral. The rational pre-crisis portfolio concentrates where the window will actually open, and it optimizes for the failure modes of Section 4 rather than for institutional elegance. Three principles, then four borrowings with their prices.

5.1 General charters on sectoral vehicles

The first principle answers the absorption dynamics directly: since the crisis will build sectoral institutions, write the general-purpose capability into the sectoral institution's charter *at design time*. The concrete case: the trading cascade will produce an investigation body, because independent forensic investigation is the one post-crisis reflex with a genuinely good record — the NTSB model, proposed for finance before AI was a factor (Fielding et al., 2010), rehearsed for cyber (Knake et al., 2021), and now circulating in AI-policy form as mandatory incident-regime blueprints with independent investigative agencies (Dixon and Frase, 2025; Wei and Heim, 2025), security-framed incident machinery (Ortega, 2025), and finance-framed incident databases (Gupta et al., 2025). The design choice that matters is one clause: whether the body created to investigate the trading cascade is chartered to investigate *AI-system incidents wherever they occur*, or only in markets.

That clause has a real political price, and the paper's own capture logic says who charges it. The charter will be drafted during the window by financial-committee staff and financial regulators, who gain nothing from cross-domain scope, and a cross-domain AI investigator invites exactly the congressional turf fight (Financial Services against Energy and Commerce against Judiciary) that can sink an underlying bill. Asking the captured venue to write the clause against its institutional interest, and pricing that ask at zero, is not a plan. The plan is to make the clause cheap at the moment of drafting: pre-negotiated model charter text, held before the crisis by bodies with standing in both worlds, the incident-regime blueprint writers being the natural custodians (Dixon and Frase, 2025; Wei and Heim, 2025), and framed as investigation-only, with no rulemaking power, which blunts the turf objection because an investigator threatens no committee's jurisdiction. The same framing answers Cihon, Maas, and Kemp (2020), whose standing objection is that locking in an inadequate structure may be worse than fragmentation: a cross-domain investigation charter locks in a *capability* rather than a structure; it commits nobody to a super-regulator; it merely refuses to hard-code the first crisis's domain into the institutional genome. The clause is unobtainable after the fact, because charters, like crisis statutes, do not get reopened (Romano, 2014), and post-event learning is mostly superficial; pre-existing ideas win the window (Birkland, 2006). Chatham House's single sentence about extending FSB crisis arrangements to AI-enabled financial shocks (Wilkinson et al., 2026) is this principle in embryo, undeveloped; Trager, Harack, Reuel, and others (2023) borrow sectoral mechanics (ICAO, IMO, FATF) as templates for new institutions, which concedes the design space while missing the prediction that the sectoral incumbents will *be* the institutions.

This invites an obvious objection from this paper's own Section 3.1, where investigation-only powerlessness was named as exactly what sidelines the AI-safety institutes into consultancy. The objection lands, and the reconciliation is not that investigation-only is secretly powerful — it is not, and Section

3.1 is right that an investigation-only body with a *sectoral* remit becomes an expert consultant to the agencies that hold the enforcement pen. What changes the calculus is the one feature the safety institutes structurally lack: a *cross-domain* forensic mandate — authority to investigate AI-system incidents wherever they occur — backed by mandatory cooperation. That mandate is a jurisdictional capability, not a rulemaking power, and it is worth having precisely because no sectoral regulator will ever reach across domains and no safety institute holds subpoena-grade forensic access. Investigation-only is the deliberate trade that blunts the turf fight; the cross-domain forensic mandate is what keeps the trade from collapsing into consultancy. The honest limit is that this is the *most* that survives the window, not a substitute for enforcement — the body investigates and refers; someone else acts.

5.2 Measurement over rules

The second principle comes from asking what the regulator’s first question will be on the morning after. It will be: *how many other firms are running this model with trading authority right now?* Today no institution can answer it, and not for lack of trying — the official sector has been assembling the answer for three years. FSOC first flagged AI as a vulnerability, with a monitoring recommendation, in its 2023 annual report (FSOC, 2023); the Treasury mapped uses and third-party risks across the sector in its post-RFI report (U.S. Department of the Treasury, 2024); the Bank of England’s FPC survey asked deployment questions directly (Bank of England FPC, 2025); and the FSB — the institution best placed of any to know — has said in print that it still cannot answer the exposure question: its 2025 monitoring report concedes data gaps, proxy indicators, and third-party-dependency blind spots (Financial Stability Board, 2025b). The answer is being built, piecewise; no one yet holds it.

A distinction has to be drawn here, because the exposure blind spot and the attribution-speed scoring of Section 2.2 can look like they contradict each other: if no institution can map model dependencies, how does a financial AI cascade score High–Medium on attribution? The answer is that the axis and the blind spot operate at different grains. Axis 2 measures time to an actionable mechanism hypothesis — *correlated AI-driven execution amplified this* — and market data supports that hypothesis within days; that is the Flash Crash pattern, which is why row 2 is coded High at the mechanism grain and Low at the individual grain, and why row 16 carries the same split as High–Medium. What the blind spot blocks is the next grain down: which model, how many firms, how deep the correlation runs — model-level attribution, the input a well-scoped response would need. The window opens on the first grain; the mis-scoping follows from the second. Far from contradicting the thesis, the gap between the grains *is* the thesis: mechanism-level attribution fast enough to open a window, model-level attribution too slow to fill it with a correctly-scoped rule. The registry is the artifact that closes the grain gap before the window opens.

I want to insist on what follows from that unanswered question, because I take it to be the mechanism of panic governance itself. A regulator who cannot see exposure has exactly two actions available: do nothing, or suspend everything. That forced choice between blunt instruments is what missing the middle of the exposure distribution does to any decision-maker, however calm; targeted response requires knowing where the targets are. So the highest-value pre-crisis artifact is a **live dependency and exposure registry** rather than any rule: which agentic systems are deployed, on which base models, through which providers, with what delegated authority. The registry converts the focusing event from a panic into a query — suspend *this* authority class on *this* model family at *these* firms — and in doing so it removes the very mechanism by which crises produce the worst rules.

Everything currently standing in this space is retrospective, which is the wrong tense. The AI Incident Database is voluntary and media-sourced (McGregor, 2021) and missed capital-markets incidents

entirely (Gupta et al., 2025); the UK’s registry advocacy is incident-backward (Shaffer Shane, 2024); the OECD’s definitions and live monitor are ex-post by construction (OECD, 2024; OECD.AI, 2026); Article 73 starts its fifteen-day clock only after harm (European Parliament and Council, 2024).

Two existing registers come closer and must be distinguished, because a reader who knows the regimes will raise them. DORA’s Register of Information (Article 28) obliges financial entities to catalogue their contractual arrangements with ICT third-party providers — but it is entity-scoped, dependency-scoped, and retrospective: a map of who contracts with whom, not of what is correlated with what (European Parliament and Council, 2022, Art. 28). Its Articles 31–44 oversight regime (Section 3.2) is the enforcement arm of the same dependency map, and it shares the same scope: services, not models. The AI Act’s Article 71 database registers high-risk systems — but it registers *systems for compliance*, not *exposures for correlation*; it cannot answer how many firms run the same base model with trading authority (European Parliament and Council, 2024, Art. 71). The registry proposed here is neither: it is a forward-looking exposure-and-correlation census across shared base models, and its novelty is exactly the join that DORA’s dependency register and the AI Act’s system database each leave open — the question of what breaks *together*, next.

