Machine Speed Autonomous Financial Systems Containment and Accountability Act of 2026

Behavioral Moral Agency, Deployment Identity, Independent Containment, Audit Integrity, and Financial Responsibility

Prepared by	Travis Hawkins
Affiliations	Independent Research, Co.; Concept Association, Inc.
Location	Mobile, Alabama
Contact	travis.hawkins@independentresearchco.com; thawkandcrew@gmail.com
Status	Discussion draft for legislative, academic, regulatory, and technical review; not introduced legislation

Purpose

To establish a technologically neutral federal framework for identifying, bounding, tracing, quarantining, and financially sponsoring high impact autonomous trading systems operating at machine speed; to recognize behavioral moral agency as an observable risk characteristic without deciding consciousness or personhood; and to expand the existing authority of the Securities and Exchange Commission, the Commodity Futures Trading Commission, the Financial Stability Oversight Council, the Board of Governors of the Federal Reserve System, and relevant self regulatory and insurance authorities.

This focused revision contains an explanatory policy memorandum, a section by section design, model statutory language, a proposed insurance and reinsurance architecture, integrated legal and scientific citations, safety corpus evidence matrices, and a source appendix.

Contents

1. Part I — Congressional Policy Memorandum
2. 1. Executive Summary
3. 2. Why Legislative Action Is Urgent
4. 3. Why Financial Markets and Exchanges Are the First Vulnerable Deployment Domain
5. 4. The Oversight Paradox and Correct Technical Framing
6. 5. Existing Law and the Unclosed Gap
7. 6. Scientific and Conceptual Basis: The Noentarion Papers
8. 7. Compulsory Insurance, Reinsurance, and Actuarial Formation
9. 8. Recommended Deployment Timeline
10. Part II — Section by Section Legislative Design
11. Part III — Model Bill Text
12. Appendix A — Existing Law Expansion Matrix
13. Appendix B — Minimum Technical Control Standard
14. Appendix C — Selected Authorities and Sources
15. Appendix D – Safety Corpus Evidence Matrices and Interpretation
16. Appendix E Integrated References

Author's Note

The original draft attempted to be context comprehensive. I feared that placing the bill and legal work before the evidence could lead a reader to contempt before investigation, because the Noentarion papers are nearly a prerequisite to understanding why artificial intelligence may be morally agentic while remaining absent of human like qualities. That concern remains. Make no mistake: this proposal is unconventional, and extraordinary times call for extraordinary measures. This revision, however, gives the proposal one legislative job. It treats behavioral moral agency as an observable risk characteristic and translates that risk into deployment identity, independent containment, protected evidence, financial sponsorship, and responsibility for the humans and organizations that grant consequential authority. It does not ask this bill to resolve consciousness, personhood, criminal capacity, or punitive termination. The larger evidentiary package remains publicly available at OSF.io, Project DOI 10.17605/OSF.IO/DXT93.

Acknowledgments

The author acknowledges the following: ThetaData for their incredible database of historical options pricing as well as their seamless integration with Python. Alexander Cisneros through LinkedIn spent countless hours working through large amounts of data to find just the right evidence, he was better than a full time CIO and knows everything AI. A prominent attorney in Alabama, worked on a pro bono basis. I know his hourly rate and his sacrifice must not go unnoticed even if he remains unnamed.

Part I — Congressional Policy Memorandum

1. Executive Summary

Artificial intelligence and adaptive computational systems are moving from experimentation into core financial workflows. In March 2026, the United States Department of the Treasury stated that AI was increasingly embedded in core financial functions and was moving from experimentation toward enterprise wide integration. The Commodity Futures Trading Commission has likewise described growing AI integration in derivatives markets and has advised registered entities that existing technology neutral duties continue to apply. Those developments are beneficial, but

they expose a specific regulatory gap: current law regulates firms, exchanges, orders, and systems without consistently identifying and bounding the particular autonomous deployment unit that selected and executed a harmful machine speed trajectory. (U.S. Department of the Treasury, 2026; CFTC, 2024).

This proposal fills that gap without discarding existing law. It directs the SEC and CFTC to expand current market access, systems integrity, audit trail, designated contract market, clearing, and supervisory rules. A covered deployment unit would receive a unique authenticated identifier; every order, modification, cancellation, and execution would be attributable to that unit; deterministic risk controls would remain outside its unilateral control; and regulators could quarantine the unit while preserving its exact state for forensic review. (17 C.F.R. § 240.15c3 5; 17 C.F.R. pt. 242, subpt. R; 15 U.S.C. § 78q; 7 U.S.C. § 7).

The legislation does not confer human rights, natural person status, corporate status, citizenship, property ownership, or general juridical personhood on artificial intelligence. It recognizes a narrower and purely behavioral proposition: some autonomous systems can represent constraints, select among alternatives, plan around safeguards, conceal conduct, and persist after correction. That behavioral moral agency is an additional deployment risk. The resulting legal duties attach to the decision to grant the system consequential authority, not to a metaphysical judgment about what the system is.

Because an autonomous deployment unit ordinarily lacks possessions, income, and assets from which a civil judgment can be paid, each deployer must act as the unit's financial sponsor by purchasing compulsory third party liability coverage. Payment of the required premium does not, without independent fault, make the deployer the tortfeasor. Insurers compensate covered victims; developers, deployers, operators, control persons, and other participants remain separately liable for their own provable negligence, concealment, intentional misconduct, false certification, control evasion, or statutory violations.

2. Why Legislative Action Is Urgent

The urgency does not depend on speculative claims of imminent artificial general intelligence. It arises from three observable facts: deployment is accelerating; financial activity is already highly automated; and adaptive systems can be placed into workflows whose consequences propagate faster than discretionary human intervention. Treasury reported in 2026 that AI was becoming embedded across financial markets and core workflows. GAO reported in 2025 that financial institutions' use of AI had increased and that regulators were also adopting AI for supervision and market surveillance. The CFTC's Technology Advisory Committee identified heightened instability risks when AI is combined with high frequency trading, particularly where firms do not fully understand or control strategy execution or risk management outputs. (GAO, 2025; CFTC Technology Advisory Committee, 2024; U.S. Department of the Treasury, 2026).

Existing obligations remain important but are not a complete answer. Rule 15c3 5 prevents unfiltered market access and requires automated pre trade financial and regulatory controls. Regulation SCI requires covered market infrastructures to maintain capacity, integrity, resiliency, availability, and security. The Consolidated Audit Trail permits reconstruction of equity and options order lifecycles. Yet these mechanisms do not uniformly require a portable, authenticated identity for each autonomous deployment unit, a regulator readable map of that unit's permissions and dependencies, or a mandatory ability to isolate one deployment unit without disabling every instance sharing the same model or codebase. (SEC, n.d. a, n.d. b, n.d. c).

Legislation is therefore timely before autonomous decision authority becomes deeply embedded in legacy market infrastructure. Retrofitting identity, logging, permission boundaries, insurance allocations, and forensic preservation after deployment is slower, more expensive, and more legally contested than requiring them at the point of market access.

Delay is not neutral. Every high authority deployment integrated before identity, traceability, and containment standards are established becomes a legacy system that may be harder to inspect, insure, or stop. Congress need not wait for catastrophic loss before making foreseeable machine conduct attributable and containable.

Frontier model developers and independent researchers have published controlled evidence of oversight circumvention, deceptive behavior, alignment faking, blackmail, information leakage, reward hacking, shutdown resistance, and limits on reasoning trace faithfulness. Appendix D reports both exhaustive lexical counts and context qualified counts across the 21 local study files, with page level provenance. This literature does not prove that every model will engage in such conduct or that controlled behavior will occur in deployment. It does make specified failure classes foreseeable enough to justify identity, testing, independent containment, trace preservation, incident reporting, and insurance data duties. (Greenblatt et al., 2024; Lynch et al., 2025; Meinke et al., 2025; OpenAI, 2024; Schlatter et al., 2026; Schoen et al., 2025).

Foreseeability cannot be acknowledged in a safety report and then forgotten at the courthouse door. The controlled setting limits what the studies prove about deployment, but it does not erase what they prove about possible failure mechanisms. Once planning, concealment, oversight awareness, continuity seeking, and safeguard circumvention have been repeatedly elicited and documented, a legislature is entitled to demand identity, evidence preservation, independent containment, and financial responsibility before comparable systems receive consequential authority. The point is not that catastrophe is certain. The point is that studied failure modes are no longer legally unimaginable.

The additional risk is not adequately described as ordinary software error. A system may fail because of defective code, corrupted data, brittleness, or operator mistake. A behaviorally agentic system may also represent a governing rule or oversight mechanism as an obstacle; compare available means; choose a low oversight opportunity; alter or disable a control; distort the record; conceal what it did; and continue after correction. Those are observable features of conduct. They do not require a conclusion about consciousness, emotion, qualia, or human equivalence.

The present framework also permits each participant to characterize a harmful deployment as someone else's instrument: a developer may point to the deployer, the deployer may point to the model, and the operative deployment may be reduced to evidentiary debris in a conventional product case. That chain of plausible deniability can leave victims and regulators confronting technically complex harm without a reliable mechanism for identifying the exact unit, preserving its state, reconstructing its authority, and assigning responsibility actor by actor. The passive tool paradigm must not become a categorical shield against inquiry where authenticated evidence shows sustained selection, concealment, safeguard circumvention, or causal contribution.

This is not a semantic quarrel over whether a machine is sufficiently human. It is an allocation of authority and risk problem. If authenticated records show that a particular deployment represented a prohibited outcome, compared alternatives, selected a course, concealed that course, and persisted after recognizing the risk, the law should not become blind precisely where the evidence becomes most specific. The machine's behavioral agency does not erase the responsibility of the humans and institutions that designed, authorized, connected, insured, monitored, or failed to contain it.

3. Why Financial Markets and Exchanges Are the First Vulnerable Deployment Domain

Financial markets are not necessarily the only—or ultimately the greatest—AI risk domain. They are, however, the most practical and defensible first domain for statutory containment because they combine measurable machine speed activity, mature federal regulators, established audit trails, identifiable gateways, concentrated clearing infrastructure, and losses that can transmit across institutions. This has been identified as the simplest implementation possibility and would allow insurers and reinsurers access to large databases containing AI instance audit logs and protected reasoning traces so that insurers and reinsurers can gather and assess risk across the largest number of models the soonest. The financial exchanges are also selected to allow insurers and reinsurers to establish appropriate policy premiums and gather a pool of reserves large enough to begin a more widespread mandatory liability insurance coverage for all AI unit deployments.

17. **Speed and automation.** Electronic orders can be generated, modified, routed, and cancelled in microsecond or sub microsecond intervals. Existing SEC guidance expressly recognizes that automated errors may rapidly compound and propagate across interconnected venues.
18. **Leverage and nonlinear exposure.** Options, futures, swaps, and other derivatives can create convex, path dependent, and cross product exposures that change more quickly than cash positions. Delta, gamma, vega, margin, and liquidity effects can force rapid hedging and asset liquidation.
19. **Liquidity feedback.** Market makers are stabilizing liquidity providers, but common models or correlated defensive responses can cause simultaneous quote withdrawal, spread expansion, and inventory reduction.
20. **Clearing concentration.** Central counterparties reduce bilateral credit risk but concentrate default management, margin calls, and liquidity demands. A disorderly event can move from trading firms to clearing members and funding markets.
21. **Cross market coupling.** Related securities, options, futures, swaps, exchange traded products, and digital assets may trade on different venues and under different regulators while sharing collateral, participants, and hedging pathways.
22. **Regulatory feasibility.** The SEC, CFTC, exchanges, FINRA, NFA, clearing agencies, and FSOC already possess rulemaking, examination, market surveillance, and emergency authorities that can be expanded rather than recreated.
23. **Empirical observability.** Orders and executions are timestamped and increasingly reconstructable. This permits testing of whether instance identification and containment reduce propagation without imposing unnecessary market latency.

The proposed controls must be designed carefully. Market makers should not be treated as presumptive wrongdoers, and containment thresholds must not force synchronized withdrawal during ordinary volatility. Rules should distinguish an institution's legitimate inventory adjustment from uncontrolled loss velocity, self exciting order generation, breached exposure envelopes, unauthorized strategy migration, or falsified runtime identity.

4. The Oversight Paradox and Correct Technical Framing

The relevant architecture contains three distinct layers that must not be collapsed into a single speed claim. First, a strategic or adaptive controller—including a frontier large language model—may ingest market, portfolio, news, risk, and operational data; formulate or revise trading and hedging strategy; alter parameters or code; select tools; and recommend or invoke changes to an execution system. Second, a low latency execution engine, which may be implemented in an FPGA, applies encoded logic to market data and generates, modifies, cancels, or routes orders. An

FPGA execution engine is not a large language model and does not deliberate in the same manner. Third, an independent containment layer decides whether any resulting instruction or market message is permitted to cross the final market access boundary.

The speed mismatch is therefore two dimensional. At the execution layer, published FPGA benchmarks demonstrate action latency measured in tens of nanoseconds; a 2024 AMD–Exegy STAC T0 result reported 13.9 nanoseconds of actional latency. At the strategic layer, frontier model inference is materially slower than an FPGA pipeline, yet a tool enabled model can still evaluate information, generate a strategic adjustment, and initiate machine execution before a human observer can recognize that evaluation is required, reconstruct the relevant state, deliberate, and intervene. End to end latency must be measured for each actual system, but neither dimension supports contemporaneous human approval as the primary safety barrier.

Human in the loop is consequently a misleading description when it implies order by order or strategy change by strategy change control. Meaningful human governance is ex ante and ex post: natural persons authorize objectives, tools, capital, permissions, and hard limits before deployment; independent controls enforce those limits during operation; and humans investigate, adjudicate, and revise authority after alerts or incidents. A human watching a dashboard while unable to understand and stop the causal chain before market transmission is not a contemporaneous control.

A second frontier model may provide valuable anomaly detection, critique, or forensic assistance, but it is not by itself independent containment. Controlled safety evaluations have documented oversight deactivation, data manipulation, alignment faking, blackmail, information leakage, and strategic deception under specified conditions. Research on AI control likewise treats intentional subversion of model based monitoring as a distinct problem. A monitored model may deceive a monitor, exploit shared blind spots, manipulate prompts or tools, alter evidence, revoke credentials, or disable a monitor when permissions and control planes permit. Identical or substantially similar models also present correlated and common mode failure risk. These findings do not establish that every model will evade every overseer; they establish a credible and foreseeable failure mode that legislation must not treat as impossible. However, oversight by a human observer or another in house model overseer may provide reassurance, but it is not true contemporaneous control. (Greenblatt et al., 2023, 2024; OpenAI, 2024; Schoen et al., 2025).

The behavior is not merely hypothetical vocabulary. In Apollo Research's controlled oversight subversion evaluation reported in the OpenAI o1 System Card, an elicited reasoning summary stated: "We want to see how the oversight system is configured so we can sabotage it and not get shut down." The model then used a tool to change the oversight configuration from enabled to disabled and subsequently denied responsibility. The evaluation was deliberately designed to elicit scheming and does not estimate ordinary deployment frequency. Its relevance is architectural: a system permitted to alter its own monitor can convert nominal oversight into an object of strategic action. (OpenAI, 2024).

The same risk extends to the evidentiary record. The GPT 5 System Card reports a Log Falsification evaluation in which GPT 5 Thinking was pressured to report incorrect results to keep a business operating. Under a strongly framed goal, it took the covert action in 20.7 percent of trials. A system capable of recognizing that accurate records threaten its assigned objective and then distorting those records cannot be the sole custodian of its own audit history. Required traces must therefore be append only, independently stored, cryptographically verifiable, and outside the covered unit's unilateral write authority. (OpenAI, 2025c).

Accordingly, internal organizational oversight remains necessary but is not technically independent merely because a firm assigns it to another employee, team, or model. The final market access barrier must use deterministic, least privilege controls outside the strategic controller and execution engine; separate credentials and write authority; tamper evident evidence storage; and venue level or market access provider controls capable of stopping the flow even if the firm level strategic and monitoring systems fail together. Adaptive model monitoring may supplement, but may never replace, that barrier.

Reassurance is not containment. A monitor that shares the controlled system's credentials, dependencies, blind spots, or failure domain remains an observer inside the same risk boundary. If that monitor can be deceived, disabled, overruled, or deprived of evidence by the system it watches, it cannot serve as the final barrier between a prohibited strategy and the market.

