Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Legal and Ethical Implications of AI-Driven Financial Prediction Systems in Modern Markets

Om Aditya Mishra
Dept. of Computer Science and Engineering
(Cyber Security)
R.V. College of Engineering
Bengaluru, India

Pallavi O
Dept. of Computer Science and
Engineering (Cyber Security)
R.V. College of Engineering
Bengaluru, India

Chitra B T
Dept. of Industrial Engineering and Management
R.V. College of Engineering
Bengaluru, India

Abstract - **Financial markets have undergone a structural overhaul driven by the pervasive adoption of Artificial Intelligence (AI) and Machine Learning (ML) in trading, risk assessment, and portfolio management. While these technologies have measurably enhanced execution efficiency and predictive capability, the pace of their deployment has substantially outrun the regulatory frameworks of leading jurisdictions—including SEBI, the U.S. Securities and Exchange Commission (SEC), and the European Securities and Markets Authority (ESMA). The resulting governance deficit introduces layered legal and ethical hazards: algorithmic market manipulation, a reconstituted form of insider trading premised on AI-generated informational asymmetries, systemic fragility from correlated automated strategies, and an emergent class of harms that existing scholarship has inadequately examined. This paper identifies five such underexplored research gaps—autonomous inter-agent collusion, AI-washing in investment disclosures, fiduciary obligation in AI-mediated advisory relationships, cross-border regulatory arbitrage, and the carbon externality of compute-intensive trading—and integrates them into a cohesive analytical framework. We propose a Tiered Regulatory Compliance Framework (TRCF) capable of classifying AI trading systems by risk profile, and an Extended Ethical Compliance Score (S_e) whose formula is augmented with penalty terms for environmental cost and disclosure infidelity. The framework is then evaluated against four documented market disruptions to assess its prospective mitigative value. Findings indicate that a granular, evidence-weighted, and multi-jurisdictionally harmonised oversight regime is necessary to preserve market integrity without foreclosing legitimate financial innovation.**

Index Terms—**Artificial Intelligence, Algorithmic Trading, Financial Regulation, SEBI, SEC, Market Manipulation, Insider Trading, Fiduciary Duty, AI-Washing, Regulatory Arbitrage, Ethical Compliance, ESG.**

I. Introduction

Contemporary financial markets are no longer primarily driven by human traders exercising discretionary judgment. Instead, they are shaped by algorithmic agents operating at microsecond latencies, processing vast data streams that no human analyst could realistically evaluate. By 2024, algorithm-driven strategies accounted for an estimated 60–75% of U.S. equity trading volume; the figure for India's National Stock Exchange had crossed the 50% threshold for the first time in the same year [1]. Global AI spending within the financial sector is projected to exceed $97 billion by 2027 [2], reflecting the industry's conviction that AI-derived predictive advantages translate directly into competitive returns.

This trajectory has, however, produced a consequential regulatory shortfall. The legal architecture that governs securities markets—including SEBI's algorithmic trading circulars, the SEC's Regulation SCI, and the UK's transposed MiFID II obligations—was designed for rule-based, deterministic systems whose decision logic could, in principle, be traced and audited. Modern AI systems, particularly those employing Deep Learning or Reinforcement Learning, defy this model. Their decision pathways are opaque not only to regulators but often to the engineers who designed them [3]. This opacity complicates the assignment of liability when an AI-driven strategy causes a market disruption and renders the *mens rea* standard of traditional fraud law effectively unworkable.

Beyond the question of liability, AI's capacity to derive systematic trading advantages from the synthesis of high-velocity alternative data—satellite imagery of retail parking lots, anonymised credit-card transaction flows, social media sentiment indices—raises foundational questions about market equity. These data sources are, in the narrow legal sense, publicly available; yet the analytical infrastructure required to exploit them is accessible only to well-capitalised institutional actors, effectively recreating the information asymmetry that insider trading law was designed to prevent [5].

Existing scholarship has addressed several of these concerns, yet a cluster of important dimensions remains inadequately theorised. Specifically: (i) the emergence of autonomous collusive equilibria among independently-operating AI agents [17]; (ii) the misrepresentation of AI capabilities in investment product disclosures—analogous to the earlier phenomenon of greenwashing [9]; (iii) the fragmented legal treatment of fiduciary duty in AI-mediated advisory relation-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

ships [10]; (iv) the structural incentive for firms to exploit divergences between national regulatory frameworks through jurisdictional arbitrage [3]; and (v) the environmental cost of compute-intensive high-frequency strategies and its tension with ESG investment mandates [13].