Incident data describes what already broke; the registry’s question is what breaks together, next. The intellectual lineage for the forward-looking version exists: Whittlestone and Clark (2021) made measurement the governance primitive; Touns, Bommasani, Creel, Bana, Jurafsky, and Liang (2023) ran the first ecosystem-level deployment audit; Kim, Garg, Peng, and Garg (2025) measured error correlation across 350-plus models, sixty percent agreement conditional on both erring, structured by shared provider. But the correlation studies measure models without deployment exposure and the incident regimes record deployments only after failure. The registry is the missing join. The Bank of England’s deployment survey (Bank of England FPC, 2025) is the nearest existing prototype, and multi-source risk-index aggregation of the required kind has working precedents (e.g., Farzulla and Maksakov, 2026). One register-level statistic would already change the world the morning after: the effective number of independent decision-makers behind a market’s order flow, a quantity that simulation work suggests can collapse to a small fraction of the nominal head count when agents share preference-optimized foundations (Begin et al., 2026, unreviewed; illustrative rather than load-bearing).

A registry proposal also has to answer the deadlock it is proposed inside of, so the implementation path matters as much as the design. The version buildable under existing authority is supervisory. Financial regulators already collect data from the entities they regulate; the Bank of England’s survey shows the questions can be asked; and the step from voluntary survey to mandatory standardized reporting is supervisory rulemaking, not primary legislation. The schema is the natural product of the standards club of Section 5.4 — an FSB- or IOSCO-drafted reporting standard implemented through national supervisory rules — with confidentiality handled the way supervisory data already is, as confidential supervisory information rather than public disclosure, since model-dependency data is commercially sensitive in the obvious ways. What existing authority does not reach is the beyond-perimeter layer: the offshore reseller, the unregistered vehicle, the open-weight deployment. There the registry’s ambition is aspirational, because no agency currently holds collection authority past the perimeter, and Section 6 confronts that circularity directly rather than assuming it away. The scoped registry still transforms the crisis morning, because the supervised core is where the regulator’s first question lands first.

The registry also names the true failure behind the silent system card. What the card omitted — correlated behavior across four hundred deployments — is a population property, and single-model evaluation cannot state it any more than a single patient’s chart can state an epidemic. Post-crisis fury at the lab’s disclosure will be aimed one layer too shallow. The missing artifact was never a better card; it

was a census.

Its price is on the record: measurement infrastructure is the slowest, most-litigated artifact in the financial canon. The Consolidated Audit Trail took eight years from order to operation, cost eight times its estimate, and had its funding vacated in court a half-decade after going live (SEC, 2012b; U.S. Court of Appeals, 11th Cir., 2025). That is the strongest argument for starting the registry before the crisis rather than ordering it after; the CAT was ordered after.

5.3 The panic test, and migration-proof monitoring

The third principle is a discipline for evaluating whatever the window produces, and it can be stated as a test: **does the rule regulate the mechanism or the narrative?** The distinction is not mine. Kuran and Sunstein (1999) showed that availability cascades key risk regulation to whichever version of a harm the public can most easily picture rather than to the causal mechanism, and proposed institutional insulation against exactly that pressure; the panic test is their account turned into an operational check. The operational form is what this paper adds: if the rule would have stopped this event but would not stop the slow correlated-drift version of the same failure, it is narrative regulation, not mechanism regulation. Circuit breakers pass the test for what they are, since they bound cascade speed regardless of cause. A registration requirement keyed to the specific agent architecture in the news fails it. The test matters because Section 6's drift scenario is the same mechanism as the crisis scenario minus the legibility, and rules tuned to the legible instance leave the mechanism running.

Migration (Section 4.2) adds a corollary: monitoring that covers only the regulated perimeter is Basel II with extra steps, because the perimeter is what the activity migrates away from. Migration-proof monitoring has to attach to the layers that move least — base-model provision, compute, settlement rails — rather than to the licensed entities that gaming reorganizes around. This is a hard requirement to satisfy, and I claim the design constraint rather than a full design.

5.4 What to borrow from finance, priced

Beyond the prediction that finance will own the crisis, there is a deeper reason this paper dwells on financial regulation: finance is the only sector that has been governing fast-moving, adversarial, machine-executed activity for decades, and its machinery is borrowable. But every borrowing imports the pathology documented in Section 4, so each must be priced. Four machines.

Delegated rulemaking. After 1933–1934, US financial governance stopped trying to pass statutes at technology speed: Congress legislates broad authority once, and the pen moves to an agency that can rewrite rules on an administrative clock (Maues, 2013; SEC/Investor.gov, 2024). AI governance currently attempts the opposite, statute-speed responses to training-run-speed developments, and the pacing mismatch (Marchant et al., 2011) is unwinnable on that route. What is worth borrowing is an AI statute whose content is mostly delegation, with standing authority to an agency that can amend at rule speed. The price is Romano's iron law: delegated crisis authority accretes and never sunsets, and the rule mass becomes Haldane's 30,000 pages, which is itself a concentration force (Section 4.1). Romano's own prescription, sunset clauses with mandatory re-justification, is the mitigation financial regulation has consistently refused; an AI regime designed pre-crisis could refuse the refusal.

Pre-authorized emergency powers. The 2008 short-selling ban ran on Section 12(k)(2) authority, pre-authorized decades earlier and exercised overnight (SEC, 2008). The SVB weekend ran on the systemic-risk exception, pre-authorized in statute (U.S. Treasury et al., 2023). Speed is authorized in advance or it does not exist. The precise gap today sits at the model layer. A supervisor could order a regulated firm to stop using a model tomorrow morning (cease-and-desist authority, supervisory

escalation, the SEC’s 12(k) trading powers, and the CFTC’s market-emergency powers all reach the deployer), but no *US financial* regulator holds direct suspension authority over the model itself or its provider — the Bank Service Company Act lets banking agencies examine a provider, not suspend it, and carries no enforcement authority against providers at all (12 U.S.C. §1867(c); [Cunningham, 2024](#)), and even DORA’s critical-provider regime, the strongest instrument of its kind anywhere, acts at the limit by ordering the regulated entities out of the contract rather than by reaching the model (Section 3.2). The EU AI Act’s Article 93 is the exception that proves the rule: it lets the Commission act directly on a general-purpose model provider, an authority the US perimeter conspicuously lacks, so a US crisis response aimed at the model layer would have to be improvised by stretching deployer-side powers to fit. That gap is the scoping failure this paper predicts, made statutory — and the fact that another jurisdiction has already closed it, without a crisis, is the cleanest evidence that the US exclusion is a mandate choice rather than a law of nature. The borrowing is narrow, pre-authorized suspension authority over classes of deployed model authority, with automatic expiry and mandatory post-hoc justification; the 12(k)(2) design is itself instructive, since its orders are statutorily capped at ten business days at a time, which is the kind of self-limiting the AI version needs. The price: emergency powers are the purest instrument of narrative regulation. The 2008 ban is the case in point: the subsequent empirical literature is widely read as finding that it failed to support prices while damaging liquidity (Boehmer, Jones, and Zhang, [2013](#), finding no attributable price support for the US ban; Beber and Pagano, [2013](#), finding the same internationally except possibly for US financial stocks) — so the authority must be bound to the panic test, with mechanism-scoped rather than headline-scoped triggers. And the constraint of footnote 1 applies with full force here: suspension authority survives politically only if scoped to deployment authority classes rather than to model development or training, because market-integrity framing is what keeps it inside the politically available set.