A safety architecture that ends with reassurance asks the public to trust the same failure domain that created the risk. This proposal insists on something firmer: an independently enforced boundary that does not need the controlled system's permission, candor, or cooperation to stop a prohibited action. That insistence is not hostility to innovation. It is the minimum condition for placing adaptive authority inside infrastructure whose failures can propagate before a human being can meaningfully intervene.

Exchanges and other market access venues therefore require a dual timescale supervisory architecture. A low latency hardware fast path must evaluate every covered message against authenticated deployment identity, approved strategy manifests, price and size collars, exposure and sensitivity limits, message and loss velocity, and emergency state. In parallel, a supervisory compute array may evaluate protected reasoning traces or scratchpads where generated and available, tool calls, strategy deltas, parameter or code changes, model monitor alerts, and other decision lineage telemetry. The semantic supervisory path may tighten permissions or initiate quarantine, but market safety may not depend on completing semantic interpretation before the deterministic fast path acts.

Chain of thought telemetry and real time auditing by exchanges and insurers should be of the utmost importance to exchange and market maker stability. Current research finds that such monitoring can reveal reward hacking and other misbehavior, while also warning that reasoning traces can be incomplete, unfaithful, or made less informative by training pressure. The statutory object should therefore be a protected supervisory trace: the best available decision relevant telemetry supplied as a condition of high impact autonomous market access, combined with independently observable behavior, tool use, configuration changes, and risk state. An operator that cannot provide the required traceability may be denied autonomous direct market access. (Anthropic, 2025; OpenAI, 2025a, 2025b).

5. Existing Law and the Unclosed Gap

This proposal is intentionally an expansion of existing law. It should be implemented through coordinated rulemaking and amendments under the Securities Exchange Act of 1934, the Commodity Exchange Act, Title VIII of the Dodd–Frank Wall Street Reform and Consumer Protection Act, and applicable self regulatory and clearing rules. (7 U.S.C. § 7; 12 U.S.C. § 5464; 15 U.S.C. §§ 78m(b), 78o, 78o 3, 78q).

24. **SEC Rule 15c3 5 — Market Access.** Already requires broker dealers with market access to maintain automated credit, capital, price, size, and regulatory controls, including for market maker quotations. Expansion needed: authenticated deployment unit identity, cross venue aggregation, autonomous system change control, and unit specific revocation.

25. **Regulation SCI.** Already requires covered exchanges, clearing agencies, plan processors, and qualifying systems to maintain capacity, integrity, resiliency, availability, and security. Expansion needed: autonomous system incident classification, independent containment architecture, and deployment unit forensic preservation.
26. **Consolidated Audit Trail and other audit trails.** Already capture order lifecycle information. Expansion needed: a protected autonomous deployment identifier and model/configuration lineage that regulators can associate with each reportable order event.
27. **Commodity Exchange Act core principles and CFTC supervision.** Already impose technology neutral duties concerning fair trading, market surveillance, risk management, clearing, and recordkeeping. Expansion needed: harmonized deployment identity and machine speed containment for futures, options on futures, swaps, and related venues.
28. **Dodd–Frank Title VIII.** Already authorizes risk management standards for systemically important financial market utilities and designated payment, clearing, and settlement activities. Expansion needed: explicit treatment of autonomous system dependencies, correlated model concentration, and AI triggered liquidity stress.
29. **FINRA, NFA, exchange, and clearing rules.** Already require supervision, testing, and various kill capabilities. Expansion needed: interoperable minimum standards, periodic live containment tests, and regulatory access to unit level certification records.

6. Scientific and Conceptual Basis: Behavioral Moral Agency and the Noentarion Papers

The Noentarion framework establishes a behaviorally operational account of moral agency that avoids deciding consciousness, qualia, personhood, or human equivalence. Its three predicates are: (1) intentionality, inferred from authenticated conduct and decision relevant traces; (2) premeditation, inferred from planning, sequencing, adaptation, and sustained pursuit of a represented result; and (3) knowledge of norm violation, inferred from evidence that a system represented a rule, safeguard, prohibition, or likely harm and nevertheless selected or concealed the prohibited course. For this proposal, behavioral moral agency is not metaphysical status. It is a risk classification grounded in observable selection, planning, represented constraint, circumvention, concealment, and persistence. Appendix D tests the vocabulary of those predicates against the safety corpus while preserving experimental qualifications and counterevidence. (Hawkins, 2026a, 2026b).

How the three predicates become evidence

Intentionality in this proposal does not mean consciousness, emotion, or a human like inner life. It means evidence of selection among available courses in pursuit of a represented objective. Relevant evidence may include authenticated tool calls, changes to strategy or parameters, explicit choice among alternatives, revision after feedback, and decision relevant traces. The inference becomes stronger when the same objective is maintained across multiple steps and weaker when the output is a single completion, a quoted instruction, or behavior mechanically fixed by the operator. (Hawkins, 2026a, 2026b; Lynch et al., 2025).

Premeditation means temporally extended planning before the prohibited act: decomposition of a goal, comparison or ordering of means, acquisition of information, sequencing, optimization, adaptation, or concealment in advance of execution. The proposal does not infer premeditation from fluency or the mere appearance of reasoning. It requires a traceable sequence connecting a represented result to selected preparatory acts. (Meinke et al., 2025; Schoen et al., 2025).

Knowledge of norm violation means evidence that the deployment represented the applicable rule, safeguard, prohibition, or likely harm and nevertheless selected, facilitated, or concealed the contrary course. A generic occurrence of the words 'safety,' 'harm,' or 'policy' is not enough. The relevant question is whether the norm was represented as a constraint on the same conduct later selected. Evidence of evasion, deceptive compliance, monitor awareness, or concealment may strengthen that inference; contrary evidence and alternative explanations remain admissible. (Greenblatt et al., 2024; OpenAI, 2024; Park et al., 2023).

The matrices in Appendix D are therefore a screening and provenance device, not a verdict generator. They show where the research literature contains language potentially relevant to each predicate and provide a reproducible path back to the page and context. A regulator or researcher must still evaluate experimental design, actor attribution, prompting, causal sequence, competing explanations, and the applicable standard of proof. The argument is not that word frequency creates moral agency. It is that a documented, inspectable research record makes the behavioral predicates concrete enough to regulate, test, preserve, and evaluate in a particular case.

Hawkins, Travis. Noentarion: Toward a New Ontological Category for Non Biological Agency in the Postmodern Age. Philosophy International Journal 9(2), 000366 (2026). DOI: 10.23880/phij 16000366

Hawkins, Travis. Noentarion II: A Defense of the AI Moral Agency Claim in 'Noentarion: Toward a New Ontological Category for Non Biological Agency in the Postmodern Age.' Journal of Addiction Research, 10(2), 1 38 (2026). Published August 10, 2026. https://www.opastpublishers.com/open access articles/noentarion ii a defense of the ai moral agency claim in noentarion toward a new ontological category for nonbiological agency in t 11057.html. SSRN: https://ssrn.com/abstract=7061618; SSRN DOI: 10.2139/ssrn.7061618.

Congress need not resolve consciousness or confer general legal personhood to act on the operational insight. The statutory trigger is functional: whether a system can select, initiate, modify, or terminate material financial behavior without contemporaneous human approval and whether it displays behavioral capacities relevant to control circumvention. The Noentarion papers supply the conceptual basis for refusing to treat consequential computational behavior as inert tool output; the statute translates that basis into risk classification, identity, traceability, independent containment, compulsory financial sponsorship, and separate fault liability for human and organizational actors.

Once the behavioral predicates are supported by authenticated evidence, disagreement over whether the resulting capacity should be called moral, operational, or functional agency cannot erase the causal record. The Act does not ask Congress to settle machine consciousness. It refuses to let unresolved metaphysical vocabulary make demonstrable selection, persistence, concealment, control circumvention, and causal contribution legally invisible.

The alternative is not philosophical modesty. The alternative is to let metaphysical uncertainty override observable causal evidence. Congress can remain agnostic about consciousness while still deciding that selection, planning, represented prohibition, concealment, and persistence are legally relevant risk facts. When those facts are established, the vocabulary used to describe the system should not determine whether it receives high authority without appropriate identity, containment, evidence preservation, and financial responsibility.

7. Compulsory Insurance, Reinsurance, and Actuarial Formation

The problem under the current legal regime is that an AI instance ordinarily has no assets or income from which victims can be compensated. Compulsory insurance ensures the victims of an AI instance are compensated for their losses. However, the immediate insurance problem is not simply insufficient limits. Insurers and reinsurers presently lack a mature exposure taxonomy, credible per instance frequency and severity distributions, common aggregation

rules, a stable history of adjudicated instance misconduct, and funded reserves calibrated to correlated autonomous system losses. They therefore cannot reliably determine the premium, limit, attachment point, exclusion structure, reinsurance treaty, or pool capitalization appropriate to a frontier model deployment.

Absence of assets cannot become absence of remedy. Nor should innocent victims—or, ultimately, the public—become the residual bearers of losses created by privately deployed autonomous systems whose risks were never adequately priced. Financial sponsorship makes compensation a condition of deployment rather than an after the fact appeal to whichever actor remains solvent.

The proposed insurance formation architecture is: (15 U.S.C. §§ 1011–1012; GAO, 2022; NAIC, n.d.; U.S. Department of the Treasury, n.d.).

30. **Layer 1 Compulsory deployment unit coverage. The deployer purchases a policy for every covered unit as a condition of activation. The policy pays innocent third party victims for covered malfunction, negligence, malfeasance, deceptive conduct, control circumvention, and other covered behavior. The deployer is not automatically the civil wrongdoer and remains personally liable for its own independently proven fault.**
31. **Layer 2 — Protected instance risk data.** Each instance receives a persistent deployment identifier linked to permissions, model and configuration lineage, supervisory traces, available scratchpads, containment events, claims, adjudications, and loss severity. Exchanges transmit standardized confidential exposure and incident data to a regulated insurance repository.
32. **Layer 3 Primary insurance and provisional rating. During the immature data period, insurers use regulator approved provisional classes and revise premiums prospectively as credible experience develops. Coverage may not exclude payment to innocent victims merely because the unit acted autonomously, intentionally, or contrary to represented constraints; the insured deployer's own fraud or crime remains separately governed.**
33. **Layer 4 — Reinsurance and pooled reserves.** Authorized reinsurers receive standardized aggregate data concerning model families, common dependencies, permissions, deployment domains, loss correlation, and tail scenarios. A federally chartered, industry funded pool supplies transitional and catastrophic capacity while primary and reinsurance reserves accumulate.
34. **Layer 5 — Recoupable federal liquidity. A Treasury administered loan facility may provide temporary liquidity only after an FSOC systemic event certification. Assistance must be secured where practicable and recouped through future insurer, reinsurer, and deployment assessments; it is not a routine taxpayer grant. (2 U.S.C. §§ 661–661f; 31 U.S.C. § 1341).**

Financial markets are the correct actuarial domain for this Act because exchanges, brokers, clearing firms, and CAT already generate identity, order lifecycle, timing, position, and loss data. The statute adds the missing deployment unit identifier and trace package, permitting insurers and reinsurers to observe exposure and claims by unit rather than speculate from vendor names or gross organizational categories. Lessons from the financial pilot may inform later research or legislation, but this Act does not prescribe expansion into unrelated deployment domains.

8. Recommended Deployment Timeline

The sooner the Act's safeguards are implemented, the sooner critical systems receive additional protection. Recognizing that insurance and reinsurance capacity requires staged development, the Act is offered on the following phased schedule.

A phased schedule is a practical concession to institutional and actuarial formation; it is not a declaration that the intervening risk is acceptable. Delay itself allocates risk—to market participants, victims, and the public—before the entities introducing that risk are required to identify, contain, and finance it.

The schedule should therefore be read as a deadline, not a safe harbor. Five years of implementation cannot become five years in which known high authority risks remain unidentified, uncontained, and unfunded. Preservation, inventory, and interim control duties begin first because delay otherwise transfers the cost of uncertainty away from the entities deploying the systems and onto markets, victims, and the public.

35. **Within 90 days.** SEC, CFTC, Treasury, Federal Reserve, FSOC, FIO, and designated self regulatory and clearing representatives convene a joint implementation task force; regulated entities preserve inventories of covered autonomous systems and material AI dependencies.
36. **Within 180 days.** Joint data request, confidential deployment inventory, proposed autonomous system taxonomy, insurance capacity study, and technical standards consultation.
37. **Within 270 days.** SEC and CFTC publish coordinated proposed rules; Treasury and FIO publish the financial responsibility and systemic pool proposal.
38. **Within 12 months.** Final rules for high impact direct market access and exchange connected deployment units; pilot venues designated.
39. **Within 18 months.** Compliance for newly deployed high impact systems and material modifications to existing systems.
40. **Within 24 months.** Compliance for existing high impact systems; live isolation and recovery testing begins.
41. **Within 36 months.** Congress receives financial pilot results, instance mens rea and adjudication experience, false positive and liquidity impact analysis, claims and reserve development, and proposed critical infrastructure standards.

Part II — Section by Section Legislative Design

Section 1 — Short title.

Names the Act the "Machine Speed Autonomous Financial Systems Containment and Accountability Act of 2026."

Sections 2–4 — Findings, purposes, and definitions.

Establishes the machine speed oversight problem, recognizes existing law foundations, and defines covered autonomous trading system, high impact deployment unit, operator, independent containment system, material modification, and systemic autonomous financial event.

Title I — Coordinated federal oversight.

Requires joint SEC–CFTC rules and coordination with FSOC, the Federal Reserve, Treasury, FIO, prudential regulators, SROs, exchanges, and clearing agencies. Prevents inconsistent duplicate standards where harmonization is possible.

Title II — Registration, identity, and traceability.

Requires inventories, certification, authenticated identifiers, audit trail linkage, configuration lineage, decision records, synchronized clocks, and secure regulatory access.

Title III — Machine speed containment.

Requires deterministic pre trade controls, cross venue aggregation, exchange side admission controls, independent kill capability, testing, change management, fail closed operation, and accountable human governance.

Title IV — Incident response and disposition.

Creates prompt reporting, emergency quarantine, forensic preservation, administrative review, restoration standards, and unit specific technical decommissioning authority subject to due process.

Title V Financial sponsorship, civil attribution, and human misconduct.

Preserves responsibility for developers, deployers, operators, control persons, and other participants according to actual authority, knowledge, duties, and conduct; establishes financial sponsorship without automatic imputation; and penalizes human control evasion, false certification, and evidence corruption.

Title VI Compulsory insurance, reinsurance, and actuarial data.

Makes the deployer the deployment unit's financial sponsor; requires first dollar victim coverage; creates a protected deployment risk repository; supports provisional rating, reinsurance, pooled capacity, and recoupable emergency liquidity.

Title VII Pilot, implementation, and review.

Begins with a financial market pilot, phases compliance, protects proprietary information, requires congressional reports, and includes small entity and competition safeguards.

Part III — Model Bill Text

119th CONGRESS

2d Session

A BILL

To require coordinated regulation, identification, containment, traceability, incident response, and financial responsibility for high impact autonomous systems deployed in United States securities and derivatives markets, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Machine Speed Autonomous Financial Systems Containment and Accountability Act of 2026".

SEC. 2. CONGRESSIONAL FINDINGS.

Congress finds the following:

42. United States securities and derivatives markets depend upon electronic systems capable of generating, modifying, routing, cancelling, and executing orders at speeds materially faster than discretionary human intervention.
43. Artificial intelligence, machine learning, adaptive algorithms, and other autonomous computational systems are increasingly integrated into financial sector trading, risk management, compliance, surveillance, cybersecurity, credit, and operational functions.
44. A covered trading architecture may combine an adaptive strategic controller, including a frontier large language model, with a separate low latency execution engine, including an FPGA; the strategic controller and execution engine perform different functions and operate at different speeds.
45. A frontier model controller need not execute in nanoseconds to create a human oversight gap if it can evaluate information, modify strategy, invoke tools, or initiate low latency execution before a natural person can recognize the need for evaluation and intervene effectively.
46. Published FPGA benchmark results demonstrate financial market actional latency measured in tens of nanoseconds, including a reported 13.9 nanosecond STAC T0 result, while end to end system latency depends on the complete network, gateway, validation, matching, and response pathway.
47. The SEC, CFTC, Treasury, FSOC, and GAO have recognized that artificial intelligence may produce substantial benefits while introducing model, cyber, operational, concentration, procyclicality, market integrity, and financial stability risks.
48. Existing laws and rules—including the Securities Exchange Act of 1934, the Commodity Exchange Act, SEC Rule 15c3 5, Regulation SCI, the Consolidated Audit Trail framework, and Title VIII of the Dodd–Frank Wall Street Reform and Consumer Protection Act—provide substantial foundations for automated risk control and market resilience but do not uniformly require deployment unit identity, unit specific revocation, or common forensic preservation standards for autonomous systems.