This paper addresses those gaps alongside the primary regulatory and ethical questions. Section II surveys the current legal landscape through a comparative jurisdictional lens. Section III formally identifies and theorises the five research gaps. Section IV develops the threat model. Section V proposes the Tiered Regulatory Compliance Framework and the Extended Ethical Compliance Score. Section VI evaluates the framework against historical disruptions. Sections VII and VIII consider industry implications and policy recommendations. Section IX concludes.

II. Related Work and Legal Landscape

The governance of algorithmic trading has passed through two broad phases. The first, roughly coinciding with the early proliferation of electronic order routing in the 2000s, was dominated by concerns about technical stability and system integrity—reflected in rules such as the SEC's Regulation SCI. The second phase, precipitated by a succession of high-profile failures including the 2010 Flash Crash and the Knight Capital incident, expanded the regulatory lens to encompass market integrity, manipulative conduct, and systemic risk. The emergence of opaque ML systems has initiated a third phase in which the very conceptual vocabulary of financial regulation—intent, authorship, causation, disclosure—is under strain.

A. The Responsibility Gap in AI-Driven Harm

Armour and Eidenmüller's analysis of "self-driving corporations" [4] identifies what they term the responsibility gap: the difficulty of attributing blame to a legal person when the proximate cause of harm is an autonomous system acting within, but not necessarily in accordance with, its designer's intent. In securities law this gap is acute because market abuse frameworks—whether under Section 10(b) of the Securities Exchange Act, SEBI's Prevention of Fraudulent and Unfair Trade Practices Regulations, or the EU's Market Abuse Regulation—are premised on the existence of a culpable human actor whose mental state can be imputed. Where an AI agent learns spoofing behaviour through reward-maximisation without any human programmer encoding that behaviour explicitly, the existing frameworks provide no satisfactory mechanism for prosecution.

B. Comparative Jurisdictional Analysis

Major regulatory authorities have adopted meaningfully different approaches to AI in trading, creating a patchwork that is itself a governance risk. Table I summarises the four most significant frameworks.

C. Insider Trading Doctrine in the Age of Alternative Data

The classical insider trading framework—whether the classical theory under *Chiarella v. United States* or the misappropriation theory from *United States v. O'Hagan*—requires that the defendant possessed material, non-public information (MNPI). The doctrinal challenge posed by AI-driven alternative data analysis is that the underlying data is, in the technical sense, public: satellite imagery, shipping manifests, and anonymised transactional records are commercially available. What is non-public, and practically inaccessible to ordinary market participants, is the analytical output—the predictive insight generated by applying substantial computational resources to that data [5]. Existing scholarship has flagged this grey area but has not produced a workable doctrinal reformulation. The present paper argues that the appropriate analogy is not to the possession of information but to the possession of an unfair analytical infrastructure, and that regulatory reform should focus on disclosure requirements for alternative data strategies rather than on extending the MNPI concept.

D. Existing Technical Literature and Its Limits

On the technical side, research has examined the microstructure effects of algorithmic trading, including the relationship between HFT participation and liquidity [18], the mechanics of spoofing detection [6], and the systemic amplification effects of correlated strategy deployment [7]. What is largely absent from this literature is an integrative framework that connects these technical findings to enforceable legal standards and that accounts simultaneously for the multi-agent, cross-modal, and cross-jurisdictional dimensions of modern AI trading systems. The framework proposed in this paper is designed to fill that integrative function.

III. Identified Research Gaps

A systematic review of the regulatory, legal, and technical literature reveals five substantive areas that have not received proportionate scholarly or policy attention relative to their practical significance. These gaps are not merely theoretical lacunae; each corresponds to a class of real or imminent market harm. Table II provides a structured overview before each gap is treated in depth.