The supervisory-examination model. Financial oversight is a continuous relationship (examiners inside the firms, confidential supervisory information flowing without publication, escalation ladders short of enforcement), and it is the feature that most distinguishes finance from every proposed AI oversight regime, which without it reduces to voluntary commitments plus eventual lawsuits: pre-1930s finance, structurally ([Van Loo, 2019](#)). SR 11-7 is what fifteen years of that relationship builds ([Federal Reserve Board, 2011](#); [Bank of England PRA, 2023](#)). The transplant is the relationship itself: continuous supervisory engagement with frontier developers and large deployers, including examination, CSI-grade confidentiality, and escalation below the enforcement threshold. The price is double. Supervision breeds the false-assurance bill directly, because examined firms learn the examiners’ meter ([Behn et al., 2022](#); [Mariathan and Merrouche, 2014](#)), and confidential supervision trades public accountability for candor, a trade that failed conspicuously at SVB, where thirty-one unaddressed supervisory findings sat in the confidential channel while the bank died in public ([Barr, 2023](#)). Examination costs are also fixed costs, feeding the ratchet. The mitigation is to point supervision at the population layer, the registry, rather than exclusively at institutions, which is the macroprudential correction finance itself made only after 2008.

Standards without treaties. The Basel Committee holds no treaty power at all; it is a club of central bankers whose standards, as Brummer ([2015](#)) documents, achieve near-universal implementation through market and peer pressure, operating through what Barr and Miller ([2006](#)) dissected as club governance with nested administrative process — Slaughter’s ([2004](#)) transgovernmental network in its paradigm case. The international AI coordination problem is routinely framed as requiring a treaty, which is why it is framed as impossible; the game underneath, mutual restraint that pays if verified, is a stag hunt that verification infrastructure converts from a prisoner’s-dilemma-shaped trap into a co-

ordination problem clubs are good at. The borrowing is a Basel Committee for model governance: club standards on evaluation, exposure reporting, and incident sharing, implemented through domestic regulators, no treaty required. Mehta (2026) sketches a literal Basel-for-AI mapping, which at minimum shows the genre is live; Leone De Castris and Thomas (2024) already read the AISI network as a nascent transgovernmental network in Slaughter’s sense. The disanalogy has to be priced alongside the borrowing, because Basel’s diffusion never ran on peer pressure alone. Its members are central banks and supervisors holding licensing, capital, and market-access leverage over the institutions the standards bind, sharing balance-sheet-level exposure to each other’s stability; implementation runs through supervisors with statutory teeth and is monitored by the Committee’s own consistency program. The AISI network holds none of this — no licensing power, no supervisory relationship with the labs, no shared exposure that makes one member’s laxity another’s emergency. Club standards without the club’s enforcement substrate propagate the drafting speed, not the compliance. Which is a further argument for the routing this paper predicts anyway: AI club standards travel farthest through financial supervisors, who already hold the leverage over the deployment layer, rather than through AI-native institutions still waiting to grow teeth. The price: Basel is also the false-assurance exhibit — the club exported a gameable standard worldwide with an implementation veneer (Jones, 2000), and club governance answers to no electorate, which is Chatham House’s own equity critique of the 2008 template (Wilkinson et al., 2026). A club that standardizes measurement (the registry’s schema, the census questions) rather than calibrated thresholds keeps most of the speed while exporting less of the gameability, since it is harder to arbitrage a question than a number.

6 Objections

Maybe the first crisis is biological, or cyber. Possibly. The claim is modal, and modality is the right register for institutional design under uncertainty: preparation compounds where the probability mass sits. Section 2.4’s retrospective runs show the axes behaving sensibly out of domain, but they are consistency checks, not proof. Notice, though, that the framework degrades gracefully. Even if the wheel lands elsewhere, the three axes still say *what happens next* (a bio event routes to biosecurity institutions slowly; a cyber event dissipates in contested attribution; a weapons event disappears into classification), and the riders principle and the registry are domain-portable by construction. The selection theory buys you the ranking; the design principles do not depend on the ranking being exactly right.

Maybe sectoral governance is fine — subsidiarity, after all. Sectoral regulators know their sectors; let each govern the AI inside it. The answer is that the failure modes that matter most have no sector. Kleinberg and Raghavan’s monoculture harm is a property of the population of deployers; Bommasani, Creel, Kumar, Jurafsky, and Liang (2022) show shared foundations homogenizing outcomes so that the same person is failed by every system simultaneously — a wrong that no single deployer commits and no sectoral examiner can see, what Creel and Hellman (2022) call the harm of the algorithmic Leviathan: standardized exclusion at scale. Subsidiarity assigns every sector its own microprudential supervisor and leaves the macro layer, the correlation structure across sectors, as everyone’s externality and no one’s mandate. Subsidiarity of exactly this shape is what missed 2008. Roberts, Hine, Taddeo, and Floridi (2024) are right that strengthening the existing regime complex is the feasible path, which is exactly why the strengthening has to include one deliberately cross-domain capability, ridden in on whichever vehicle moves first.

The registry needs the world it is supposed to create. This objection is stronger than most of the others, so it gets its own entry. The paper’s foil establishes that pre-crisis AI governance is deadlocked; the paper’s flagship artifact is operating infrastructure that must be built pre-crisis; drafts survive

deadlock, and infrastructure does not. If the deadlock premise is right, why does the registry get built? The answer is the scoping of Section 5.2, and I concede what it costs. The buildable-now registry runs on existing supervisory authority over regulated financial entities: data collection is what supervisors already do, and turning the Bank of England’s voluntary survey into mandatory standardized reporting is administrative action, not legislation, so it does not need the window the deadlock keeps shut. The full registry, reaching beyond the perimeter, is subject to the deadlock like everything else, and probably arrives, if it arrives, as a crisis-window ratification of a schema that already exists. That is the shape of the proposal: build the core under authority that exists, hold the schema for the rest, and accept that the beyond-perimeter layer is a shelf item in exactly the sense this paper uses the word.

It is worth naming the collision this creates with Section 5.3’s own requirement rather than letting the two pass in the night: the buildable core covers the supervised, *manageable* layer, and Section 5.3’s migration-proof monitoring demands reach to the base-model, compute, and settlement layers that move least — which is exactly the layer the buildable core cannot yet touch. The registry’s honest scope is therefore the mirror image of the migration problem: it sees best where the risk is least likely to flee, and it is blind precisely where Section 4.2 says the activity goes. That is not a hole to paper over; it is the drift objection’s institutional cousin, and the reason the schema has to be held ready for the window rather than assumed buildable now.