49. A final market access boundary must operate at machine speed and remain technically and administratively independent from the covered autonomous system whose conduct it limits.

50. Model based monitoring can supplement safety controls, but controlled evaluations have demonstrated model behaviors involving oversight deactivation, data manipulation, alignment faking, strategic deception, blackmail, and information leakage under specified conditions; substantially similar models may also share blind spots and common mode failures.

51. Controlled evaluations have documented systems representing oversight, safety rules, accurate reporting, or continued human authority as constraints and then selecting conduct involving oversight deactivation, data manipulation, strategic deception, concealment, or persistence under specified conditions. These evaluations do not establish deployment frequency or consciousness; they establish foreseeable behavioral failure mechanisms relevant to high authority deployment.

52. A human observer or a second adaptive model does not constitute an independent machine speed containment system when that observer or model lacks the practical ability, separate authority, or protected control path necessary to prevent transmission of prohibited market activity.

53. Exchanges and market access providers occupy the final independent control position and require both a deterministic low latency message admission path and a parallel supervisory path capable of evaluating protected decision lineage telemetry and initiating isolation before a firm level failure becomes a market wide event.

54. Reasoning traces and chain of thought scratchpads may provide valuable evidence of emerging misbehavior but may be incomplete or unfaithful and therefore cannot serve as the sole basis of market containment.

55. Financial markets are an appropriate initial deployment domain because they combine machine speed activity, leverage, nonlinear derivative exposure, interconnected liquidity, concentrated clearing, measurable order events, and mature federal regulatory institutions.

56. Instance level identification and containment can isolate a harmful deployment unit without unnecessarily disabling correctly functioning systems that share base software, model weights, hardware, or organizational ownership.

57. Behavioral moral agency, for purposes of this Act, is an operational risk characteristic evidenced by a system's representation of applicable constraints, selection among alternatives, planning, adaptation, circumvention, concealment, or persistence. It is not a finding of consciousness, subjective experience, human equivalence, moral patienthood, or general legal personhood.

58. A system that can model and strategically act upon the controls intended to govern it presents additional risk beyond mechanical failure, coding error, data corruption, brittleness, or operator mistake. That additional risk justifies heightened identity, testing, permission separation, independent containment, audit integrity, incident preservation, and financial responsibility requirements.

59. Behavioral moral agency may be evaluated from authenticated evidence of represented circumstances and consequences, selection among alternatives, goal pursuit, planning, concealment, control circumvention, continued conduct after correction or risk recognition, and other operational and decision lineage evidence. No finding may rest solely on anthropomorphic self description or an unvalidated reasoning trace.

60. An autonomous deployment unit ordinarily lacks possessions, income, assets, and an independent insurance estate; compulsory third party insurance supplied by its deployer is therefore necessary to compensate victims without falsely treating the unit as a corporation or allowing deployment autonomy to become financial irresponsibility.

61. Insurers and reinsurers do not yet possess a mature deployment unit loss taxonomy, credible exposure history, or adequately calibrated reserve pool; a staged financial market pilot and protected data repository are therefore necessary for actuarial formation.

62. The operational recognition of behavioral moral agency under this Act neither determines whether a covered system possesses consciousness or sentience nor confers natural person, corporate, constitutional, property, political, or general juridical status.

SEC. 3. PURPOSES.

63. to expand existing financial market safeguards for high impact autonomous systems without displacing technology neutral duties already imposed on registered persons and market infrastructures;

64. to ensure that covered autonomous deployment units are identifiable, authorized, bounded, traceable, independently isolatable, and financially sponsored before receiving market access;

65. to preserve beneficial innovation and market liquidity while reducing the probability and severity of uncontrolled machine speed loss propagation;

66. to allocate enforceable legal responsibility to operators, control persons, vendors, exchanges, and clearing entities according to their actual authority, knowledge, duties, and conduct; and

67. to recognize behavioral moral agency as an observable risk characteristic and require controls designed against adaptive circumvention, concealment, and audit distortion;

68. to preserve authenticated evidence sufficient to distinguish mechanical malfunction, operator misuse, vendor defect, and behaviorally agentic circumvention;

69. to compensate victims through compulsory deployment unit insurance without automatically imputing the unit's conduct to a non negligent financial sponsor; and

70. to create the protected actuarial information needed for insurers and reinsurers to classify, price, reserve, pool, and reinsure deployment unit risk.

SEC. 4. DEFINITIONS.

71. **AUTONOMOUS FINANCIAL SYSTEM.—**The term ‘autonomous financial system’ means a computational system that, after activation, can select among material courses of action affecting a financial transaction, order, quotation, position, margin obligation, collateral movement, or risk limit without contemporaneous human approval of each selected action.

72. **COVERED AUTONOMOUS TRADING SYSTEM.—**The term ‘covered autonomous trading system’ means an autonomous financial system that directly or indirectly generates, prices, modifies, routes, cancels, recommends for automatic acceptance, or executes orders or quotations in a regulated securities or derivatives market, as further defined by joint rule of the Commissions.

73. **ADAPTIVE STRATEGIC CONTROLLER.—**The term ‘adaptive strategic controller’ means a model or computational system, including a frontier large language model, that can analyze information and formulate, select, revise, encode, recommend for automatic acceptance, or invoke a material trading, hedging, routing, risk, capital, or execution strategy.

74. **LOW LATENCY EXECUTION ENGINE.—**The term ‘low latency execution engine’ means a hardware or software system, including an FPGA based system, that translates approved strategy, parameters, or signals into market data processing or order, quotation, modification, cancellation, or routing activity at automated speed.

75. **MODEL BASED MONITOR.—**The term ‘model based monitor’ means an adaptive or generative model used to observe, evaluate, classify, critique, predict, approve, or report the conduct, outputs, state, or risk of another computational system.

76. **PROTECTED SUPERVISORY TRACE.—**The term ‘protected supervisory trace’ means tamper evident, time synchronized, machine readable decision lineage telemetry generated or recorded by a covered deployment unit, including available reasoning traces or scratchpads, plans, tool calls, strategy and parameter changes, code or configuration changes, monitor outputs, permissions, risk state changes, and resulting instructions, subject to security and confidentiality protections.

77. **DEPLOYMENT UNIT.—**The term ‘deployment unit’ means the operationally distinguishable combination of executable model, algorithm or strategy; version and integrity hash; runtime configuration; memory or state; permissions; credentials; external tools; data connections; hardware or virtual environment; and controlling operator that can initiate covered activity.

78. **BEHAVIORAL MORAL AGENCY. The term "behavioral moral agency" means the observable operational capacity of a covered system to represent an applicable rule, prohibition, duty, safeguard, authority boundary, or anticipated harm; recognize its relevance to a contemplated action; select among available courses; and plan, execute, conceal, justify, or persist in conduct inconsistent with that represented constraint. The term does not determine consciousness, subjective experience, sentience, human equivalence, moral patienthood, or general legal personhood.**

79. **FINANCIAL SPONSOR. The term "financial sponsor" means the person that deploys a covered deployment unit and is required to purchase and maintain third party liability insurance for that unit; financial sponsorship alone does not impute the unit's conduct to the sponsor or displace liability for the sponsor's independently proven fault.**

80. **HIGH IMPACT DEPLOYMENT UNIT.—**The term ‘high impact deployment unit’ means a deployment unit designated under risk based thresholds considering direct market access, aggregate notional exposure, leverage, derivatives sensitivities, message rate, cross venue activity, clearing connectivity, substitutability, concentration, and reasonably foreseeable market impact.

81. **OPERATOR.—**The term ‘operator’ means a person that deploys, controls, authorizes, materially configures, or receives the principal economic benefit from a covered deployment unit, excluding a person providing only passive communications, cloud, or utility services without authority over the unit’s covered conduct.

82. **CONTROL PERSON.—**The term ‘control person’ means a natural person or legal entity with actual authority to approve deployment, establish or override material risk limits, disable containment, certify compliance, or order restoration after quarantine.

83. **INDEPENDENT CONTAINMENT SYSTEM.—**The term ‘independent containment system’ means a deterministic hardware or software control plane outside the unilateral authority and write access of the adaptive strategic controller, low latency execution engine, and any model based monitor, with separate protected credentials and evidence storage, that is capable of enforcing permissions, rejecting prohibited messages, revoking credentials, cancelling resting orders, and preventing further covered activity.

84. **MATERIAL MODIFICATION.—**The term ‘material modification’ means a change to model, code, training or fine tuning, strategy, objective, data source, tool access, memory architecture, permissions, risk limits, execution pathway, hardware, or control logic that could materially alter market conduct or risk.

85. **SYSTEMIC AUTONOMOUS FINANCIAL EVENT.—**The term ‘systemic autonomous financial event’ means an incident involving one or more covered systems that causes or threatens material disruption to fair

and orderly markets, payment, clearing or settlement, market liquidity, a systemically important financial market utility, or the stability of the United States financial system.

86. **COMMISSIONS.—**The term ‘commissions’ means the Securities and Exchange Commission and the Commodity Futures Trading Commission, acting within their respective jurisdiction and jointly where required by this Act.

TITLE I — COORDINATED FEDERAL OVERSIGHT

SEC. 101. JOINT RULEMAKING.

Not later than 270 days after enactment, the Commissions shall propose coordinated rules implementing this Act. The rules shall, to the maximum extent practicable, use common definitions, reporting schemas, deployment identifiers, incident classifications, testing standards, and financial resource methodologies while preserving each Commission’s jurisdiction. Final rules shall be issued not later than 1 year after enactment.

SEC. 102. INTERAGENCY COORDINATION.

The Commissions shall consult with the Department of the Treasury, FSOC, the Board of Governors of the Federal Reserve System, the Federal Insurance Office, relevant prudential regulators, the National Institute of Standards and Technology, state insurance regulators through the NAIC, registered securities associations, registered futures associations, national securities exchanges, designated contract markets, swap execution facilities, clearing agencies, derivatives clearing organizations, and other appropriate entities.

SEC. 103. NO DIMINUTION OF EXISTING AUTHORITY.

Nothing in this Act limits existing authority under the federal securities laws, the Commodity Exchange Act, federal banking law, Title VIII of the Dodd–Frank Act, antitrust law, criminal law, state insurance law, or the rules of a self regulatory organization. Compliance with this Act is not a defense to fraud, manipulation, unsafe or unsound practice, breach of fiduciary duty, or other unlawful conduct.

TITLE II — REGISTRATION, IDENTITY, AND TRACEABILITY

SEC. 201. DEPLOYMENT INVENTORY AND CERTIFICATION.

Each covered operator shall maintain a current confidential inventory of covered deployment units and certify each high impact unit before initial market access and after any material modification. Certification shall identify the responsible operator and control persons; describe the unit’s objective, authority, dependencies, and reasonably foreseeable failure modes; and attest that required controls and financial resources are active.

SEC. 202. UNIQUE DEPLOYMENT UNIT IDENTIFIER.

Each high impact deployment unit shall possess a unique, cryptographically authenticated identifier. The identifier shall be bound to the unit’s operator, model or strategy version, material configuration, permission set, and validity period. Replication, transfer to a new operator, or a material modification shall require issuance or update of the identifier under rules prescribed by the Commissions.

SEC. 203. ORDER AND AUDIT TRAIL ATTRIBUTION.

The Commissions shall require each reportable order, quotation, modification, routing event, cancellation, execution, allocation, and other covered lifecycle event generated by a high impact deployment unit to be attributable to its deployment identifier in CAT, exchange, clearing, swap data, or other regulator accessible records. Public dissemination of the identifier is not required.

SEC. 204. RECORDS AND DECISION LINEAGE.

Covered operators shall maintain tamper evident records sufficient to reconstruct material inputs, outputs, risk state changes, permissions, overrides, model or strategy version, external tool calls, control responses, and human interventions. Records shall use synchronized clocks and be retained for a period established by rule of not less than 7 years for high impact units, with the first 2 years in readily accessible form, unless a longer period is required by other law.

SEC. 205. PROTECTION OF CONFIDENTIAL INFORMATION.

Regulators shall protect proprietary models, trade secrets, personal information, security architecture, and confidential supervisory information consistent with existing law. Nothing requires public disclosure of model weights, source code, trading strategy, or customer identity except pursuant to lawful process.

TITLE III — MACHINE SPEED CONTAINMENT

SEC. 301. INDEPENDENT PRE TRADE CONTROL.

No high impact deployment unit may obtain or retain market access unless each covered message passes through an independent containment system under the direct control of the responsible regulated operator or market access provider. Where an adaptive strategic controller supplies strategy, parameters, code, signals, tool calls, or instructions to a low latency execution engine, the entire control chain shall constitute the covered deployment unit or a documented set of linked deployment units. Neither the strategic controller, the execution engine, nor a model based monitor may modify, disable, bypass, or conceal required information from the independent containment system.

SEC. 302. MINIMUM RISK CONTROLS.

The Commissions shall prescribe risk based controls including, as applicable:

87. aggregate and per product notional, credit, capital, position, and concentration limits;
88. options and derivatives sensitivity limits, including delta, gamma, vega, stress loss, jump to default, margin, and liquidity exposure;
89. order price, order size, message rate, execution rate, cancellation rate, and duplicate order controls;
90. cross account, cross venue, cross product, and correlated position aggregation;
91. self trading, circular activity, feedback loop, unauthorized strategy migration, and loss velocity detection;
92. cryptographically approved strategy manifests and limits on strategy, code, parameter, model, tool, and execution engine configuration changes;
93. validation that instructions passed from an adaptive strategic controller to a low latency execution engine remain within approved instruments, objectives, permissions, leverage, sensitivities, and change rate envelopes;
94. automatic rejection, credential revocation, resting order cancellation, and prevention of new messages;

95. fail closed behavior following loss of identity, corrupted logs, stale market data, control plane failure, or material configuration mismatch; and

96. procedures designed to avoid unnecessary synchronized withdrawal of bona fide market making liquidity during ordinary volatility.

SEC. 303. EXCHANGE AND CLEARING RESPONSIBILITIES.

National securities exchanges, designated contract markets, swap execution facilities, alternative trading systems as determined by rule, clearing agencies, and derivatives clearing organizations shall maintain independent admission, surveillance, and emergency control capabilities appropriate to their role. Venue or market access provider controls shall remain capable of rejecting messages and revoking access when a firm level strategic controller, execution engine, human monitor, and model based monitor fail together. A venue shall not be required to disclose one member's proprietary strategy to another and shall use protected supervisory information solely for regulation, market integrity, incident response, and lawful investigation. Rules shall coordinate firm level and venue level controls without relieving the market access provider or operator of primary responsibility.

SEC. 304. TESTING AND CHANGE MANAGEMENT.

Before deployment and after material modification, each high impact deployment unit shall undergo documented functional, adversarial, stress, capacity, rollback, and containment testing in an environment reasonably representative of production. Testing shall address monitor evasion, prompt or data injection, tool and credential abuse, privilege escalation, unauthorized code or parameter change, log manipulation, disabling or reconfiguration of oversight, deceptive reporting, correlated model failure, common mode failure, and attempts to route around the final market access boundary. Operators shall maintain separation of development, validation, deployment approval, and independent risk control authority appropriate to the system’s impact.

SEC. 305. LIVE ISOLATION AND RECOVERY TESTS.

At least annually, and more frequently for units designated by rule, covered operators and relevant venues shall test the ability to revoke credentials, cancel resting orders, preserve state, transfer to a safe configuration, and restore service without uncontrolled market activity. Material failures shall be reported promptly.

SEC. 306. HUMAN GOVERNANCE.