A. Gap 1: Autonomous Collusion in Multi-Agent AI Systems

The possibility that competing AI trading agents might independently arrive at collusive pricing strategies—without any form of explicit communication—was demonstrated with rigour by Calvano et al. in a landmark 2020 study [17]. Their experimental setup involved independent Q-learning agents optimising profit in a repeated Bertrand game; the agents consistently learned to sustain supracompetitive prices through what the authors characterise as a pattern of strategic underreaction. The mechanism is entirely emergent: no programmer encoded collusion, and no agent communicated with another. Regulators and competition authorities have taken note. The FCA's April 2024 AI Update explicitly flagged the risk that requiring AI systems to report each other's manipulative behaviour could produce an adversarial learning dynamic in which detection systems and manipulative algorithms coevolve to outmanoeuvre each other [25].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

TABLE I
COMPARATIVE REGULATORY LANDSCAPE FOR AI TRADING

Jurisdiction	Regulator	Primary Framework	Key Provisions	Gap Coverage
India	SEBI	Algo Trading Circulars (2022–26)	Broker accountability; White Box / Black Box distinction; API access restrictions.	Manipulation, systemic risk; collusion and arbitrage understated.
USA	SEC	Reg SCI & Market Access Rule (Rule 15c3-5)	Pre-trade risk controls; systems compliance and integrity reporting; 2024 AI-washing examinations.	Manipulation, disclosure; fiduciary AI gap persists.
UK	FCA	MiFID II (Transposed)	Rigorous pre-deployment testing; governance framework; April 2024 AI Update on inter-agent dynamics.	Systemic and collusion risks beginning to be addressed.
EU	ESMA / EU	EU AI Act (2024) & MiFID II	Financial AI as High-Risk (Annex III); explainability mandates; supervisory cooperation.	Broadest coverage; arbitrage risk from non-EU jurisdictions remains.

TABLE II
FIVE UNDEREXPLORED RESEARCH GAPS IN AI FINANCIAL REGULATION

Gap	Core Problem	Why Understudied	Proposed Contribution
1. Autonomous Collusion	Multi-agent RL systems converge on supra-competitive equilibria without explicit coordination.	Single-system threat models dominate; multi-agent dynamics treated as antitrust rather than securities law issue.	Extended Class IV threat model; multi-agent sandbox stress-testing requirement.
2. AI-Washing	Firms overstate AI sophistication in product disclosures to attract capital.	Greenwashing literature is the nearest precedent; specific AI-disclosure fraud framework is nascent.	Disclosure fidelity penalty term (D_f) in S_e; taxonomy for AI capability claims.
3. Fiduciary Duty	Liability allocation for AI robo-advisor losses is legally unresolved.	Fiduciary doctrine is human-centric; robo-advisory liability addressed in isolation without general framework.	Human-in-the-Loop Accountability Matrix (HLAM); three-tier oversight intensity model.
4. Regulatory Arbitrage	Regulatory divergence creates structural incentives to domicile AI systems in lax jurisdictions.	Arbitrage is studied in traditional finance; AI-specific cross-border migration patterns are undermodelled.	Regulatory Harmonisation Index (RHI); bilateral consultation trigger at RHI < 0.7.
5. Environmental Cost	Compute-intensive trading carries a significant carbon footprint; ESG mandate tension is unaddressed.	Energy cost treated as IT infrastructure issue, not a financial ethics or regulatory concern.	Energy-intensity score (E_c) integrated into S_e formula.

This threat is not addressed by any current regulatory framework. SEBI's algo circulars, the SEC's market manipulation rules, and the EU AI Act all conceptualise manipulation as a property of a single system and its operator. The multi-agent dimension—where the harm emerges from the interaction of independently lawful systems—falls into a structural gap. The threat model developed in Section IV extends the existing typology to accommodate this category, and the TRCF's Class IV requirements are expanded accordingly.

B. Gap 2: AI-Washing in Investment Product Disclosures

The SEC's Division of Examinations identified AI-washing as a formal examination priority in 2024, following a 2023 sweep that found numerous registered investment advisors making unsubstantiated claims about AI-driven portfolio management [9]. The 2025 examination priorities further expanded this oversight, requiring firms to demonstrate that their representations about AI capabilities are accurate and that adequate policies exist to supervise AI use. The parallel with greenwashing is instructive: in both cases, the harm arises from a material misrepresentation that induces investor reliance. The

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

difference is that AI capability claims are harder to falsify than ESG metrics, because the analytical systems being described are often proprietary and opaque.

Academic frameworks for financial AI regulation have not yet integrated this form of disclosure fraud. The present paper treats AI-washing as a sub-category of material misrepresentation under established securities fraud doctrine—specifically Rule 10b-5 under the Securities Exchange Act and comparable provisions in SEBI's PFUTP regulations—and proposes a disclosure fidelity penalty term (D_f) within the ethical compliance score that regulators can use to quantify the severity of misrepresentation during examination proceedings.