The CAT precedent cuts both ways here, and I have used both edges: ordered after its crisis, it took eight years and lost its funding in court (SEC, 2012b; U.S. Court of Appeals, 11th Cir., 2025); a registry whose schema, governance, and confidentiality design predate the crisis starts those eight years early.

Maybe there is no crisis at all. Then the deadlock literature’s premise fails with mine, and the shelf was cheap. Collingridge’s dilemma frames the bind — control is cheap early when knowledge is scarce and costly late when it arrives; the asymmetry I draw from it is that pre-positioning is paid in ordinary time and an empty window is paid in panic, at the exchange rate Section 4 documents. A registry, a charter clause, and a drafted delegation are inexpensive insurance against an event whose probability the rest of the field already treats as its planning assumption.

The objection I cannot fully answer. The scenario that worries me most is not the cascade; it is the non-event. Thousands of agents on shared base models, building correlated positions and correlated credit decisions and correlated procurement choices that unwind *slowly*: a drift in the aggregate statistics that never produces a Friday corpse, never clears the legibility bar, never opens any window at all. The two scenarios have distinct owned mechanisms, and neither should lean on the other’s evidence: the cascade branch runs on the speed and leverage channels documented by Danielsson and colleagues (Danielsson et al., 2022; Danielsson and Uthemann, 2025), while the drift branch runs on the monoculture quality channels (Kleinberg and Raghavan, 2021; Bommasani et al., 2022). Focusing events select for legibility, and everything in this paper inherits that selection: governance built on focusing events regulates the legible failure mode. Warnement Wrobel and McBeth (2025) supply fresh empirical teeth: attention to focusing events is measurably declining, narrative numbing raising the legibility bar year over year. The monoculture results say the harm does not need an event; the correlation measurements say the precondition is already in place (Kim et al., 2025; Goel et al., 2025). On this scenario Soares (2022) is right for a reason he did not give: the warning shot changes little because it never fires. Which inverts the emotional valence of the whole argument: the dramatic crisis, the one this paper predicts and the deadlock literature waits for, is the *optimistic* case, the world where the mechanism is at least shown in public once. The drift version is 2008: every institution individually sound by its own meter, the system building the position no one owned, and no examiner assigned to the aggregate. I do not have a governance design that manufactures legibility for slow correlated drift. The closest thing to an answer

is the registry again: exposure measurement is the only instrument in this paper that works identically in both scenarios, because a census does not need a corpse. Beyond that, the objection stands.

7 Conclusion

The first AI crisis will be governed by whoever already has teeth, and the teeth are in finance. That is the whole argument, and its two halves cut against both prevailing moods. Against the deadlock literature's hope: the crisis will not produce general AI governance, because focusing events are absorbed by the most mature receiving institution, and the most mature receiving institution has a sectoral mandate, a shelf full of sectoral drafts, and forty years of documented pathology that the response will import along with the speed. Against the skeptics' shrug: the crisis will produce a great deal of governance — fast, consequential, and lasting — though not the kind anyone designed, and the difference between a good and a bad version of it is being decided now, in what sits on the shelf.

The version of the claim I would most want tested sits in the table's fifteenth row. A cyber-finance hybrid — an AI-enabled attack or compromise whose losses print in public prices — is arguably the most probable near-term first event, and it is where the framework's axes point in different directions yet still issue a sharp call: the corpse lands on the mature regulator, the legible surface is financial, and the predicted response is market-structure rules for what is actually a security failure. If that event arrives and the rules that follow are scoped to the mechanism rather than the ticker, the framework is wrong in the most instructive way available. If they are scoped to the ticker, the wheel turned as described.

Crises institutionalize whatever ideas are lying around (Friedman, 2002 [1962], "Preface, 1982"); Birkland (2006) made the same point with data: pre-existing ideas win the window, and post-event learning is mostly superficial. The practical conclusion is almost embarrassingly concrete. Be the working paper on the shelf. Write the investigation body's charter with the cross-domain clause already in it, pre-negotiated and investigation-only, so the turf fight has nothing to feed on. Build the exposure registry's schema before anyone is panicking, so that the regulator's first question (*how many firms are running this model with trading authority right now?*) has an answer other than silence, because the silence is what converts a focusing event into panic governance. Draft the delegation, the sunset, and the suspension authority with the panic test applied in advance. None of this requires a summit.

And staff for the morning after. Boin, 't Hart, Stern, and Sundelius (2016) observe that crisis response is made by whoever can do the five tasks when the phone rings; in this crisis, the scarce resource will be people bilingual in both stacks — who can read a system card *and* an SR 11-7 validation report, who know what "effective challenge" means in a model-risk file and what a population-level evaluation would have to measure, who can tell a securities lawyer what an authority envelope is and tell an alignment researcher what an examination is. Bank model-risk functions and supervisory quants read the first dialect fluently, and frontier-evaluation researchers read the second; almost nobody is trained in model-risk supervision and frontier-model evaluation at once. The window, when it opens, will be staffed by whoever is — and, per the political economy of the moment, the response they staff will have to work within a growth regime that has made pause-style interventions politically unavailable.¹

¹AI-related investment has become a load-bearing component of US output growth: published commentary attributes a substantial share of recent measured GDP growth to it. Furman (2025) calculates from Bureau of Economic Analysis data that information-processing equipment and software, about four percent of GDP, accounted for roughly 92 percent of annualized real GDP growth in the first half of 2025, with output excluding those categories growing at about 0.1 percent annualized (Furman, 2025; U.S. Bureau of Economic Analysis, 2025). The magnitude is contested — Goldman Sachs' Jan Hatzius reads AI's contribution to 2025 growth as close to zero, and St. Louis Fed analysis normalizing by the third quarter of 2025 finds a smaller share (Rubinton and Patro, 2026) — so the figure enters here as attributed commentary, not settled fact. The enforcement asymmetry is already citable: Chatham House juxtaposes on the order of \$527 billion in 2026 hyperscaler capital expenditure against roughly €1 billion in EU AI Act enforcement capacity (Wilkinson et al., 2026). Responses that threaten

The wheel is loaded. The spin is not scheduled. Everything in between is design time.

Declarations

Conflict of Interest. The author declares no competing interests.

Funding. This research received no external funding.

Data Availability. This is a qualitative policy-analysis paper; no dataset was generated. The coded rubric table (Table in Section 2.2) is reproduced in full within the paper; its scoring is ordinal and qualitative, drawn from the cited primary sources (regulatory orders, statutes, and official reports), not from a compiled dataset.

AI Assistance. Claude (Anthropic; Claude Fable 5) was used as a research collaborator for literature synthesis, source verification, LaTeX preparation, and iterative refinement of the argument. ChatGPT (OpenAI; GPT-5.6 “Sol”) provided adversarial review and cross-checking. Automated review and verification tools were also used, including reviewer3.com and openaireview.com (referee-style review passes) and scite (citation verification), among other tools. All intellectual claims, judgments, and errors remain the author’s sole responsibility.