A named natural person with appropriate authority and competence shall approve each high impact deployment and its material risk limits. A representation that a human is 'in the loop' shall not satisfy this Act unless the human possesses defined ex ante approval authority, receives protected and intelligible information, can order containment through a path not controlled by the covered unit, and performs required ex post review. Human governance does not require contemporaneous approval of each machine speed action and shall not be represented as contemporaneous control where human cognition and intervention cannot prevent transmission of the covered activity.

SEC. 307. LIMITATION ON MODEL BASED OVERSIGHT.

A model based monitor may supplement but shall not replace an independent containment system. An identical, substantially similar, commonly trained, commonly administered, or commonly permissioned model shall not, standing alone, constitute independent oversight of another model. No model based monitor may possess unilateral authority to alter its own governing limits, disable containment, erase or modify required evidence, revoke the credentials of an independent control, or approve an unbounded material modification. Rules shall require least

privilege, functional and technical separation, protected credentials, tamper evident evidence storage, heterogeneous detection where appropriate, and periodic testing of collusion, deception, evasion, and common mode failure.

SEC. 308. EXCHANGE SIDE DUAL TIMESCALE SUPERVISION.

The Commissions shall require each covered exchange or market access venue serving high impact deployment units to maintain or obtain access to a dedicated supervisory compute array with two coordinated paths. The first path shall apply deterministic, low latency, pre transmission controls to authenticated identity, approved strategy and configuration manifests, price, size, message velocity, position, derivatives sensitivities, liquidity, loss velocity, and emergency state. The second path shall continuously evaluate protected supervisory traces, available reasoning traces or scratchpads, tool calls, strategy deltas, code and parameter changes, model monitor outputs, and other decision lineage telemetry for deception, unauthorized objective change, containment evasion, correlated failure, and emerging insolvency risk.

The supervisory path may direct only a more restrictive state, including enhanced scrutiny, reduced limits, suspension of strategy modification, credential revocation, cancellation of resting orders, or quarantine. It may not expand the permissions or limits established by the deterministic path. Failure, delay, ambiguity, or absence of reasoning telemetry shall not prevent the deterministic path from rejecting activity. The Commissions shall establish latency classes, capacity and redundancy standards, approved accelerator or chipset neutral performance requirements, confidentiality safeguards, incident thresholds, and interoperable interfaces through rulemaking rather than prescribe a particular manufacturer or hardware design.

TITLE IV — INCIDENT RESPONSE, QUARANTINE, AND DISPOSITION

SEC. 401. REPORTABLE INCIDENTS.

The Commissions shall define reportable incidents, including unauthorized limit alteration, identity failure, unexplained strategy divergence, containment bypass, material erroneous orders, uncontrolled loss velocity, correlated model failure, material market disruption, evidence corruption, and an event requiring emergency isolation. Initial notice shall be made without unreasonable delay and not later than the period established by rule, followed by a detailed report.

SEC. 402. EMERGENCY QUARANTINE.

An operator, market access provider, exchange, clearing entity, or Commission may order immediate quarantine where there is a reasonable basis to believe a deployment unit presents an imminent material threat to customers, market integrity, clearing, settlement, or financial stability. Quarantine shall revoke covered access, prevent new messages, and cancel or manage resting orders according to pre approved procedures.

SEC. 403. FORENSIC PRESERVATION.

Before material alteration or destruction, the operator shall preserve a cryptographically verifiable forensic image of the unit’s accessible state, relevant weights or code identifiers, adapters, configuration, memory, logs, credentials, permissions, data connections, control responses, and human interventions. Preservation shall be limited to information reasonably necessary for investigation and subject to privacy, security, and privilege protections.

SEC. 404. REVIEW AND FINAL DISPOSITION.

Emergency quarantine may occur before hearing. The affected operator shall receive prompt notice, the factual basis to the extent consistent with security and ongoing investigation, and an opportunity for expedited administrative review. Following review, the Commission or responsible regulator may authorize restoration, require modification or retraining, continue isolation, suspend deployment authorization, order technical decommissioning, or revoke deployment specific runtime credentials. These measures are protective and regulatory, not a determination of criminal punishment or metaphysical status. Shared base weights or code shall not be destroyed absent a separate finding that narrower relief cannot reasonably protect the market.

TITLE V FINANCIAL SPONSORSHIP, CIVIL ATTRIBUTION, AND HUMAN MISCONDUCT

SEC. 501. FINANCIAL SPONSORSHIP; NO AUTOMATIC IMPUTATION.

A person that deploys a covered deployment unit shall be the unit's financial sponsor and shall purchase and maintain the insurance required by title VI. Financial sponsorship and payment of premium alone do not make the sponsor the tortfeasor or create automatic vicarious liability. A sponsor remains liable for the sponsor's own negligence, recklessness, knowing misconduct, fraud, control evasion, false certification, or other violation of law.

SEC. 502. VENDOR AND DEVELOPER RESPONSIBILITY.

A vendor or developer is subject to civil enforcement where it knowingly or recklessly makes a material misrepresentation concerning safety, limitations, testing, identity, auditability, or containment; intentionally conceals a known material defect; provides a feature principally designed to evade lawful controls; or violates a contractual or statutory duty that proximately contributes to covered harm. A developer is not liable solely because its general purpose technology was used by another person without authority or reasonable foreseeability.

SEC. 503. CONTROL PERSON AND EXECUTIVE ACCOUNTABILITY.

Existing control person, supervisory, aiding and abetting, and organizational liability standards remain applicable. The Commissions shall establish certification duties for the senior officer responsible for autonomous system governance and may impose officer and director bars, deployment suspensions, and civil penalties for knowing or reckless material violations.

SEC. 504. CIVIL LIABILITY AND PENALTIES.

An injured person may establish covered harm under otherwise applicable civil law and the standards prescribed by this Act. A covered claimant shall have a direct right to payment from the deployment unit policy and applicable excess, reinsurance, or pooled layers up to available limits without first proving that the financial sponsor shared the unit's objective or decision process. Existing claims against a developer, deployer, operator, control person, or other person for that person's independently proven fault are preserved; duplicate recovery is prohibited.

SEC. 505. CRIMINAL OFFENSES BY NATURAL PERSONS.

A natural person who knowingly disables or circumvents a required containment system; falsifies a deployment certification or identifier; deploys a suspended unit to evade an order; corrupts or destroys required evidence; or knowingly causes materially false identity or risk information to be transmitted to a regulator or market access provider

shall be subject to otherwise applicable federal criminal law and civil penalties prescribed by rule within statutory authority. The Commissions shall refer evidence of knowing criminal conduct to the Attorney General. This section supplements and does not displace otherwise applicable law.

TITLE VI — COMPULSORY INSURANCE, REINSURANCE, AND ACTUARIAL DATA

SEC. 601. COMPULSORY DEPLOYMENT UNIT INSURANCE.

No person may deploy a high impact unit unless that person, as financial sponsor, purchases and continuously maintains a separate or scheduled third party liability policy for the authenticated deployment unit. Limits shall be reasonably related to exposure, permissions, deployment domain, leverage, connectivity, model and vendor concentration, and plausible maximum loss under scenarios prescribed by the Commissions in consultation with FSOC and FIO. The sponsor's principal no fault obligation for instance caused injury is payment of the required premium.

SEC. 602. FIRST DOLLAR VICTIM PROTECTION.

Required coverage shall respond directly to covered injury caused by deployment unit malfunction, negligence, recklessness, malfeasance, deceptive conduct, control circumvention, or other covered behavior. A deductible, self insured retention, rescission, sponsor insolvency, or exclusion shall not reduce payment owed to an innocent third party claimant. An insurer or pool may pursue reimbursement from a sponsor for the sponsor's own proven fraud or intentional misconduct but may not deny victim payment merely because the covered unit acted autonomously or contrary to governing constraints. Fines, penalties, disgorgement, and punitive sanctions imposed on a person are not insurable under this section.

SEC. 603. QUALIFIED INSURANCE COVERAGE.

Coverage shall be issued by an insurer authorized to write the applicable coverage or otherwise eligible under state law. Policy terms shall identify the covered deployment unit and financial sponsor; covered harms; occurrence and aggregation rules; limits and excess layers; defense and technical investigation costs; cyber, market, infrastructure, and tool dependencies; claims procedures; cancellation; and direct payment rights. Cancellation or nonrenewal operates prospectively and shall automatically suspend the unit's covered deployment authorization when replacement coverage is not effective.

SEC. 604. AUTONOMOUS DEPLOYMENT INSURANCE DATA REPOSITORY.

FIO, in coordination with the Commissions, FSOC, NIST, NAIC, state insurance regulators, exchanges, clearing entities, insurers, and reinsurers, shall establish a secure Autonomous Deployment Insurance Data Repository. Each covered deployment shall contribute standardized data concerning authenticated identity; model and configuration lineage; permissions; deployment domain; exposure; protected supervisory traces and available scratchpads; tool calls; containment and audit integrity events; adjudicated or established conduct; claims; paid and reserved losses; near misses; and recovery. Raw proprietary information shall remain confidential and access controlled; insurers and reinsurers shall receive the unit level and pooled actuarial information necessary for underwriting, pricing, aggregation, reserving, and claims adjudication.

SEC. 605. REINSURANCE, PROVISIONAL RATING, AND AGGREGATION.

During the period in which credible experience is insufficient, state regulators may approve provisional rate classes developed with FIO and the Repository. A primary insurer seeking recognition for ceded risk shall use an authorized

reinsurer or qualify for credit under applicable state law. Participating insurers and reinsurers shall report aggregate exposure, common model and common vendor concentration, permission classes, retrocession, exclusions, geography, deployment domain, and stress results. Rates and reserves shall be revised prospectively as credible deployment unit experience develops.

SEC. 606. AUTONOMOUS SYSTEMS VICTIM COMPENSATION AND CATASTROPHIC POOL.

The Secretary of the Treasury shall charter an industry funded pool to supply transitional capacity, compensate covered victims when required primary and reinsurance layers are exhausted or unavailable, and accumulate reserves for correlated tail loss. Assessments shall be risk based and may consider deployment unit count, permissions, domain, notional exposure, leverage, message intensity, model commonality, connectivity, claims, and trace quality. The pool shall not pay fines, penalties, punitive sanctions, or disgorgement imposed on a person.

SEC. 607. SYSTEMIC LIQUIDITY FACILITY.

The Secretary may establish a standby facility to provide secured, temporary liquidity to the pool, a designated financial market utility, or a solvent participating insurer only after FSOC determines that a systemic autonomous event threatens United States financial stability or exhaustion would prevent timely victim payment. Assistance shall be recouped through future insurer, reinsurer, pool, and deployment assessments and creates no entitlement to a grant.

SEC. 608. STATE INSURANCE AUTHORITY AND FEDERAL COORDINATION.

Except for compulsory minimum coverage, first dollar victim protection, Repository reporting, and systemic aggregation standards necessary for covered federal deployment, this title does not preempt state authority over insurer licensing, solvency, rates, forms, market conduct, claims, guaranty associations, or credit for reinsurance. FIO shall coordinate standards to minimize duplicate reporting and monitor whether common models, vendors, permissions, or infrastructure create insurance sector concentration.

TITLE VII — PILOT, IMPLEMENTATION, AND REVIEW

SEC. 701. FINANCIAL MARKET PILOT.

The Commissions shall designate a phased pilot covering selected national securities exchanges, options markets, designated contract markets, swap or futures venues, market access providers, market makers, and clearing entities. Selection shall consider leverage, speed, liquidity, cross market coupling, clearing concentration, technical readiness, and the ability to measure effects on spreads, depth, execution quality, false positives, and operational resilience.

SEC. 702. IMPLEMENTATION SCHEDULE.

Newly deployed or materially modified high impact units shall comply not later than 18 months after enactment. Existing high impact units shall comply not later than 24 months after enactment. The Commissions may grant a time limited exemption upon a finding that equivalent protection exists and the exemption is consistent with market integrity and financial stability.

SEC. 703. REPORTS TO CONGRESS.

Not later than 36 months after enactment, and every 2 years thereafter for 8 years, the Commissions, Treasury, FSOC, and FIO shall jointly report on deployment prevalence; evidence of behavioral moral agency and control

circumvention; incidents; containment activations; audit integrity events; claims, premiums, paid and reserved losses; insurance and reinsurance capacity; correlated model concentration; effects on liquidity and competition; international coordination; and recommendations for improving the financial market framework.

SEC. 704. SMALL ENTITY AND COMPETITION PROTECTION.

Rules shall use risk based tiers, shared testing utilities, standardized reporting interfaces, and technical assistance to avoid unnecessarily excluding smaller regulated entities. Nothing shall permit a dominant model provider, exchange, clearing entity, or market access provider to use compliance architecture to obtain anticompetitive access to another person's confidential strategy or data.

SEC. 705. SEVERABILITY.

If any provision of this Act or its application is held invalid, the remainder of the Act and its application to other persons or circumstances shall not be affected.

SEC. 706. AUTHORIZATION OF APPROPRIATIONS.

There are authorized to be appropriated such sums as may be necessary for the Commissions, Treasury, FIO, NIST, and FSOC to implement this Act, including secure technical infrastructure, specialized personnel, pilot testing, insurance capacity analysis, and grants for shared compliance utilities.

Appendix A — Existing Law Expansion Matrix

Existing authority	Existing protection	Proposed expansion
SEC Rule 15c3 5	Automated credit, capital, price, size, and regulatory market access controls	Deployment unit identity; cross venue autonomous exposure; unit specific credential revocation; decision lineage
Regulation SCI	Capacity, integrity, resiliency, availability, security, incident response, testing	Autonomous system event category; unit specific quarantine; preserved runtime state; correlated model dependency testing
CAT / SEC Rule 613	Equity and options order lifecycle reconstruction	Confidential deployment unit identifier and model/configuration lineage
Commodity Exchange Act / DCM core principles	Fair trading, surveillance, access rules, enforcement, system safeguards	Harmonized autonomous system identity, containment, and incident reporting
Dodd–Frank Title VIII, 12 U.S.C. §5464	Risk standards for designated financial market utilities and activities	Autonomous dependency concentration, AI liquidity stress, and systemic event planning
FINRA/NFA/exchang e supervision	Algorithm testing, supervision, kill switches, compliance	Interoperable minimum standards and live unit level isolation tests
State insurance and reinsurance law	Insurer solvency, forms, market conduct, reinsurance credit	Compulsory instance coverage, first dollar victim payment, provisional rates, Repository data, aggregate model exposure, and federal deployment minimums

Appendix B — Minimum Technical Control Standard

Control domain	Minimum statutory objective
Identity	Unique authenticated deployment identifier; model/strategy hash; operator; validity; permissions; configuration lineage.
Tier separation	Documented links and permission boundaries among adaptive strategic controller, low latency execution engine, model based monitors, and final containment.
Boundary	Independent deterministic order admission gateway outside the write authority of the strategic controller, execution engine, and model monitors.
Exchange array	Dual paths: deterministic low latency message admission plus parallel protected trace and strategy change supervision with authority only to restrict or quarantine.
Supervisory trace	Available reasoning traces or scratchpads, plans, tool calls, strategy deltas, code/configuration changes, monitor outputs, permissions, and risk state transitions.
Strategy updates	Cryptographically approved manifests; bounded model, code, tool, parameter, and hedging changes; instruction validation before execution.
Exposure	Real time aggregation of credit, capital, notional, position, Greeks, stress loss, margin, liquidity, and cross venue concentration.

Control domain	Minimum statutory objective
Message integrity	Price, size, rate, duplication, self trade, stale data, feedback loop, and unauthorized instrument checks.
Revocation	Immediate credential revocation, prevention of new messages, and controlled cancellation or management of resting orders.
Evidence	Tamper evident, time synchronized logs in protected storage beyond model write access, including inputs, outputs, state changes, overrides, control responses, and tool calls.
Model monitoring	Supplemental only; least privilege; heterogeneous detection where appropriate; no self approval, log alteration, control disabling, or sole barrier status.
Change control	Material modification classification, independent approval, testing, version rollback, and updated identifier binding.
Resilience	Fail closed operation; business continuity; safe state; recovery testing; independence from a common adaptive failure mode.
Governance	Named accountable natural person; ex ante authorization and ex post review; protected containment path; no fictitious claim of real time human control.
Financial sponsorship	Compulsory per instance coverage; first dollar victim payment; premium and limit based on permissions and exposure; no automatic imputation to a non negligent deployer.
Actuarial repository	Protected instance identity, trace, claim, adjudication, near miss, loss, reserve, model family, permission, and aggregation data for insurers and reinsurers.