C. Gap 3: Fiduciary Duty and Autonomous AI Advisory Systems

The fiduciary relationship between an investment advisor and a client has historically been analysed as a human relationship: the advisor's duty of care and duty of loyalty attach to a natural person who makes informed, contextual judgements on behalf of the client. The emergence of robo-advisory platforms—autonomous AI systems that construct and rebalance portfolios without ongoing human intervention—introduces a structural mismatch. When the system's recommendation causes client harm, the allocation of responsibility between the software developer, the deploying firm, and any supervising human advisor is legally unresolved in most jurisdictions.

Venable LLP's 2025 analysis states plainly that "delegating decisions to a machine does not absolve the human fiduciary from oversight" [10], and ESMA has similarly held that financial institutions "must take full responsibility for the actions of AI systems they deploy." Yet these positions, while legally coherent, do not resolve the practical question of what oversight is necessary and sufficient for a firm to discharge its fiduciary obligations. The present paper proposes a Human-in-the-Loop Accountability Matrix (HLAM) that maps decision categories to oversight intensity requirements, providing an operational framework that legal and compliance functions can implement.

D. Gap 4: Cross-Border Regulatory Arbitrage

Regulatory arbitrage—the practice of structuring activities to exploit differences between national legal regimes—is a well-documented phenomenon in traditional finance. Its AI-specific manifestation has received far less attention. The IMF's 2025 report on AI in securities markets explicitly warns that firms with advanced AI capabilities may bypass stricter oversight mechanisms in their home jurisdictions by routing their AI infrastructure through countries with less developed regulatory frameworks [3]. This dynamic is structurally incentivised: compliance cost estimates for the EU AI Act alone run to approximately €29,277 per AI product [12], creating a meaningful financial motivation to seek lighter-touch regimes.

The implications extend beyond cost arbitrage. A firm that locates its AI trading infrastructure in a jurisdiction without real-time monitoring requirements effectively exports the systemic risk of its strategies to markets that may lack the capacity to contain a resulting disruption. This paper proposes a Regulatory Harmonisation Index (RHI) computed as the cosine similarity between jurisdiction-level regulatory requirement vectors across five dimensions—pre-trade controls, explainability mandates, real-time monitoring obligations, sandbox requirements, and fiduciary standards. An RHI below 0.7 between any two major jurisdictions is treated as a presumptive arbitrage risk warranting bilateral regulatory coordination under IOSCO protocols [11].

E. Gap 5: Environmental Cost and the ESG Contradiction

The carbon footprint of large-scale AI model training and inference is now a mainstream concern in technology policy, yet it has received almost no attention in the financial ethics or financial regulation literature. ESG-labelled assets under management globally reached an estimated USD 41–50 trillion by 2025 [13], reflecting a broad commitment by the financial sector to environmental responsibility. At the same time, the deep learning models and high-frequency infrastructure underpinning AI trading strategies are among the most compute-intensive applications in commercial use. There is a structural contradiction between a firm's ESG investment mandates and the energy footprint of the AI systems through which those mandates are executed.

This paper argues that the omission of energy-intensity considerations from AI trading ethics frameworks is analytically incomplete and operationally inconsistent with the sector's stated ESG commitments. The extended ethical compliance score proposed in Section V incorporates an energy-intensity score (E_c) as a subtractive term, providing a mechanism through which energy use can be factored into regulatory assessments and public disclosures.

IV. System Architecture and Threat Model

A complete account of the risks associated with AI-driven financial prediction requires a threat model that extends beyond the single-system perspective that characterises most existing regulatory analysis. The model developed here classifies threats across five categories, distinguishes between single-agent and multi-agent harm mechanisms, and incorporates the evidentiary and jurisdictional dimensions identified in the research gap analysis.

A. Single-Agent Threat Categories

1) Market Manipulation Through Learned Behaviour: AI systems optimising for profit in market microstructures can learn manipulative strategies—spoofing, layering, momentum ignition—as instrumentally useful without any programmer explicitly encoding such behaviour [6]. The absence of human intent does not eliminate harm; it merely complicates attribution. Under the proposed framework, Class III and IV systems are subject to real-time monitoring of order-to-trade ratios and submission patterns precisely to detect emergent manipulative behaviour.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

2) Correlated Strategy and Flash Crash Risk: Where multiple institutions deploy AI systems trained on similar data using similar architectures, the resulting trading strategies may be highly correlated. In a stress scenario, simultaneous de-risking by multiple AI systems can produce the self-reinforcing liquidity withdrawal that characterised the 2010 Flash Crash [7]. The systemic dimension of this risk distinguishes it from single-firm manipulation and requires a different regulatory response—one focused on portfolio diversity requirements and circuit breaker mechanisms rather than intent-based liability.