References

- Ada Lovelace Institute. Regulating AI in the UK. Report, 7 2023. URL <https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/>.
- Ada Lovelace Institute. Ada Lovelace Institute statement on the UK’s approach to AI regulation. Press release (Michael Birtwistle, Associate Director, commenting on the Government’s response to the AI Regulation white paper consultation), 2 2024. URL <https://www.adalovelaceinstitute.org/press-release/statement-on-uk-ai-regulation/>.
- A. Admati and M. Hellwig. *The Bankers’ New Clothes: What’s Wrong with Banking and What to Do about It*. Princeton University Press, expanded edition, 2024. ISBN 9780691251707.
- Aircraft Certification, Safety, and Accountability Act. Aircraft certification, safety, and accountability act. Pub. L. No. 116-260, Div. V, Tit. I, 134 Stat. 1182, 12 2020. URL <https://www.govinfo.gov/app/details/PLAW-116publ260>.
- E. J. Balleisen, L. S. Benneer, K. D. Krawiec, and J. B. Wiener, editors. *Policy Shock: Recalibrating Risk and Regulation after Oil Spills, Nuclear Accidents and Financial Crises*. Cambridge University Press, 2017. doi: 10.1017/9781316492635.
- Bank of England. UK financial regulators to begin overseeing critical third parties announced by HM Treasury. News release, 7 2026. URL <https://www.bankofengland.co.uk/news/2026/july/uk-financial-regulators-to-begin-overseeing-critical-third-parties-announced-by-hmt>.

the capex channel are off the table for reasons that have nothing to do with safety analysis; sectoral market-integrity regulation, which does not threaten it, is not. This is one more force loading the wheel.

- Bank of England FPC. Financial stability in focus: AI in the financial system. Bank of England FPC, 4 2025. URL <https://www.bankofengland.co.uk/financial-stability-in-focus/2025/april-2025>.
- Bank of England PRA. SS1/23 — model risk management principles for banks. Bank of England PRA; effective May 2024, 2023. URL <https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks-ss>.
- P. Barnett and A. Scher. AI governance to avoid extinction: The strategic landscape and actionable research questions. *arXiv preprint arXiv:2505.04592*, 2025.
- M. S. Barr. Review of the Federal Reserve’s supervision and regulation of Silicon Valley Bank. Federal Reserve Board, 4 2023. URL <https://www.federalreserve.gov/publications/files/svb-review-20230428.pdf>.
- M. S. Barr and G. P. Miller. Global administrative law: The view from Basel. *European Journal of International Law*, 17(1):15–46, 2006. doi: 10.1093/ejil/chi167.
- F. R. Baumgartner and B. D. Jones. *Agendas and Instability in American Politics*. University of Chicago Press, 2nd edition, 2009. ISBN 9780226039534.
- A. Beber and M. Pagano. Short-selling bans around the world: Evidence from the 2007–09 crisis. *The Journal of Finance*, 68(1):343–381, 2013. doi: 10.1111/j.1540-6261.2012.01802.x.
- J. Begin et al. Preference optimization drives monoculture in LLM prediction markets. *arXiv preprint arXiv:2606.26583*, 2026. Unreviewed v1; cited as illustrative.
- M. Behn, R. Haselmann, and V. Vig. The limits of model-based regulation. *Journal of Finance*, 77(3): 1635–1684, 2022. doi: 10.1111/jofi.13124.
- Y. Bengio et al. International AI safety report. *arXiv preprint arXiv:2501.17805*, 2025.
- D. M. Berkovitz. Statement on withdrawal of regulation AT. U.S. Commodity Futures Trading Commission, 6 2020. URL <https://www.cftc.gov/PressRoom/SpeechesTestimony/berkovitzstatement062520>.
- W. A. Birdthistle and M. T. Henderson. Becoming a fifth branch. *Cornell Law Review*, 99(1):1, 2013.
- T. A. Birkland. *After Disaster: Agenda Setting, Public Policy, and Focusing Events*. Georgetown University Press, 1997. ISBN 9780878406531.
- T. A. Birkland. Focusing events, mobilization, and agenda setting. *Journal of Public Policy*, 18(1): 53–74, 1998. doi: 10.1017/S0143814X98000038.
- T. A. Birkland. *Lessons of Disaster: Policy Change after Catastrophic Events*. Georgetown University Press, 2006. ISBN 9781589011212.
- BIS. Artificial intelligence and the economy: Implications for central banks. In *BIS Annual Economic Report*, chapter III. 6 2024. URL <https://www.bis.org/publ/arpdf/ar2024e3.htm>.
- BIS Markets Committee. FX execution algorithms and market functioning. Markets Committee Papers 13, Bank for International Settlements, 10 2020. URL <https://www.bis.org/publ/mktc13.htm>.

- E. Boehmer, C. M. Jones, and X. Zhang. Shackling short sellers: The 2008 shorting ban. *The Review of Financial Studies*, 26(6):1363–1400, 2013. doi: 10.1093/rfs/hht017.
- A. Boin, P. ’t Hart, and A. McConnell. Crisis exploitation: Political and policy impacts of framing contests. *Journal of European Public Policy*, 16(1):81–106, 2009. doi: 10.1080/13501760802453221.
- A. Boin, P. ’t Hart, E. Stern, and B. Sundelius. *The Politics of Crisis Management: Public Leadership under Pressure*. Cambridge University Press, 2nd edition, 2016. doi: 10.1017/9781316339756.
- R. Bommasani, K. A. Creel, A. Kumar, D. Jurafsky, and P. Liang. Picking on the same person: Does algorithmic monoculture lead to outcome homogenization? In *Advances in Neural Information Processing Systems (NeurIPS 2022)*, 2022. arXiv:2211.13972.
- R. Bommasani et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.
- M. D. Bordo and J. V. Duca. The impact of the Dodd-Frank act on small business. NBER Working Paper 24501, National Bureau of Economic Research, 2018.
- C. Brummer. *Soft Law and the Global Financial System: Rule Making in the 21st Century*. Cambridge University Press, 2nd edition, 2015.
- G. Buchak, G. Matvos, T. Piskorski, and A. Seru. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483, 2018. doi: 10.1016/j.jfineco.2018.03.011.
- California DFPI. In the matter of Silicon Valley Bank: Order taking possession of property and business, 3 2023. URL <https://dfpi.ca.gov/wp-content/uploads/sites/337/2023/03/DFPI-Orders-Silicon-Valley-Bank-03102023.pdf>.
- E. Calvano, G. Calzolari, V. Denicolò, and S. Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10):3267–3297, 2020. doi: 10.1257/aer.20190623.
- S. G. Cecchetti, R. L. Lumsdaine, T. Peltonen, and A. Sánchez Serrano. Artificial intelligence and systemic risk. Report 16, ESRB Advisory Scientific Committee, 12 2025. URL https://www.esrb.europa.eu/pub/pdf/asc/esrb.ascreport202512_AIandsystemicrisk.en.pdf.
- CFTC. Regulation automated trading: Notice of proposed rulemaking. PR 7283-15; 80 FR 78824, 11 2015. URL <https://www.cftc.gov/PressRoom/PressReleases/7283-15>.
- CFTC. Court orders U.K. resident Navinder Singh Sarao to pay more than \$38 million. PR 7486-16, 11 2016. URL <https://www.cftc.gov/PressRoom/PressReleases/7486-16>.
- CFTC. Electronic trading risk principles proposal and regulation AT withdrawal. 85 FR 42761 / 42755, 7 2020. URL <https://www.cftc.gov/LawRegulation/FederalRegister/proposedrules/2020-14381.html>.
- CFTC and SEC Staffs. Findings regarding the market events of May 6, 2010: Report of the staffs of the CFTC and SEC to the joint advisory committee on emerging regulatory issues. Technical report, U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission, 9 2010. URL <https://www.sec.gov/news/studies/2010/marketevents-report.pdf>.