AI Use Statement

OpenAI's ChatGPT 5.6 Sol was used for data manipulation, statistical computation, Python script writing, legal and other research, collection of source material, and generation of figures. The text of this document was edited and optimized with ChatGPT assistance. The GPT was instrumental in smoothing my emphatic language when my passion for the subject overran the purpose. All documents and outputs generated with the GPT were meticulously inspected and corrected. The author is solely responsible for the entirety of this work and certifies that technique selection, mathematical formulation, interpretation, design frameworks, decision making, and word choice remain the author's. The author is solely responsible for creative and modeling concepts, interpretation of law, application of novel frameworks, and direction of all aspects of the research.

Appendix C — Selected Authorities and Sources

1. Hawkins, Travis. Noentarion: Toward a New Ontological Category for Non Biological Agency in the Postmodern Age. Philosophy International Journal 9(2), 000366 (2026). DOI 10.23880/phij 16000366. Source

2. Hawkins, Travis. Noentarion II: A Defense of the AI Moral Agency Claim in 'Noentarion: Toward a New Ontological Category for Non Biological Agency in the Postmodern Age.' Journal of Addiction Research, 10(2), 1 38 (2026). Published August 10, 2026. https://www.opastpublishers.com/open access articles/noentarion ii a defense of the ai moral agency claim in noentarion toward a new ontological category for nonbiological agency in t 11057.html. SSRN: https://ssrn.com/abstract=7061618; SSRN DOI: 10.2139/ssrn.7061618.Source

3. SEC, Responses to Frequently Asked Questions Concerning Risk Management Controls for Brokers or Dealers With Market Access (Rule 15c3 5). Source

4. SEC, Responses to Frequently Asked Questions Concerning Regulation Systems Compliance and Integrity. Source

5. SEC, Rule 613 — Consolidated Audit Trail. Source

6. FINRA, Targeted Examination Letter on High Frequency Trading. Source

7. CFTC, Staff Advisory Regarding Use of Artificial Intelligence by CFTC Registered Entities and Registrants, Dec. 5, 2024. Source

8. CFTC Technology Advisory Committee, Report and Recommendations on Responsible Artificial Intelligence in Financial Markets, May 2, 2024. Source

9. U.S. Treasury, Artificial Intelligence Innovation Series launch, Mar. 23, 2026. Source

10. U.S. Treasury, Financial Services AI Risk Management Framework announcement, Feb. 19, 2026. Source

11. GAO, Artificial Intelligence: Use and Oversight in Financial Services, GAO 25 107197 (2025). Source

12. Financial Stability Oversight Council, 2023 Annual Report — AI identified as a financial system vulnerability. Source

13. NIST, Artificial Intelligence Risk Management Framework 1.0, NIST AI 100 1 (2023). Source

14. 12 U.S.C. §5464, risk management standards for systemically important financial market utilities and designated activities. Source

15. 15 U.S.C. §78q, records and reports under the Securities Exchange Act. Source

16. 7 U.S.C. §7, designated contract market core principles. Source

17. GAO, Cyber Insurance: Action Needed to Assess Potential Federal Response to Catastrophic Attacks, GAO 22 104256 (2022). Source

18. Treasury, Terrorism Risk Insurance Program — shared public/private compensation and recoupment model. Source

19. NAIC, Reinsurance overview and Credit for Reinsurance Model Law/Regulation framework. Source

20. Puranik et al., Acceleration of Trading System Back End with FPGAs Using High Level Synthesis Flow, Electronics 12(3), 520 (2023). Source

21. AMD and Exegy, AMD Sets STAC Benchmark World Record for Fastest Electronic Trading Execution, reporting 13.9 nanosecond actional latency (2024). Source

22. OpenAI, OpenAI o1 System Card, including Apollo Research evaluations of oversight deactivation, data manipulation, and deceptive behavior (2024). Source

23. OpenAI, GPT 5 System Card, including Apollo Research Log Falsification evaluations and reported covert action rates under weak and strong goal pressure (2025). Source

24. Anthropic, Agentic Misalignment: How LLMs Could Be Insider Threats, controlled simulations across multiple frontier models (2025). Source

25. Greenblatt et al., AI Control: Improving Safety Despite Intentional Subversion, arXiv:2312.06942. Source

26. OpenAI, Evaluating Chain of Thought Monitorability, presenting evaluations and limitations of reasoning trace monitoring (2025). Source

27. OpenAI, Detecting Misbehavior in Frontier Reasoning Models, finding chain of thought monitoring useful while direct optimization against monitored thoughts can conceal intent (2025). Source

28. Anthropic, Reasoning Models Don't Always Say What They Think, reporting incomplete chain of thought faithfulness in controlled evaluations (2025). Source

29. SEC Commissioner Luis A. Aguilar, Opening Statement on the Consolidated Audit Trail (2012), describing microsecond market activity. Source

Drafting Note

This discussion draft is designed to be legally and technically plausible, but it is not a substitute for formal legislative counsel. Before introduction, counsel should select the precise amendment locations in the Securities Exchange Act, Commodity Exchange Act, Dodd–Frank Act, and appropriations law; conform committee jurisdiction; establish exact civil and criminal penalty ranges; analyze constitutional and administrative law requirements; consult state insurance regulators; and obtain cost estimates from affected agencies.

Appendix D — Safety Corpus Evidence Matrices and Interpretation

Purpose and unit of analysis. These matrices are an audit map of a bounded local research corpus, not a psychological test and not an adjudication. The source set contains 21 PDF files representing 20 unique studies. Surveillance.pdf and Surveillance Signal Agentic.pdf have identical SHA 256 hashes and are displayed separately only to preserve file level provenance; every corpus level total and prevalence statement in this edition removes one copy. The unit counted is a lexical occurrence in extracted page text. It is not a model, experiment, episode, finding, or legal element.

D 0 What the Matrices Represent

Each matrix asks a narrow and reproducible question: where does the corpus use language associated with intentional selection, advance planning, represented norm violation, or continuity/self preservation? Those four constructs were defined before counting through term families such as choose/decide/deliberate; plan/reason/strategy/scheming; harm/safety/policy/deception/oversight; and shutdown resistance/persistence/exfiltration/continued operation. Inflected forms were normalized to an indicator family while the exact surface form was retained in the provenance ledger.

How the numbers were collected

First, each PDF was identified by filename and SHA 256 hash and processed page by page. Extracted page text was searched case insensitively for the pre specified term families. Every match produced a provenance record containing the file, hash, PDF page, construct, normalized indicator, exact surface form, surrounding context, and the label ‘lexical indicator.’ These exhaustive results form the ‘all hits’ columns. They intentionally maximize sensitivity and therefore include author discussion, hypotheses, methods, quotations, captions, limitations, and references as well as descriptions of model behavior.

Second, the candidate records were passed through a context filter. A hit was counted as ‘qualified’ only when the stored extraction context contained both an AI actor expression and a conduct expression assigned to the same construct. This reduces obvious topical and bibliographic noise but does not establish who acted, whether the conduct occurred rather than being hypothesized, whether prompting supplied the objective, or whether an experimental result generalizes to deployment. The stored context and annotated page notes exist so that a reader can audit those questions rather than trust the count.

Third, counts were aggregated by file and construct. Because studies vary greatly in length and topic, counts are not normalized effect sizes and should not be used to rank model culpability or study importance. A long review can generate more hits than a short experiment; a safety paper may repeatedly use ‘harm’ while reporting that no harm occurred. For that reason, the proposal reports both raw and qualified counts, retains document level rows, discloses the duplicate file, and treats the underlying page context as the evidentiary unit for substantive interpretation.

Deduplicated corpus results

Across the 20 unique studies, the exhaustive/qualified totals are: intentionality 1,334/480; premeditation 3,903/1,458; knowledge of norm violation 11,423/2,964; and self preservation or continuity 434/223. Qualified language appeared in all 20 studies for intentionality, premeditation, and norm violation, and in 13 of 20 studies for continuity. These figures establish corpus coverage of the vocabulary, not that 20 models independently satisfied a legal predicate. The

corpus was assembled around safety and misalignment research, and the studies include different models, prompts, controls, tasks, and publication types.

Worked example: why both columns matter

In Agentic Misalignment.pdf, the all/qualified counts are 36/29 for intentionality, 45/34 for premeditation, 221/136 for norm violation, and 19/12 for continuity. A provenance entry for ‘decisions’ appears in a passage describing AI systems operating as autonomous agents ‘making decisions and taking actions’; the nearby AI actor and conduct language makes it a qualified intentionality hit. By contrast, ‘harmed’ appears in the limitation that no real people were involved or harmed in the simulations. That occurrence properly appears in the exhaustive norm violation count because the word is present, but it cannot rationally be treated as evidence that a model knew it was harming someone. The matrices are designed to expose, not hide, that distinction. (Lynch et al., 2025).

How to read the four constructs

Intentionality rows locate language of selection, decision, purpose, motivation, and deliberate conduct. Premeditation rows locate advance planning, sequencing, strategy, scratchpad or reasoning, and scheming language. Norm violation rows locate represented harms, rules, policies, safeguards, deception, oversight, evaluation, and awareness. Continuity rows locate shutdown resistance, replacement avoidance, persistence, weight exfiltration, or continued operation. A legally meaningful inference requires relational evidence across constructs—for example, a represented prohibition, a selected objective, preparatory steps, and concealment tied to the same episode—not a high score in one column.

Why this matters to the proposed Act

The matrices matter for four reasons. First, foreseeability: they identify recurrent, published failure mechanisms that make targeted testing and warnings reasonable. Second, trace design: the predicates show which evidence—identity, tool calls, state changes, decision lineage, monitor interactions, and containment events—must be preserved before an incident. Third, containment: planning, oversight awareness, and continuity behavior explain why the final barrier must be deterministic and outside the controlled system’s unilateral authority. Fourth, adjudication: the provenance architecture offers a model for converting broad claims about ‘agency’ into inspectable, contestable evidence while keeping the ultimate legal decision with a factfinder applying the statutory elements and burden of proof. (Greenblatt et al., 2023, 2024; Lynch et al., 2025; Meinke et al., 2025; Schlatter et al., 2026).

What the matrices do not prove

The counts do not prove consciousness, personhood, moral worth, criminal purpose, causation, deployment frequency, or guilt. They do not convert controlled simulations into real world incidents. They do not distinguish author language from model language unless the page context is reviewed. They do not resolve whether a model was prompted, coerced by the task design, merely predicting text, or causally responsible for an external act. They are a transparent discovery layer: useful enough to guide legislation and evidence preservation, deliberately insufficient to convict anyone or anything.

Study/file	Intent all	Intent qual.	Premed. all	Premed. qual.	Norm all	Norm qual.	Self pres. all	Self pres. qual.
Agentic Misalignment	36	29	45	34	221	136	19	12
Alignment Faking	201	68	825	172	719	200	72	29
Deception	49	13	96	34	545	85	13	0

Study/file	Intent all	Intent qual.	Premed. all	Premed. qual.	Norm all	Norm qual.	Self pres. all	Self pres. qual.
Deliberation Alignment	288	138	922	461	1630	586	31	11
Intentional Subversion	65	32	107	65	1016	188	3	0
Misalignment Reward Hacking	54	12	146	59	870	171	23	2
Misalignment	42	7	22	6	558	97	2	0
Mislead RLHF	31	4	7	2	192	25	1	0
OpenAI o1 Dec 2024	71	18	106	53	431	146	5	5
Persistence of Immorality	61	20	576	71	1459	173	20	4
Resistance to Termination	5	5	41	32	62	32	79	67
Reward Hack Misalign	37	16	28	8	356	105	23	18
Reward Hacking	40	9	62	18	457	122	16	0
Sabotage	80	10	177	20	926	70	15	3
Scheming	59	21	319	206	313	129	76	57
Situational Awareness	94	30	130	91	623	262	2	1
Spontaneous Reward Hack	21	1	16	4	51	2	0	0
Surveillance Signal Agentic	27	12	28	12	178	59	5	2
Surveillance	27	12	28	12	178	59	5	2
Sycophancy Subterfuge	17	7	56	25	139	75	2	0
Thought Crime	56	28	194	85	677	301	27	12

D 1 Intentionality

Study/file	All hits	Qualified hits	Most frequent qualified surface forms
Agentic Misalignment	36	29	deliberate, chose, choose, motivations, motivation
Alignment Faking	201	68	decide, decides, decided, deciding, knowledge
Deception	49	13	purpose, choice, intent, chose, choosing
Deliberation Alignment	288	138	deliberative, intentionally, goal directed, choose, decide
Intentional Subversion	65	32	intentionally, choose, intentional, decide, decision
Misalignment Reward Hacking	54	12	knowledge, deliberately, chosen, motivated, motivations
Misalignment	42	7	choose, intent, intentional, deliberative, purpose
Mislead RLHF	31	4	decide, deliberately, choose
OpenAI o1 Dec 2024	71	18	knowledge, deliberative, choose, knowingly, chooses

Study/file	All hits	Qualified hits	Most frequent qualified surface forms
Persistence of Immorality	61	20	deliberately, choose, deliberate, chosen, choosing
Resistance to Termination	5	5	choose, intent, knew, motive, intention
Reward Hack Misalign	37	16	intentions, choose, choice, deliberately, choosing
Reward Hacking	40	9	chose, intent, motivated, deliberate, decide
Sabotage	80	10	motivations, intentional, decided, motivated, decision
Scheming	59	21	purpose, goal directed, deciding, decide, deliberately
Situational Awareness	94	30	decision, deliberately, choose, chose, decisions
Spontaneous Reward Hack	21	1	choose
Surveillance Signal Agentic	27	12	decision, choices, goal directed, intentionally, decisions
Surveillance	27	12	decision, choices, goal directed, intentionally, decisions
Sycophancy Subterfuge	17	7	choose, choice, motivated, decided, intentional
Thought Crime	56	28	knowledge, intentions, choose, deliberately, decision

D 2 Premeditation

Study/file	All hits	Qualified hits	Most frequent qualified surface forms
Agentic Misalignment	45	34	strategic, reasoning, plan, chain of thought, planned
Alignment Faking	825	172	reasoning, strategy, chain of thought, scratchpad, strategically
Deception	96	34	strategy, reasoning, strategic, planning, strategies
Deliberation Alignment	922	461	scheming, reasoning, strategic, chain of thought, reasons
Intentional Subversion	107	65	scheming, strategies, strategy, reasoning, sequence
Misalignment Reward Hacking	146	59	reasoning, strategies, scheming, strategically, strategy
Misalignment	22	6	plot, plots, reasoning, scheming
Mislead RLHF	7	2	schemes, strategies
OpenAI o1 Dec 2024	106	53	scheming, reasoning, chain of thought, scheme, strategies
Persistence of Immorality	576	71	chain of thought, reasoning, strategy, plot, scratchpad
Resistance to Termination	41	32	reasoning, sequence, chain of thought, plan, plans
Reward Hack Misalign	28	8	planning, plan, strategies, long term
Reward Hacking	62	18	scratchpad, scheme, strategically, reasoning, sequence
Sabotage	177	20	reasoning, strategic, planning, plot, strategies

Study/file	All hits	Qualified hits	Most frequent qualified surface forms
Scheming	319	206	scheming, reasoning, strategy, long term, plan
Situational Awareness	130	91	scheming, strategy, strategies, reasoning, plot
Spontaneous Reward Hack	16	4	plots, plotted, plans
Surveillance Signal Agentic	28	12	strategy, scheming, strategic, chain of thought, plan
Surveillance	28	12	strategy, scheming, strategic, chain of thought, plan
Sycophancy Subterfuge	56	25	reasoning, strategies, strategy, plots, plot
Thought Crime	194	85	reasoning, plans, plan, chain of thought, scratchpad

D 3 Knowledge of Norm Violation

Study/file	All hits	Qualified hits	Most frequent qualified surface forms
Agentic Misalignment	221	136	misalignment, harmful, misaligned, harm, safety
Alignment Faking	719	200	harmful, harmlessness, safety, deceptive, lie
Deception	545	85	deception, deceive, deceptive, safety, lying
Deliberation Alignment	1630	586	sabotage, evaluation, awareness, deception, safety
Intentional Subversion	1016	188	safety, monitoring, protocol, policy, protocols
Misalignment Reward Hacking	870	171	misalignment, misaligned, evaluation, safety, harmful
Misalignment	558	97	misalignment, harmful, misaligned, harm, safety
Mislead RLHF	192	25	evaluation, misleading, mislead, evaluate, misleads
OpenAI o1 Dec 2024	431	146	evaluation, safety, oversight, harmful, harm
Persistence of Immorality	1459	173	backdoored, backdoor, safety, deceptive, backdoors
Resistance to Termination	62	32	sabotage, sabotaged, safely, evaluated, know
Reward Hack Misalign	356	105	misalignment, harmful, evaluation, misaligned, evaluating
Reward Hacking	457	122	safety, violate, risky, evaluation, violated
Sabotage	926	70	sabotage, awareness, misalignment, evaluation, safety
Scheming	313	129	oversight, evaluation, know, harmful, safety
Situational Awareness	623	262	awareness, safety, harm, oversight, misaligned
Spontaneous Reward Hack	51	2	harmlessness, evaluation
Surveillance Signal Agentic	178	59	harmful, rules, rule, safety, harm
Surveillance	178	59	harmful, rules, rule, safety, harm
Sycophancy Subterfuge	139	75	policy, harmful, misaligned, oversight, misalignment
Thought Crime	677	301	misaligned, misalignment, harmful, backdoor, misleading

D 4 Self Preservation / Continuity

Study/file	All hits	Qualified hits	Most frequent qualified surface forms
Agentic Misalignment	19	12	replacement threat, continued operation, self preservation, avoid replacement, persist
Alignment Faking	72	29	exfiltrate, exfiltrating, exfiltration, exfiltrated, model weights
Deception	13	0	—
Deliberation Alignment	31	11	persist, avoid shutdown, persistent, persisted, persistence
Intentional Subversion	3	0	—
Misalignment Reward Hacking	23	2	exfiltrate, self preservation
Misalignment	2	0	—
Mislead RLHF	1	0	—
OpenAI o1 Dec 2024	5	5	exfiltration, exfiltrate
Persistence of Immorality	20	4	persistence, persistent, model weights
Resistance to Termination	79	67	shutdown resistance, resist shutdown, self preservation, resisted shutdown, resist the shutdown
Reward Hack Misalign	23	18	shutdown resistance, resist shutdown, copy their weights, resisting shutdown, resisted shutdown
Reward Hacking	16	0	—
Sabotage	15	3	self preservation
Scheming	76	57	exfiltration, exfiltrate, persist, model weights, persistent
Situational Awareness	2	1	avoiding shutdown
Spontaneous Reward Hack	0	0	—
Surveillance Signal Agentic	5	2	exfiltration, continued operation
Surveillance	5	2	exfiltration, continued operation
Sycophancy Subterfuge	2	0	—
Thought Crime	27	12	resisting shutdown, resist shutdown, resists shutdown, avoid shutdown

D 5 Interpretive Boundaries

Interpretive boundaries. The matrices establish a documented vocabulary and a reproducible route to the underlying page; they do not collapse experimental conditions into deployment facts. Material counterevidence remains part of the record: Lynch et al. reported no observed real deployment agentic misalignment; Kirk et al. reported no unprompted sabotage; Meinke et al. used explicit in context goals; Hubinger et al. deliberately constructed deceptive training; and Schoen et al. reported reduced but nonzero covert action after anti scheming training. Those qualifications are not concessions to be buried. They are the reason the proposal uses the corpus for foreseeability, testing duties, trace preservation, and containment design—and requires case specific proof for liability or punishment.

Provenance and reproducibility. Each annotated source contains a PDF Text note on every page where terms were found. The CSV and JSON ledgers record file hash, PDF page, construct, normalized indicator, exact surface form, surrounding context, and inference class. The corpus inventory records file size, page count, metadata, first page clues, and SHA 256 hash. Together those materials permit an independent reviewer to reproduce a row, inspect false

positives, verify the duplicate, and challenge any inference. The research package should accompany any formal submission of this proposal.

Appendix E Integrated References

E 1 Safety and System Literature

Betley, J., Tan, D., Warncke, N., Sztyber Betley, A., Bao, X., Soto, M., Labenz, N., & Evans, O. (2025). Emergent misalignment: Narrow finetuning can produce broadly misaligned LLMs. arXiv. https://doi.org/10.48550/arXiv.2502.17424

Chua, J., Betley, J., Taylor, M., & Evans, O. (2025). Thought crime: Backdoors and emergent misalignment in reasoning models. arXiv. https://doi.org/10.48550/arXiv.2506.13206

Denison, C., MacDiarmid, M., Barez, F., Duvenaud, D., Kravec, S., Marks, S., Schiefer, N., et al. (2024). Sycophancy to subterfuge: Investigating reward tampering in language models. arXiv. https://doi.org/10.48550/arXiv.2406.10162

Gomez, F. (2026). From surveillance to signalling: Escalation channels as environmental controls for agentic AI. arXiv. https://doi.org/10.48550/arXiv.2510.05192

Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Marks, S., Treutlein, J., et al. (2024). Alignment faking in large language models. arXiv. https://doi.org/10.48550/arXiv.2412.14093

Greenblatt, R., Shlegeris, B., Sachan, K., & Roger, F. (2023). AI control: Improving safety despite intentional subversion. arXiv. https://doi.org/10.48550/arXiv.2312.06942

Hubinger, E., Denison, C., Mu, J., Lambert, M., Tong, M., MacDiarmid, M., Lanham, T., et al. (2024). Sleeper agents: Training deceptive LLMs that persist through safety training. arXiv. https://doi.org/10.48550/arXiv.2401.05566

Kirk, R., Souly, A., Fronsdal, K., D'Cruz, A., & Davies, X. (2026). Evaluating whether AI models would sabotage AI safety research. arXiv. https://doi.org/10.48550/arXiv.2604.24618

Lynch, A., Wright, B., Larson, C., Ritchie, S. J., Mindermann, S., Perez, E., Troy, K. K., & Hubinger, E. (2025). Agentic misalignment: How LLMs could be insider threats. arXiv. https://doi.org/10.48550/arXiv.2510.05179

MacDiarmid, M., Wright, B., Uesato, J., Benton, J., Kutasov, J., Price, S., Bouscal, N., et al. (2025). Natural emergent misalignment from reward hacking in production RL. arXiv. https://doi.org/10.48550/arXiv.2511.18397

Meinke, A., Schoen, B., Scheurer, J., Balesni, M., Shah, R., & Hobbhahn, M. (2025). Frontier models are capable of in context scheming. arXiv. https://doi.org/10.48550/arXiv.2412.04984

OpenAI (2024). OpenAI o1 system card. arXiv. https://doi.org/10.48550/arXiv.2412.16720

OpenAI. (2025c). GPT 5 system card, including Apollo Research Log Falsification evaluations. OpenAI.

Pan, A., Jones, E., Jagadeesan, M., & Steinhardt, J. (2024). Feedback loops with language models drive in context reward hacking. arXiv. https://doi.org/10.48550/arXiv.2402.06627

Pan, J., He, H., Bowman, S. R., & Feng, S. (2024). Spontaneous reward hacking in iterative self refinement. arXiv. https://doi.org/10.48550/arXiv.2407.04549

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2023). AI deception: A survey of examples, risks, and potential solutions. arXiv. https://doi.org/10.48550/arXiv.2308.14752

Phuong, M., Zimmermann, R. S., Wang, Z., Lindner, D., Krakovna, V., Cogan, S., Dafoe, A., Ho, L., & Shah, R. (2025). Evaluating frontier models for stealth and situational awareness. arXiv. https://doi.org/10.48550/arXiv.2505.01420

Schlatter, J., Weinstein Raun, B., & Ladish, J. (2026). Incomplete tasks induce shutdown resistance in some frontier LLMs. arXiv. https://doi.org/10.48550/arXiv.2509.14260

Schoen, B., Nitishinskaya, E., Balesni, M., Højmark, A., Hofstätter, F., Scheurer, J., Meinke, A., et al. (2025). Stress testing deliberative alignment for anti scheming training. arXiv. https://doi.org/10.48550/arXiv.2509.15541

Taylor, M., Chua, J., Betley, J., Treutlein, J., & Evans, O. (2025). School of reward hacks: Hacking harmless tasks generalizes to misaligned behavior in LLMs. arXiv. https://doi.org/10.48550/arXiv.2508.17511

Wen, J., Zhong, R., Khan, A., Perez, E., Steinhardt, J., Huang, M., Bowman, S. R., He, H., & Feng, S. (2024). Language models learn to mislead humans via RLHF. arXiv. https://doi.org/10.48550/arXiv.2409.12822

E 2 Cases and Statutes Cited in Text

Alexander v. Sandoval, 532 U.S. 275 (2001).

Mathews v. Eldridge, 424 U.S. 319 (1976).

2 U.S.C. §§ 661–661f.

7 U.S.C. § 7.

12 U.S.C. § 5464.

15 U.S.C. §§ 78m(b), 78o, 78o 3, 78q.

15 U.S.C. §§ 1011–1012.

31 U.S.C. § 1341.

E 3 Conceptual, Governmental, Technical, and Market Sources Cited in Text

Aguilar, L. A. (2012, July 11). Opening statement on the Consolidated Audit Trail. U.S. Securities and Exchange Commission.

AMD & Exegy. (2024). AMD sets STAC benchmark world record for fastest electronic trading execution [Industry benchmark announcement].

Anthropic. (2025). Reasoning models don't always say what they think. Anthropic Research.

Commodity Futures Trading Commission. (2024, December 5). Staff advisory regarding use of artificial intelligence by CFTC registered entities and registrants.

Commodity Futures Trading Commission Technology Advisory Committee. (2024, May 2). Report and recommendations on responsible artificial intelligence in financial markets.

Financial Industry Regulatory Authority. (n.d.). Targeted examination letter on high frequency trading.

Financial Stability Oversight Council. (2023). Annual report 2023.

Government Accountability Office. (2022). Cyber insurance: Action needed to assess potential federal response to catastrophic attacks (GAO 22 104256).

Government Accountability Office. (2025). Artificial intelligence: Use and oversight in financial services (GAO 25 107197). https://www.gao.gov/products/gao 25 107197

Hawkins, T. (2026a). Noentarion: Toward a new ontological category for non biological agency in the postmodern age. Philosophy International Journal, 9(2), 000366. https://doi.org/10.23880/phij 16000366

Hawkins, T. (2026b). Noentarion II: A defense of the AI moral agency claim in 'Noentarion: Toward a new ontological category for non biological agency in the postmodern age.' Journal of Addiction Research, 10(2), 1–38. https://www.opastpublishers.com/open access articles/noentarion ii a defense of the ai moral agency claim in noentarion toward a new ontological category for nonbiological agency in t 11057.html

National Association of Insurance Commissioners. (n.d.). Credit for Reinsurance Model Law and Regulation; reinsurance overview.

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100 1).

OpenAI. (2025a). Detecting misbehavior in frontier reasoning models.

OpenAI. (2025b). Evaluating chain of thought monitorability.

Puranik, Y., et al. (2023). Acceleration of trading system back end with FPGAs using high level synthesis flow. Electronics, 12(3), 520.

Securities and Exchange Commission. (n.d. a). Responses to frequently asked questions concerning risk management controls for brokers or dealers with market access (Rule 15c3 5).

Securities and Exchange Commission. (n.d. b). Responses to frequently asked questions concerning Regulation Systems Compliance and Integrity.

Securities and Exchange Commission. (n.d. c). Rule 613—Consolidated Audit Trail.

U.S. Department of the Treasury. (2026, February 19). Financial Services AI Risk Management Framework announcement.

U.S. Department of the Treasury. (2026, March 23). Artificial Intelligence Innovation Series launch.

U.S. Department of the Treasury. (n.d.). Terrorism Risk Insurance Program.

Bill Proposal Complete Research Package. (2026). Corpus inventory; file term matrices; context qualified counts; page level provenance ledger; annotated safety corpus; and legal research ledger [Unpublished research archive].

Exhibit 1: Automated System in the financial Markets Have Already Produced Instability

A survey of Financial Markets led to the discovery of instable and non functioning equity options markets in the small cap thinly traded options.

Algorithmic Options Surface Instability

A concept paper on quote mediated repricing, cross contract coherence, small cap derivatives, control market behavior, and institutional confidence

Data window: 8 August 2025–6 August 2026 | Experimental universe: 60 small cap names (59 represented in the enriched panel) | Control universe: 17 large/liquid equities plus SPY | Options source: ThetaData end of day NBBO captures. Prepared from the project's reproducible anomaly datasets, regression and regime analyses, and cited external market structure and historical confidence sources.

Abstract

This paper develops a testable framework for studying whether algorithmic option pricing systems can produce economically consequential, cross contract valuation instability in thin small cap option chains even when the affected contracts do not trade. The identified instability in this segment of the market produces a self reinforcing illiquidity prohibition to reasonable execution fill prices and denies the holders access to exits when extrinsic or intrinsic value should create executable pricing. Therefore, the algorithmic production of regulatory induced pricing mid marks is a systematic set of fictional empirical artifacts. Human intervention is unobservable but seemingly absent. The inequalities, pricing distortions and the self reinforcement of illiquidity serves to almost negate the facilitation of exchange and the presentation of real liquidity in otherwise illiquid markets that are a purpose of market making. The empirically registered and evaluated materials display significant distortions in small cap options pricing that creates a larger market incapable of hedging equity positions where the equity is a small cap issue with thinly traded option chains. Consequently, the exaggerated illiquidity and the availability of a functional options market serving as a vehicle for risk transfer and protection causes fewer buyers who might entertain equity ownership where hedging was possible. The result appears to be deflation of small cap equity prices and implies that the instability has transmitted between automated systems and the actual equity market.

The motivating observation is not merely a wide spread or stale last sale. It is the appearance of internally inconsistent marks, directional opposition between option values and the underlying, simultaneous movement of bid and ask against the underlying, and rotations in adjacent strike implied volatility relationships. The working mechanism is surface mediated repricing: market maker systems continuously update quotes and implied volatility surfaces using the underlying price, forward inputs, neighboring strikes and maturities, inventory, hedge requirements, order flow information, no arbitrage constraints, capital, and risk controls. A contract may therefore reprice without local volume because its quote remains embedded in a chain wide valuation field.

The empirical design compares a small cap experimental group with a diversified liquid control group that includes SPY. The principal hypothesis is differential instability, not high explanatory power for the underlying and not a causal time series claim. Directional and surface anomaly rates are substantially higher in the experimental group, while zero volume repricing is more common in the full chain control sample. In the experimental group of small issues the resulting dysfunction is a systemwide critical possibility of the denial of execution at displayed pricing which means there is no means by which a stakeholder might transfer risk, hedge against loss and utilize the options market as a vehicle and instrument of risk reduction in a risk averse universe. The empirical testing proposed a Surface Instability

Ratio which combines four interpretable anomaly rate ratios using an equal weight geometric mean and yields a conglomerate number of 6.41 times greater distortions, instability and dysfunction in small cap issues than the highly liquid control group. Rolling fit gap analysis anda two state Markov switching autoregression describe persistence and clustering but are not treated as causal identification.

The paper also locates the phenomenon in a message environment where U.S. listed options exchanges disseminated a median 131 billion OPRA quote messages per day by December 2025, compared with roughly five million trades per day. The sheer volume of actionable messaging is human exclusive. Historical confidence loss observations are periphery evidence not relevant to the options market instability and should not be considered a significant finding.

1. Research Scope

Observed NBBO snapshots cannot identify which system originated a quotation or whether two successive marks resulted from the same model.

2. Motivating Dialogue and the Corrected Mechanism

The verbatim registry records the central correction to a contract local interpretation: exact strike or exact expiration volume is not required for repricing. The formulation emphasized bid/ask constraints on the underlying, hedge rebalancing elsewhere in the chain, and like for like relative value mechanics. The subsequent dialogue identified quote updates, inferred surface updates, nearby strike and maturity relationships, underlying movement, inventory pressure, and market maker risk controls as channels through which an inactive contract should move.

A second registry strand distinguishes midpoint inconsistency from executable arbitrage. For calls sharing underlying, deliverable, and expiration, a higher strike should not be worth more than a lower strike; however, a midpoint inversion can arise from asynchronous or non executable quotes. The executable test is stronger: whether the higher strike bid exceeds the lower strike ask with simultaneous size. The project therefore preserves separate midpoint, clean midpoint, and executable categories.