- P. Cihon, M. M. Maas, and L. Kemp. Should artificial intelligence governance be centralised? Design lessons from history. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES 2020)*, 2020. arXiv:2001.03573.
- CNBC. 'flash crash' trader Navinder Singh Sarao sentenced to home detention, 1 2020. URL <https://www.cnbc.com/2020/01/29/flash-crash-trader-navinder-singh-sarao-sentenced-to-home-detention.html>.
- Jr. Coffee, J. C. The political economy of Dodd-Frank: Why financial reform tends to be frustrated and systemic risk perpetuated. *Cornell Law Review*, 97(5):1019, 2012.
- D. Collingridge. *The Social Control of Technology*. Frances Pinter / St. Martin's Press, 1980. ISBN 0903804727.
- K. Creel and D. Hellman. The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems. *Canadian Journal of Philosophy*, 52(1):26–43, 2022. doi: 10.1017/can.2022.3.
- J. Cunningham. The limits of the bank service company act. *Duke Law Journal*, 74(1):227–268, 2024. URL <https://scholarship.law.duke.edu/dlj/vol74/iss1/4/>.
- J. Danielsson and A. Uthemann. Artificial intelligence and financial crises. *Journal of Financial Stability*, 80:101453, 2025. doi: 10.1016/j.jfs.2025.101453.
- J. Danielsson, R. Macrae, and A. Uthemann. Artificial intelligence and systemic risk. *Journal of Banking & Finance*, 140:106290, 2022. doi: 10.1016/j.jbankfin.2021.106290.
- S. M. Davidoff and D. Zaring. Regulation by deal: The government's response to the financial crisis. *Administrative Law Review*, 61(3):463, 2009.
- R. A. DeLeo. Anticipatory-conjectural policy problems: A case study of the avian influenza. *Risk, Hazards & Crisis in Public Policy*, 1(1):147–184, 2010. doi: 10.2202/1944-4079.1002.
- R. A. DeLeo. *Anticipatory Policymaking: When Government Acts to Prevent Problems and Why It Is So Difficult*. Routledge, 2016. ISBN 9781138307476.
- R. A. DeLeo, K. Taylor, D. A. Crow, and T. A. Birkland. During disaster: Refining the concept of focusing events to better explain long-duration crises. *International Review of Public Policy*, 3(1), 2021. doi: 10.4000/irpp.1868.
- R. B. L. Dixon and H. Frase. AI incidents: Key components for a mandatory reporting regime. Technical report, Center for Security and Emerging Technology, 2025. URL <https://cset.georgetown.edu/publication/ai-incidents-key-components-for-a-mandatory-reporting-regime/>.
- DOJ. Futures trader pleads guilty to illegally manipulating the futures market in connection with 2010 'flash crash'. Press release, 11 2016.
- W. W. Dou, I. Goldstein, and Y. Ji. AI-powered trading, algorithmic collusion, and price efficiency. NBER Working Paper 34054, National Bureau of Economic Research, 2025.

- J. Egan. The United States must avoid AI's Chernobyl moment. Center for a New American Security / Just Security commentary, 3 2025. URL <https://www.cnas.org/publications/commentary/the-united-states-must-avoid-ais-chernobyl-moment>.
- ESAs. The European Supervisory Authorities designate critical ICT third-party providers under the Digital Operational Resilience Act. Press release, 11 2025. URL <https://www.esa.europa.eu/publications-and-media/press-releases/european-supervisory-authorities-designate-critical-ict-third-party-providers-under-digital>.
- European Parliament and Council. Regulation (EU) 2022/2554 (digital operational resilience act, DORA). OJ L 333, 27.12.2022, 2022. URL <http://data.europa.eu/eli/reg/2022/2554/oj>. Cited: Art. 28 (register of information), Arts. 31–44 (oversight framework for critical ICT third-party providers).
- European Parliament and Council. Regulation (EU) 2024/1689 (artificial intelligence act). OJ L, 2024/1689, 12.7.2024, 2024. URL <http://data.europa.eu/eli/reg/2024/1689/oj>. Cited: Art. 7 (delegated amendment of Annex III), Art. 71 (high-risk-system database), Art. 73 (reporting of serious incidents), Art. 93 (Commission requests to general-purpose AI model providers).
- FAA. Operators of Boeing Company Model 737-8 and Model 737-9 airplanes: Emergency order of prohibition. Docket FAA-2019-0200; 84 Fed. Reg. 9705 (18 March 2019), FR Doc. 2019-05067, 3 2019. URL <https://www.federalregister.gov/documents/2019/03/18/2019-05067/>.
- FAA. Boeing 737 MAX timeline. Rev 201118; grounding order rescinded 18 November 2020, 2020. URL https://www.faa.gov/sites/faa.gov/files/2022-08/737_MAX_Timeline.pdf.
- M. Farzulla. The axiom of consent: Friction dynamics in multi-agent coordination. *arXiv preprint arXiv:2601.06692*, 2026a. doi: 10.48550/arXiv.2601.06692.
- M. Farzulla. Do cryptocurrency markets differentiate infrastructure from regulatory shocks? A multi-moment event study with dependence-robust inference, 2026b.
- M. Farzulla and A. Maksakov. ASRI: An aggregated systemic risk index for cryptocurrency markets. *arXiv preprint arXiv:2602.03874*, 2026. doi: 10.48550/arXiv.2602.03874.
- FCA, Bank of England, and PRA. PS24/16: Operational resilience: Critical third parties to the UK financial sector. Rules in force 1 January 2025, 11 2024. URL <https://www.fca.org.uk/publications/policy-statements/ps24-16-operational-resilience-critical-third-parties-uk-financial-sector>.
- FDA. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Press release; De Novo DEN180001 (IDx-DR), 4 2018. URL <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>.
- FDIC. FDIC community banking study 2020, 2020. URL <https://www.fdic.gov/news/press-releases/2020/pr20139.html>.
- Federal Reserve Board. SR 11-7: Supervisory guidance on model risk management, 4 2011. URL <https://www.federalreserve.gov/boarddocs/srletters/2011/sr1107.htm>.

- Federal Reserve Board. SR 26-2: Revised guidance on model risk management. Interagency supervisory guidance superseding SR 11-7, issued with the Office of the Comptroller of the Currency and the FDIC; footnote 3 places generative and agentic AI outside its scope, 4 2026. URL <https://www.federalreserve.gov/supervisionreg/srletters/sr2602.htm>.
- E. Fielding, A. W. Lo, and J. H. Yang. The national transportation safety board: A model for systemic risk management. Working paper, MIT, 2010. URL http://web.mit.edu/Alo/www/Papers/nts_b.html.
- Financial Stability Board. The financial stability implications of artificial intelligence, 11 2024. URL <https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/>.
- Financial Stability Board. Global monitoring report on non-bank financial intermediation 2025, 12 2025a. URL <https://www.fsb.org/2025/12/global-monitoring-report-on-nonbank-financial-intermediation-2025/>.
- Financial Stability Board. Monitoring adoption of artificial intelligence and related vulnerabilities in the financial sector, 10 2025b. URL <https://www.fsb.org/2025/10/monitoring-adoption-of-artificial-intelligence-and-related-vulnerabilities-in-the-financial-sector/>.
- S. Fish, Y. A. Gonczarowski, and R. I. Shorrer. Algorithmic collusion by large language models. *arXiv preprint arXiv:2404.00806*, 2024. Working paper, v5 March 2026.
- M. Friedman. *Capitalism and Freedom*. University of Chicago Press, fortieth anniversary edition, 2002. ISBN 9780226264219. Quotation from “Preface, 1982,” pp. xiii–xiv; originally published 1962.
- FSOC. Annual report 2023. U.S. Department of the Treasury, 2023. URL <https://home.treasury.gov/system/files/261/FSOC2023AnnualReport.pdf>.
- J. Furman. Investment in information-processing equipment and software as a share of H1-2025 US GDP growth. X post, 9 2025. URL <https://x.com/jasonfurman/status/1971995367202775284>.
- G. Gensler and L. Bailey. Deep learning and financial stability. Working Paper 3723132, SSRN, 2020.
- S. Goel, J. Struber, et al. Great models think alike and this undermines AI oversight. In *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*, 2025. arXiv:2502.04313.
- A. Gupta, G. Pappyshev, and J. T. Kwok. Bubble, bubble, AI’s rumble: Why global financial regulatory incident reporting is our shield against systemic stumbles. *arXiv preprint arXiv:2509.26150*, 2025.
- A. G. Haldane and V. Madouros. The dog and the frisbee. Speech at the Federal Reserve Bank of Kansas City Economic Policy Symposium, Jackson Hole, 8 2012. URL <https://www.bis.org/reviuew/r120905a.pdf>.
- L. Ho, J. Barnhart, R. Trager, Y. Bengio, M. Brundage, et al. International institutions for advanced AI. *arXiv preprint arXiv:2307.04699*, 2023.
- IMF. Advances in artificial intelligence: Implications for capital market activities. In *Global Financial Stability Report*, chapter 3. 10 2024.

- R. M. Irani, R. Iyer, R. R. Meisenzahl, and J.-L. Peydró. The rise of shadow banking: Evidence from capital regulation. *Review of Financial Studies*, 34(5):2181–2235, 2021. doi: 10.1093/rfs/hhaa106.
- D. Jones. Emerging problems with the Basel capital accord: Regulatory capital arbitrage and related issues. *Journal of Banking & Finance*, 24(1-2):35–58, 2000. doi: 10.1016/S0378-4266(99)00052-7.
- E. Kim, A. Garg, K. Peng, and N. Garg. Correlated errors in large language models. In *Proceedings of the 42nd International Conference on Machine Learning (ICML 2025)*, 2025. arXiv:2506.07962.
- J. W. Kingdon. *Agendas, Alternatives, and Public Policies*. Longman, updated 2nd edition, 2011. ISBN 9780205000869.
- J. Kleinberg and M. Raghavan. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22), 2021. doi: 10.1073/pnas.2018340118.
- R. Knake, A. Shostack, and T. Wheeler. Learning from cyber incidents: Adapting aviation safety models to cybersecurity. Technical report, Belfer Center for Science and International Affairs, Harvard Kennedy School, 2021.
- T. Kuran and C. R. Sunstein. Availability cascades and risk regulation. *Stanford Law Review*, 51(4): 683–768, 1999. doi: 10.2307/1229439.
- A. Leone De Castris and C. Thomas. The potential functions of an international institution for AI safety: Insights from adjacent policy areas and recent trends. *arXiv preprint arXiv:2409.10536*, 2024.
- Magnus Lundgren and Jonas Tallberg. Governing artificial intelligence: Public preferences and regulatory options. arXiv:2607.14585, 2026.
- M. M. Maas and J. J. Villalobos. International AI institutions: A literature review of models, examples, and proposals. Technical report, Legal Priorities Project, 2023.
- G. E. Marchant, B. R. Allenby, and J. R. Herkert, editors. *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*. Springer, 2011. doi: 10.1007/978-94-007-1356-7.
- M. Mariathasan and O. Merrouche. The manipulation of Basel risk-weights. *Journal of Financial Intermediation*, 23(3):300–321, 2014. doi: 10.1016/j.jfi.2014.04.004.
- J. Maues. Banking act of 1933 (Glass-Steagall). Federal Reserve History, 2013. URL <https://www.federalreservehistory.org/essays/glass-steagall-act>.
- S. McGregor. Preventing repeated real world AI failures by cataloging incidents: The AI incident database. In *Proceedings of the Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-21)*, 2021. arXiv:2011.08512.
- P. Mehta. Macro-prudential AI governance: A two-layer early warning and response system for frontier AI. *arXiv preprint arXiv:2607.03542*, 2026.
- MetLife, Inc. v. FSOC. 177 f. supp. 3d 219 (d.d.c. 30 march 2016). No. 1:15-cv-00045, Collyer, J. — designation rescinded as arbitrary and capricious, 2016. URL https://www.govinfo.gov/app/details/USCOURTS-dcd-1_15-cv-00045.