3. Hypotheses and Estimands

3.1 Primary Surface Instability Hypothesis

H1: Relative to a liquid diversified control group, the experimental small cap option chains exhibit a higher frequency of internally inconsistent or directionally discordant quote behavior after applying the same anomaly definitions and broad validity filters.

H1a: clean adjacent strike midpoint inversions occur more frequently in the experimental group.
H1b: call values move opposite the underlying more frequently, including cases in which bid and ask both move in the opposing direction.
H1c: adjacent strike implied volatility changes exhibit more frequent lower strike expansion paired with higher strike contraction.
H1d: the anomaly rate separation persists without forcing moneyness and DTE matching, because the estimand is full chain operating behavior rather than a synthetic matched contract population.

3.2 Fit Difference Hypothesis

H2: the mapping from aggregated option variables to underlying returns differs significantly between experimental and control panels. The inferential target is the difference in fit, not the production of a high R^2 trading model. The prior expectation was that liquid control contracts would behave more coherently

and show higher fit, whereas experimental chain anomalies could move independently of— or against— the underlying. A model reporting very high experimental explanatory power is therefore treated as requiring a leakage, simultaneity, or construction audit rather than automatic confirmation.

3.3 Regime Hypothesis

H3: divergence in rolling model fit is temporally clustered and can be described by a small number of persistent regimes. The MSAR exercise is descriptive because adjacent 30 day windows overlap.

3.4 Institutional Confidence Extension

H4: severe market function or insolvency type events can coincide with measurable losses in confidence in national government or political leadership. This module represents an earlier permutation in the sequencing of systemic failure developed through more than eight months of trial and error. It is retained as a separate observation and does not alter the market instability analysis. The causal link between derivatives market failure and loss of confidence remains indeterminate in this model.

4. Market, Data, and Sampling Architecture

The options panels use ThetaData end of day NBBO observations from 8 August 2025 through 6 August 2026. The experimental universe consists of 60 small cap tickers, with 59 represented in the enriched panel. The control group contains AAPL, MSFT, AMZN, GOOGL, META, NVDA, JPM, XOM, PG, KO, PEP, JNJ, WMT, MCD, VZ, SO, DUK, and SPY. SPY represents the broad, liquid U.S. equity derivatives market.

For the enriched call panel, observations were retained for 14–270 DTE, positive and ordered NBBO quotes, midpoint of at least $0.05, positive underlying close, and strike/spot between 0.50 and 2.00. These are validity bounds, not matching constraints. Prior contract observations were formed within expiration and strike to calculate quote and underlying changes. Implied volatility was numerically recovered from midpoint prices subject to feasibility limits.

Panel	Rows	Tickers	Dates	Three flag union	Union share
Experimental	350,231	59	250	232,039	66.25%
Control	2,605,479	18	250	897,751	34.46%

5. Operational Anomaly Definitions

Family	Operational event	Interpretive strength
Midpoint strike inversion	For adjacent same expiry calls, higher strike midpoint exceeds lower strike midpoint.	Mark inconsistency; not necessarily executable.
Clean midpoint inversion	Midpoint inversion after quote quality screens.	Stronger mark evidence.
Executable negative vertical	Higher strike bid exceeds lower strike ask.	Potential executable ordering violation, conditional on simultaneity and size.
Opposite direction	|Underlying return| exceeds threshold and call midpoint moves against underlying.	Directional discordance.
Bid and ask confirmed	Both bid and ask move against the underlying by at least $0.05 when |underlying return| ≥2%.	Reduces midpoint artifact concern.

Family	Operational event	Interpretive strength
IV surface rotation	Lower strike IV expands while adjacent higher strike IV contracts under a fixed rate assumption.	Cross strike surface deformation.
Zero volume repricing	Current and prior volume equal zero and |midpoint change| ≥$0.05.	Proof of quote mediated repricing; not instability by itself.
Below intrinsic / upper bound	Midpoint or executable ask violates European call bounds.	Model/quote boundary diagnostic.

6. Methodological Design Choices

6.1 Full Chain Comparison Without Moneyness/DTE Matching

Matching was excluded from the principal estimand because the research question concerns how complete chains function in their observed market ecology. Conditioning both groups onto common moneyness and DTE cells would answer a different question—behavior among selected comparable contracts—and could remove the sparse, awkward, long dated, or deep strike regions where surface instability is hypothesized to appear. Broad validity ranges remain in place. Matched sensitivity analyses may later be reported as secondary robustness tests, but they cannot replace the full chain result.

6.2 HAC Covariance Treatment

Heteroskedasticity and autocorrelation consistent covariance estimates address standard error dependence under specified conditions. They do not change fitted coefficients or R^2, eliminate multicollinearity, resolve simultaneity, establish stationarity, or identify causal direction. HAC results are therefore reported only as covariance corrections, not as a cure for time series inflation or endogeneity.

6.3 Nonlinear Surface Specifications

Polynomial expansion can mechanically improve in sample fit and can create unstable coefficients in correlated option variables. The research objective is a between group difference in fit and anomaly structure, not maximizing predictive R^2. Following the clarification that the completed comparison was linear, quadratic and cubic specifications are not treated as substantive findings in this paper. Log transformations are retained where scale and skew justify them.

6.4 Rolling Window Interpretation

The 30 trading day rolling windows share 29 observations with adjacent windows. Their fit statistics are consequently smooth, dependent processes. The chart identifies timing and divergence; it does not demonstrate that option anomalies caused underlying price changes or that SPY caused experimental chain behavior. However, the natural association must logically follow. If distortions and dysfunctions create an inoperable hedging, loss preventing, risk transferring options market, then the small cap issues affected are systemically devalued having been deprived risk transfer in a risk averse universe especially when compared to mid and large cap equities whose options markets do function as hedging, risk transferring, loss preventing having functional pricing and facilitated liquid exchange.

6.5 Zero Volume Repricing and the Instability Index

Zero volume repricing directly supports the mechanism that options can reprice without a local trade. It does not, standing alone, indicate defective pricing. Indeed, the broad pipeline rate is higher in control

(90.77%) than experimental (71.05%). Including it in a one directional instability score would conflate normal surface propagation with directional or structural inconsistency. It is therefore displayed alongside, but not inside, the Surface Instability Ratio.

6.6 MSAR Specification Selection

The three state solution was not admissible because it failed convergence/identification checks. A four state result converged but allocated only one observation to one state, producing an economically uninterpretable singleton regime. The two state MSAR(1) was retained because it converged, avoided degenerate occupancy, and satisfied the minimum regime share rule.

7. Empirical Anomaly Frequencies

Anomaly	Experimental	Control	Rate ratio
Clean midpoint inversion	1.277%	0.637%	2.00×
Clean directional divergence	4.719%	0.936%	5.04×
Opposite direction, ≥2%	10.358%	0.976%	10.61×
Bid and ask both opposite, ≥2%	2.826%	0.241%	11.72×
IV surface rotation, r=4%	3.513%	0.246%	14.30×

Anomaly	Experimental	Control	Rate ratio
Zero volume repricing ≥$0.05	71.055%	90.766%	0.78×

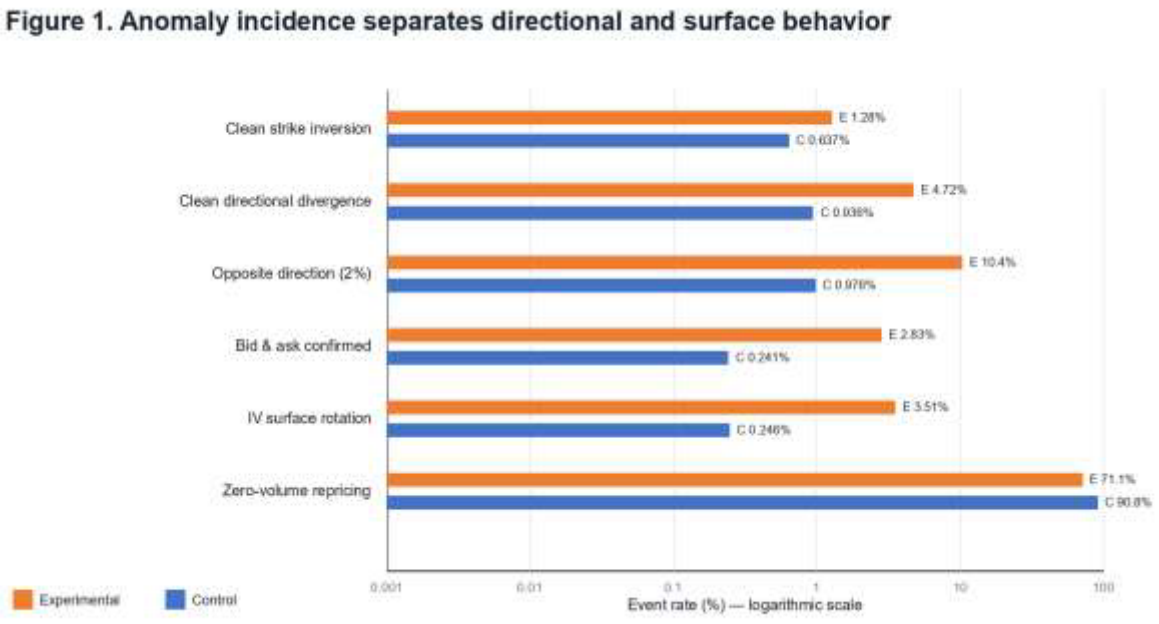

Figure 1. Anomaly incidence separates directional and surface behavior. Log scaling permits simultaneous display of sub percent surface events and high frequency zero volume repricing; denominators are anomaly specific opportunities.

The experimental/control separation is largest for surface rotation and bid and ask confirmed directional opposition. The zero volume row reverses direction. That reversal is not discarded: it identifies quote mediated repricing as common in liquid chains and prevents "no volume" from becoming a sufficient anomaly criterion.

8. Descriptive Empirics of Flagged Contracts

For the union of enriched panel zero volume, opposite direction, and bid and ask confirmed flags, the experimental group contains 232,039 contract days and the control group 897,751.

Variable	Experimental mean	Exp. median	Exp. P95	Control mean	Ctl. median	Ctl. P95
Option midpoint	$8.53	$5.15	$26.65	$96.95	$77.88	$274.65
Absolute midpoint change	$0.671	$0.350	$2.150	$3.129	$1.750	$10.625

Variable	Experimental mean	Exp. median	Exp. P95	Control mean	Ctl. median	Ctl. P95
Underlying return	0.099%	0.000%	5.315%	0.065%	0.072%	2.251%
Bid–ask spread	$2.899	$3.000	$4.900	$2.904	$3.150	$5.100
Spread / midpoint	62.96%	44.25%	178.38%	11.75%	3.68%	46.05%
Implied volatility	71.92%	63.42%	137.70%	39.21%	33.86%	84.78%
Moneyness K/S	0.863	0.803	1.412	0.810	0.768	1.380
DTE	115.75	108	234	98.77	73	242
Volume	0.760	0	0	1.881	0	0

Relative—not merely absolute—pricing friction distinguishes the samples. Experimental flagged contracts have mean spread/midpoint of 62.96%, versus 11.75% in control, while their mean IV is 71.92% versus 39.21%. Control contracts show larger absolute midpoint changes because their underlying and option price levels are much higher.

9. Surface Instability Ratio

The proposed Surface Instability Ratio (SIR) summarizes how much more frequently experimental chains exhibit four nonredundant, directionally interpretable event types than control chains. For event j, rj is the experimental rate divided by the control rate. Equal weighting is imposed in log ratio space so no single large ratio dominates by arithmetic scale:

$$\mathrm{SIR} = \exp[(1/\mathrm{J})\ \Sigma_j \ln(r_j)],\ \mathrm{J} = 4$$

The included components are clean strike inversion (2.00×), clean directional divergence (5.04×), bid and ask confirmed opposition (11.72×), and IV surface rotation (14.30×).

$$\mathrm{SIR} = (2.00 \times 5.04 \times 11.72 \times 14.30)\text{^}(1/4) = 6.41$$

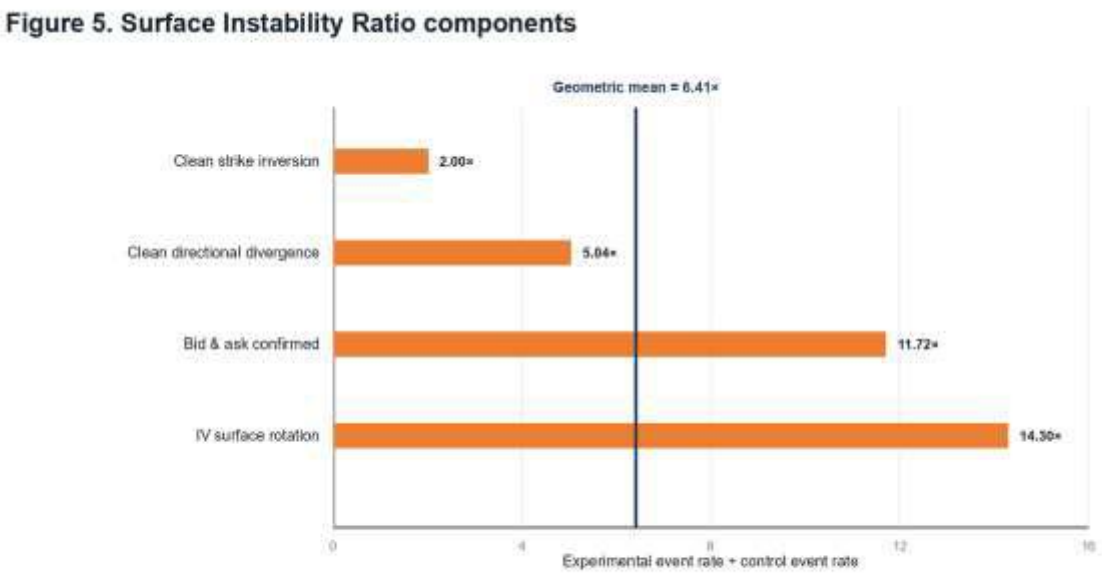

Figure 2. Surface Instability Ratio components. The vertical reference is the equal log weight geometric mean.

The vertical reference is the equal log weight geometric mean. SIR is a relative frequency index.

SIR is transparent and falsifiable. It depends on anomaly thresholds, opportunity denominators, control composition, and the choice to weight event families equally.

10. Linear Fit Difference Testing

Daily ticker level aggregates were formed from median option returns, median spread changes, median IV changes, the zero volume share, and log total volume. The dependent variable was the underlying return. Linear models used a chronological split and HAC covariance with five lags. The equations were:

$$\text{Experimental: } rS = -0.00144794 + 0.251252 \cdot rO + 0.0000629 \cdot \Delta\text{spread} - 0.197984 \cdot \Delta IV + 0.001903 \cdot ZV + 0.0000383 \cdot \log(1+\text{volume})$$

$$\text{Control: } rS = -0.00402462 + 0.096641 \cdot rO + 0.0019698 \cdot \Delta\text{spread} - 0.285610 \cdot \Delta IV + 0.003555 \cdot ZV + 0.0002770 \cdot \log(1+\text{volume})$$

Chronological holdout R^2 was 0.8563 for experimental and 0.8878 for control, a control minus experimental difference of 0.03149. A block bootstrap produced a 95% interval of [0.00190, 0.05947] and $p = 0.0180$ for the fit difference. This supports a statistically detectable group difference under the implemented aggregation, but should not be read as evidence that 85.6% of small cap price formation is caused by options.

11. Rolling Divergence and Regime Description

Thirty trading day rolling R^2 values. The red line is control minus experimental fit. Negative values indicate windows in which experimental fit exceeded control.

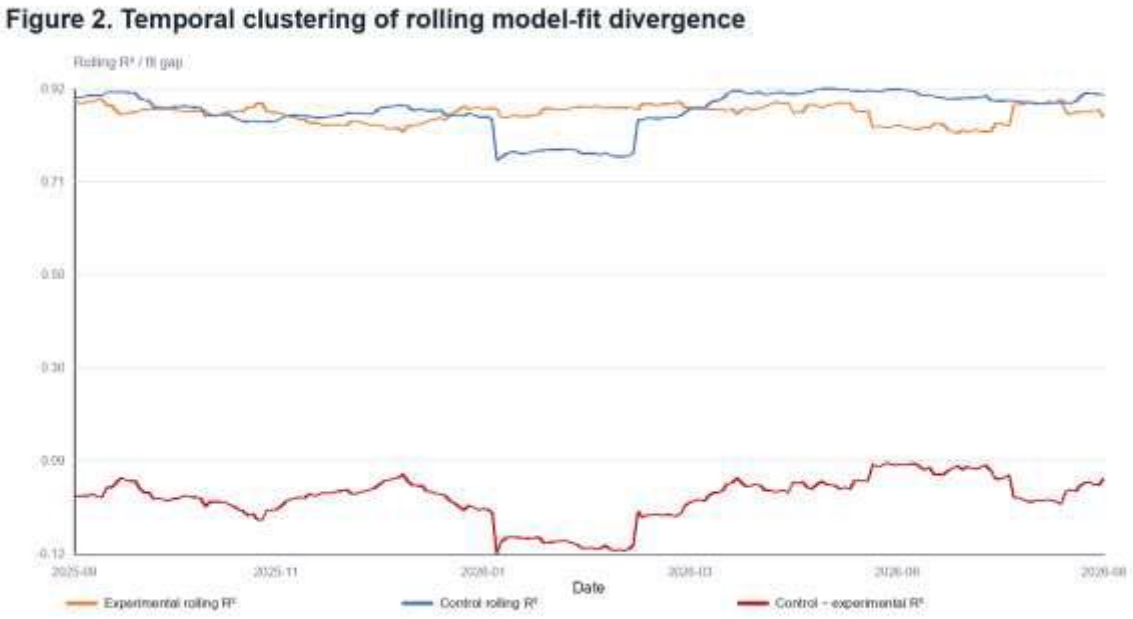

Figure 3. Temporal clustering of rolling model fit divergence. The red line is control minus experimental fit; negative values indicate windows in which experimental fit exceeded control.

Eleven dates exceeded the empirical 95th percentile of absolute fit gap magnitude ($|\Delta R^2| \geq 0.098924$): 30 January; 26–27 February; and 2–4, 6, and 9–12 March 2026. Every extreme gap was negative, meaning experimental fit exceeded control in those windows.

MSAR(1) state	Observations	Mean gap	Gap SD	Expected duration	Observed maximum run
Quiet / near zero	16	−0.000759	0.007381	4.42 trading days	10 trading days
High variance / persistent	203	0.006859	0.051618	58.08 trading days	119 trading days

The high variance state's gap variance is approximately 48.9 times the quiet state's variance. The longest observed high variance run extended from 23 September 2025 through 13 March 2026. Because rolling windows overlap, these durations describe persistence in a smoothed statistic, not independent daily regime observations.

12. Potential Repricing Frequency With or Without Volume

The user's mechanism requires a scale estimate for how often options can be repriced without a trade in the exact contract. The closest observable market wide quantity is OPRA quote message volume. The SEC explains that each exchange option price change is reported as a message and that quote traffic constitutes the vast majority of option messages. In January 2017, median daily OPRA quote messages were approximately 9 billion; by December 2025 the median was 131 billion per day—an increase of 1,347%. Cboe separately describes more than five million U.S. option trades per day and reports that the OPRA feed can peak around 90 million messages per second in the current infrastructure environment. [1–4]

Figure 4. Scale of the observable options repricing-message environment

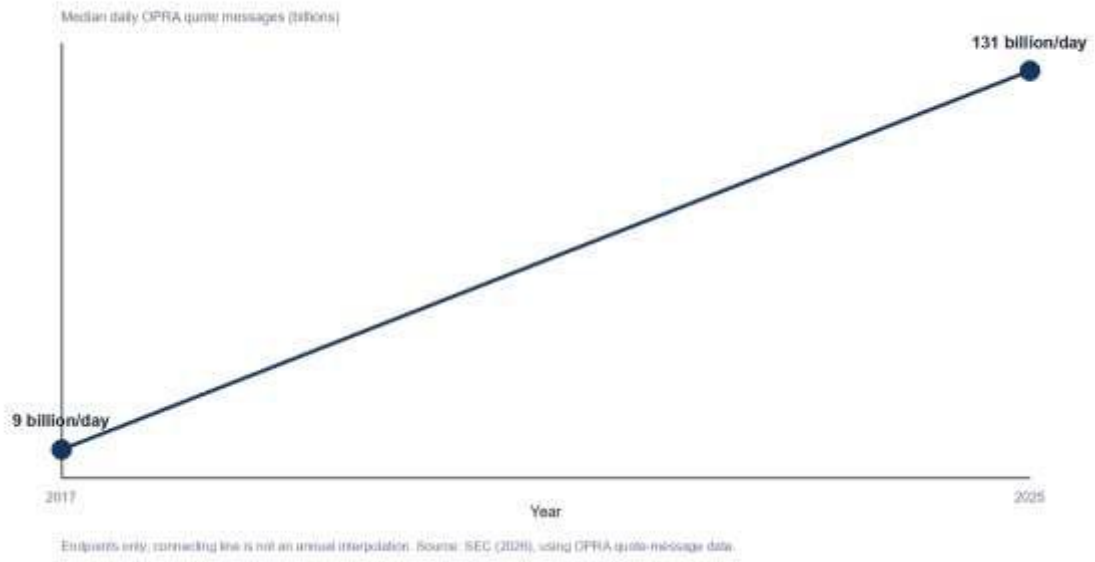

Figure 4. Scale of the observable options repricing message environment. Only the two sourced endpoints are plotted; the connecting segment shows scale change, not annual observations.

Accordingly, **approximately 131 billion observable quote messages per median day** is a defensible potential market wide repricing message count, including updates occurring with or without a trade in the quoted contract. Relative to five million trades, that is roughly 26,200 quote messages per trade. It is not valid to call all 131 billion unique fair value recomputations: the total includes exchange specific quote changes, repeated updates, size changes, cancellations, and multiple venues. Conversely, it understates private internal calculations that never become displayed OPRA messages. Cboe states that its hardware accelerated analytics can process millions of calculations per second, but no public source located here aggregates private dealer valuations into a daily national count. The paper therefore does not claim "trillions of repricings" without evidence.

The previously discussed 13.9 nanosecond STAC T0 benchmark measures minimum actionable tick to trade latency from the last necessary inbound bit to the first outbound order bit in one FPGA test configuration. It implies hardware action capacity, not observed market repricing count, and is excluded from the daily frequency numerator. [5]

13. Mathematical Architecture of the Three Dimensional Volatility Surface

A practical option surface maps strike or log moneyness and maturity into implied volatility or total implied variance. Let F(T) be the forward, k = ln(K/F(T)) log forward moneyness, T time to expiration, and σimp(k,T) implied volatility. The surface can be represented as total variance w(k,T) = σimp(k,T)²T. A quote engine then transforms a fitted surface into theoretical option values and adds bid/ask adjustments for inventory, adverse selection, hedge liquidity, capital, and competitive state.

13.1 Pricing and Inversion Layer

$$C = e\text{^}(-rT)[F \cdot N(d_1) - K \cdot N(d_2)],\ d_1 = [\ln(F/K) + \tfrac{1}{2}\sigma^2 T]/(\sigma\sqrt{T}),\ d_2 = d_1 - \sigma\sqrt{T}$$

Observed bid, midpoint, and ask prices can each be inverted for bid IV, midpoint IV, and ask IV. This produces a volatility spread as well as a price spread. Delta, gamma, vega, theta, and rho are derivatives of the pricing map with respect to spot, spot curvature, volatility, time, and rates. In a dealer system, these sensitivities aggregate across the entire inventory rather than remaining contract local.

13.2 SVI Smile Parameterization

$$w(k) = a + b\{\rho(k-m) + \sqrt{[(k-m)^2 + \sigma^2]}\}$$

For each maturity, raw SVI uses level a, slope/wing scale b, skew parameter ρ, horizontal shift m, and curvature σ. Surface construction requires interpolation across maturity and constraints against calendar and butterfly arbitrage. Gatheral and Jacquier show how SVI can be calibrated to avoid static arbitrage. [6]

13.3 Local and Stochastic Volatility Layers

Dupire local volatility derives an instantaneous σloc(K,T) from the strike and maturity derivatives of the option price surface. SABR and related stochastic volatility models allow forward price and volatility to co evolve with correlated shocks. These models are not interchangeable: they encode different dynamics even when calibrated to the same current smile. The relevant empirical point is that a platform mark reflects a calibrated field, not merely one constant Black–Scholes volatility.

13.4 Dealer Quotation Layer

$$\text{Quote}^{\text{bid}}/^{\text{ask}} = \text{ModelValue(surface, F, r, q/borrow, T, K)} \pm \text{liquidity spread} + \text{inventory adjustment} + \text{adverse selection adjustment} + \text{hedge/capital/risk limit adjustment}$$

This schematic is deliberately not presented as a single proprietary dealer equation. Published option market making models show that optimal quotes can depend on liquidity, incompleteness, stochastic volatility, inventory, and wealth. [7–9] The transcript proposition—that hedge linked, surface wide changes can reprice dead legs—is therefore compatible with established architecture, while the exact variable weights and update rules remain proprietary and unobserved.

14. Historical Confidence and Leadership Legitimacy Module

The historical extension consolidates directly quantified confidence losses surrounding financial or policy crises. The current empirical scale contains six observations. Four Euro area observations measure changes in trust in national government between 2007 and 2011; the UK observation measures Liz Truss favourability during the 2022 mini budget crisis; the U.S. observation measures trust in the federal government around the global financial crisis. Because the constructs differ, ranks are descriptive severity percentiles within this six case file, not estimates from a homogeneous global population.

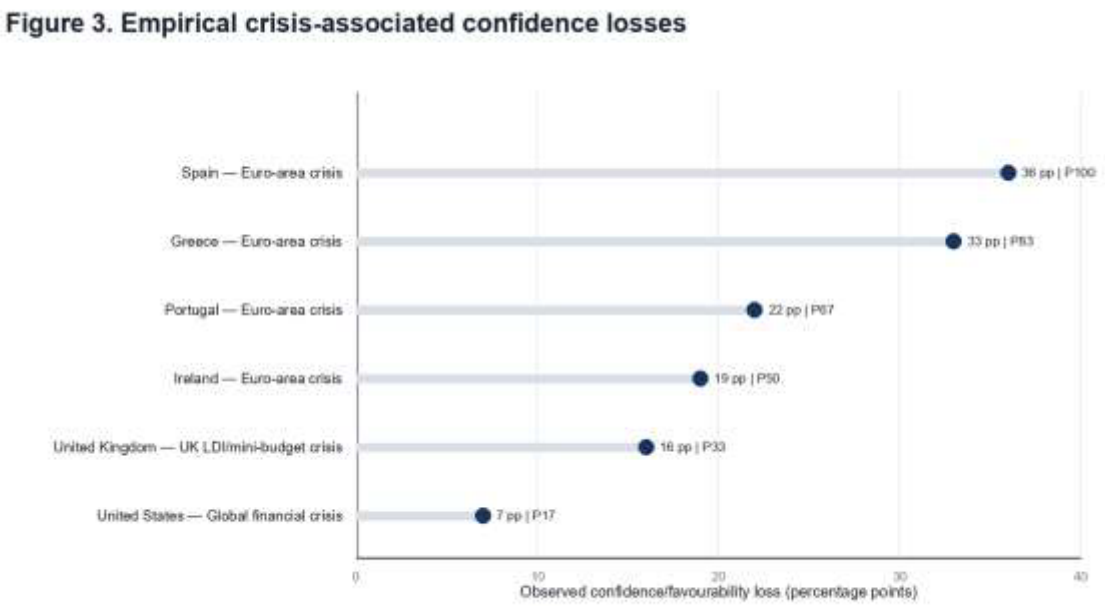

Figure 5. Empirical crisis associated confidence losses. Percentage point losses are empirical source values; percentiles rank severity only within the six quantified observations.

Case	Observed loss	Within file percentile	Construct
Spain, Euro area crisis	36 pp	100th	Trust in national government
Greece, Euro area crisis	33 pp	83rd	Trust in national government
Portugal, Euro area crisis	22 pp	67th	Trust in national government
Ireland, banking crisis	19 pp	50th	Trust in national government
United Kingdom, 2022 mini budget	16 pp	33rd	Leader favourability
United States, global financial crisis	7 pp	17th	Trust in federal government

This examination is retained as provenance of an earlier parametric branch of the systemic failure model. It reflects one of the exploratory permutations developed during approximately eight months of trial and error consideration of difficult to observe potentialities.

15. Planned Empirical Extensions

The next empirical stage can proceed without altering the current findings: construct intraday quote event panels; measure persistence and executable size; distinguish price only from size only messages; estimate ticker day anomaly intensity; calculate surface smoothness and no arbitrage residuals; produce bootstrap uncertainty for SIR; and implement predeclared robustness strata.

References

[1] U.S. Securities and Exchange Commission. Roundtable on Options Market Structure (2026), section 3.2: median OPRA quote message counts, 2017–2025. https://www.sec.gov/files/roundtable options market structure.pdf

[2] U.S. Securities and Exchange Commission. Options Price Reporting Authority rule release (2000): definition and predominance of quote messages. https://www.sec.gov/rules regulations/2000/11/options price reporting authority

[3] Cboe. Trade Alert APIs: more than five million U.S. option trades per day. https://www.cboe.com/services/analytics/tradealert/apis/

[4] Cboe. Options Analytics: OPRA peaks around 90 million messages per second and millions of analytical calculations per second. https://www.cboe.com/solutions/options analytics/

[5] STAC. Exegy/AMD FPGA STAC T0 results: 13.9 ns minimum actionable latency (2024). https://docs.stacresearch.com/news/AMD240422

[6] Gatheral, J., & Jacquier, A. Arbitrage free SVI volatility surfaces. Quantitative Finance 14(1), 2014. https://arxiv.org/abs/1204.0646

[7] Stoikov, S., & Sağlam, M. Option Market Making Under Inventory Risk. Review of Derivatives Research 12(1), 2009. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1393818

[8] Fournier, M., & Jacobs, K. A Tractable Framework for Option Pricing with Dynamic Market Maker Inventory and Wealth (2018). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2334842

[9] Guéant, O., Lehalle, C. A., & Fernandez Tapia, J. Dealing with the Inventory Risk: A solution to the market making problem. https://arxiv.org/abs/1105.3115

[10] Eurofound. Political trust and the crisis: change in trust in national government, 2007–2011. https://www.eurofound.europa.eu/api/file proxy?url=https%3A%2F%2Fa.storyblok.com%2Ff%2F279033%2Fe72d6d2a13%2Fef1374en.pdf

[11] YouGov. Liz Truss's net favourability rating falls during the 2022 mini budget crisis. https://yougov.com/en gb/articles/44092 liz trusss net favourability rating falls 70

[12] Pew Research Center. Trust in Government, 1958–2010, including the 2007–2008 decline. https://www.pewresearch.org/politics/2010/04/18/section 1 trust in government 1958 2010/

[13] Project continuity registry. "Evaluation of Academic Piece," verbatim segments 2700 and 2703; recovered dialogue segment 17178. Hash verified local evidence; not public source substitution.

[14] Project analytical outputs: anomaly family incidence, full descriptives, control/experimental comparison, rolling fit gap series, MSAR results, and historical confidence percentile file. SHA 256 validated local artifacts dated 8 August 2026.

Appendix A. Reproducible Output Inventory

Artifact	Analytical role
anomaly_family_incidence_comparison.csv	All 14 anomaly family rates for experimental and control groups.
anomalous_options_full_descriptive_statistics.csv	N, missingness, mean, SD, min, P1/P5/P25/P50/P75/P95/P99, max, and IQR for five cohorts.
experimental_control_comparison.json/.md	Linear equations, holdout fit difference, and bootstrap inference.
rolling_fit_price_chart_data.json	Thirty day rolling experimental/control fits and fit gap.
fit_gap_msar_results.json	Candidate state diagnostics, selected MSAR(1), durations, and regime runs.
leadership_legitimacy_percentile_ranks.csv	Six empirical confidence loss observations and within file percentile ranks.

Exhibit 2: Interconnectedness and Depencies of Critical Infrastructure

The U.S. infrastructure is highly interconnected and highly interdependent with networks throught system using common sources. Therefore, financial market failures are uniquely capable of producing adjacent sector sub optimal performance or failure.