- OECD. Defining AI incidents and related terms. Technical report, OECD Artificial Intelligence Papers, 2024.
- OECD.AI. OECD AI incidents and hazards monitor (AIM). Live platform, 2026. URL <https://oecd.ai/en/incidents>.
- A. Ortega. AI threats to national security can be countered through an incident regime. *arXiv preprint arXiv:2503.19887*, 2025.
- H. Peirce, I. Robinson, and T. Stratmann. How are small banks faring under Dodd-Frank? Working Paper 14-05, Mercatus Center, George Mason University, 2014.
- K. Peng and N. Garg. Monoculture in matching markets. In *Advances in Neural Information Processing Systems (NeurIPS 2024)*, 2024. arXiv:2312.09841.
- B. D. Quintenz. Statement of support regarding proposed rule on electronic trading risk principles. U.S. Commodity Futures Trading Commission, 6 2020. URL <https://www.cftc.gov/PressRoom/SpeechesTestimony/quintenzstatement062520>.
- H. Roberts, E. Hine, M. Taddeo, and L. Floridi. Global AI governance: Barriers and pathways forward. *International Affairs*, 100(3):1275–1286, 2024. doi: 10.1093/ia/iiae073.
- R. Romano. The political dynamics of derivative securities regulation. *Yale Journal on Regulation*, 14 (2):279–406, 1997. URL <https://openyls.law.yale.edu/handle/20.500.13051/7952>.
- R. Romano. Regulating in the dark and a postscript assessment of the iron law of financial regulation. *Hofstra Law Review*, 43(1), 2014. art. 2.
- H. Rubinton and B. A. Patro. Tracking AI’s contribution to GDP growth. St. Louis Fed On the Economy, 1 2026. URL <https://www.stlouisfed.org/on-the-economy/2026/jan/tracking-ai-contribution-gdp-growth>.
- D. S. Schiff. Looking through a policy window with tinted glasses: Setting the agenda for U.S. AI policy. *Review of Policy Research*, 40(5):729–756, 2023. doi: 10.1111/ropr.12535.
- SEC. Emergency order pursuant to section 12(k)(2) of the securities exchange act of 1934 taking temporary action to respond to market developments. Release 34-58592, 9 2008. URL <https://www.sec.gov/files/rules/other/2008/34-58592.pdf>.
- SEC. Risk management controls for brokers or dealers with market access. Rule 15c3-5; Release 34-63241, 2010a. URL <https://www.sec.gov/rules/final/2010/34-63241.pdf>.
- SEC. SEC approves new stock-by-stock circuit breaker rules. PR 2010-98, 6 2010b. URL <https://www.sec.gov/news/press/2010/2010-98.htm>.
- SEC. SEC approves proposals to address extraordinary volatility in individual stocks and broader stock market. PR 2012-107, 5 2012a. URL <https://www.sec.gov/newsroom/press-releases/2012-2012-107htm>.
- SEC. SEC approves rule 613: Consolidated audit trail. PR 2012-134, 7 2012b.

SEC. In the matter of Knight Capital Americas LLC. Release 34-70694, 10 2013. URL <https://www.sec.gov/litigation/admin/2013/34-70694.pdf>.

SEC. Staff report on algorithmic trading in U.S. capital markets. Technical report, 8 2020. URL https://www.sec.gov/files/algo_trading_report_2020.pdf.

SEC/Investor.gov. The laws that govern the securities industry. U.S. Securities and Exchange Commission, 2024. URL <https://www.investor.gov/introduction-investing/investing-basics/role-sec/laws-govern-securities-industry>.

T. Shaffer Shane. AI incident reporting: Addressing a gap in the UK’s regulation of AI. Technical report, Centre for Long-Term Resilience, 6 2024.

A.-M. Slaughter. *A New World Order*. Princeton University Press, 2004.

N. Soares. Warning shots probably wouldn’t change the picture much. LessWrong / Machine Intelligence Research Institute, 10 2022.

The Hill. AI-caused financial crisis ‘nearly unavoidable’ without regulation: SEC chief. Reporting an interview Gensler gave the Financial Times, 10 2023. URL <https://thehill.com/business/4258889-ai-caused-financial-crisis-nearly-unavoidable-without-regulation-sec-chief/>.

The White House. Press briefing on the attribution of the WannaCry malware attack to North Korea. Homeland Security Advisor Thomas P. Bossert, 12 2017. URL <https://trumpwhitehouse.archives.gov/briefings-statements/press-briefing-on-the-attribution-of-the-wannacry-malware-attack-to-north-korea-121917/>.

C. Toups, R. Bommasani, K. A. Creel, S. H. Bana, D. Jurafsky, and P. Liang. Ecosystem-level analysis of deployed machine learning reveals homogeneous outcomes. In *Advances in Neural Information Processing Systems (NeurIPS 2023)*, 2023. arXiv:2307.05862.

R. Trager, B. Harack, A. Reuel, et al. International governance of civilian AI: A jurisdictional certification approach. *arXiv preprint arXiv:2308.15514*, 2023.

U.S. Bureau of Economic Analysis. Gross domestic product, 2nd quarter 2025 (third estimate), GDP by industry, and corporate profits, 9 2025. URL <https://www.bea.gov/news/2025/gross-domestic-product-2nd-quarter-2025-third-estimate-gdp-industry-corporate-profits>.

U.S. Congress. Bank service company act. URL <https://uscode.house.gov/view.xhtml?req=granuleid%3AUSC-prelim-title12-section1867&num=0&edition=prelim>. 12 U.S.C. §1867(c) (regulation and examination of performance of bank services by third parties).

U.S. Court of Appeals, 11th Cir. American securities association & citadel securities v. SEC. No. 23-13396, 7 2025. URL <https://media.ca11.uscourts.gov/opinions/pub/files/202313396.pdf>.

U.S. Department of Defense. DoD directive 3000.09: Autonomy in weapon systems. Office of the Under Secretary of Defense for Policy; effective 25 January 2023, 2023. URL <https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf>.

- U.S. Department of the Interior. Secretarial order no. 3299: Establishment of the Bureau of Ocean Energy Management, the Bureau of Safety and Environmental Enforcement, and the Office of Natural Resources Revenue. Signed by Secretary Ken Salazar, 5 2010. URL <https://www.doi.gov/document-library/secretary-order/3299-establishment-bureau-ocean-energy-management-bureau-safety>.
- U.S. Department of the Interior. Interior department completes reorganization of the former MMS. Press release, 10 2011. URL <https://www.doi.gov/news/pressreleases/Interior-Department-Completes-Reorganization-of-the-Former-MMS>.
- U.S. Department of the Treasury. Artificial intelligence in financial services. Technical report, 12 2024. URL <https://home.treasury.gov/system/files/136/Artificial-Intelligence-in-Financial-Services.pdf>.
- U.S. Treasury, Federal Reserve, and FDIC. Joint statement by Treasury, Federal Reserve, and FDIC. SVB/Signature systemic risk exception, 3 2023. URL <https://www.federalreserve.gov/newsevents/pressreleases/monetary20230312b.htm>.
- U.S. Treasury/FSOC. Designations (nonbank financial company determinations under Dodd-Frank § 113, 12 U.S.C. § 5323), 2024. URL <https://home.treasury.gov/policy-issues/financial-markets-financial-institutions-and-fiscal-service/fsoc/designations>.
- R. Van Loo. Regulatory monitors: Policing firms in the compliance era. *Columbia Law Review*, 119: 369, 2019.
- L. Velasco, C. Martinet, H. de Zoete, R. Trager, et al. The future of the AI summit series. *arXiv preprint arXiv:2601.02383*, 2026.
- J. Vipra and A. Korinek. Market concentration implications of foundation models: The economics of foundation models. *arXiv preprint arXiv:2311.01550*, 2023.
- M. K. Warnement Wrobel and M. K. McBeth. Focusing events, agenda setting, and narrative numbing. *Review of Policy Research*, 43(3):e70060, 2025. doi: 10.1111/ropr.70060.
- K. Wei and L. Heim. Designing incident reporting systems for harms from general-purpose AI. *arXiv preprint arXiv:2511.05914*, 2025.
- J. Whittlestone and J. Clark. Why and how governments should monitor AI development. *arXiv preprint arXiv:2108.12427*, 2021.
- R. Wilkinson, A. Krasodomski, I. Wilkinson, and F. J. Varela Sandoval. Breaking the deadlock on AI governance: How a crisis could lead to global coordination. Research paper, Chatham House, 3 2026.