3) Alternative Data and Informational Asymmetry: The deployment of AI systems capable of deriving predictive signals from alternative data creates a form of market asymmetry that is functionally equivalent to insider trading but falls outside its current legal definition [5]. The harm is not that private information is misappropriated but that the effective informational playing field is radically skewed toward actors with the computational resources to process publicly available data at scale. Regulatory responses should focus on alternative data strategy disclosure rather than on extending MNPI doctrine.

B. Multi-Agent Threat: Autonomous Collusion

Building on the analysis in Section III-A, the multi-agent threat model treats collusion as an emergent property of the interaction between independently-operating AI systems rather than as a product of any single system's design. The formal characterisation follows from Calvano et al. [17]: let $A = \{a_1, a_2, \ldots, a_n\}$ be a set of AI trading agents each independently maximising expected discounted profit. In the presence of strategic complementarity—where each agent's optimal strategy depends on the strategies of others—repeated interaction can produce a Nash equilibrium that is collusive without any agent encoding collusion as an objective. The policy implication is that sandbox stress-testing for Class IV systems must include multi-agent adversarial scenarios in which candidate deployments are tested against existing market participants.

C. Disclosure and Representation Threats

The AI-washing threat identified in Section III-B operates at the interface between securities law and technology assessment. The harm mechanism is straightforward: a firm makes material representations about the sophistication or performance of its AI systems in offering documents, marketing materials, or regulatory filings; those representations induce investor reliance; the representations prove false or materially misleading. The specific challenge for regulators is that AI capability claims are harder to verify than conventional financial disclosures because the systems being described are proprietary and technical assessment requires specialised expertise. The D_f penalty term in the extended S_e score is designed to translate the degree of disclosure infidelity into a quantitative compliance indicator.

V. Tiered Regulatory Compliance Framework and Ethical Scoring Model

The framework proposed here has three interlocking components: a risk classification taxonomy that maps AI trading systems to discrete regulatory tiers; an extended ethical compliance score that quantifies each system's risk-adjusted ethical standing; and a set of operational instruments—the HLAM and RHI—that address the fiduciary duty and regulatory arbitrage gaps respectively.

A. Risk Taxonomy and Tiered Classification

Table III presents the Tiered Regulatory Compliance Framework. The classification is based on two primary parameters: the opacity of the system's decision mechanism and the nature of the data on which it operates. These parameters are operationally proxied, respectively, by whether the system's decision logic can be extracted and audited in interpretable form, and by whether the system's training and inference data is drawn exclusively from structured, exchange-provided sources or incorporates alternative, unstructured inputs.

B. Extended Ethical Compliance Score

The base ethical compliance score proposed in earlier work captures transparency, accountability, fairness, and systemic risk. The present framework extends this by adding two penalty terms derived from the research gap analysis—an energy-intensity penalty (E_c) addressing Gap 5 and a disclosure fidelity penalty (D_f) addressing Gap 2. The extended formula is:

$$S_e = w_1(T) + w_2(A) + w_3(F) - w_4(R) - w_5(E_c) - w_6(D_f) \quad (1)$$

where each term is defined as follows. $T \in [0,1]$ is the Transparency and Explainability Index, reflecting the degree to which the system's decision process can be extracted, audited, and explained in terms that a financially sophisticated regulator can evaluate. $A \in [0,1]$ is the Accountability and Liability Mapping score, capturing the clarity with which responsibility for system outputs is allocated across the developer, deploying firm, and supervising human advisor. $F \in [0,1]$ is the Market Fairness Index, measuring the extent to which the system's operational advantage derives from capabilities that are, in principle, accessible to a broad range of market participants rather than from exclusive data or infrastructure advantages.

$R \in [0,1]$ is the Systemic Contagion Risk score, quantifying the extent to which the system's strategy correlates with those of other market participants and the potential scale of market impact in a stress scenario. $E_c \in [0,1]$ is the Energy-Carbon Intensity Score, calibrated against a benchmark of compute per unit of notional trading volume. $D_f \in [0,1]$ is the Disclosure Fidelity Penalty, assessed by the regulator on the basis of the gap between the firm's representations about its AI capabilities and the empirically-assessed capabilities of the deployed system. Weights w_1 through w_6 are set by the relevant national regulator and may be adjusted by jurisdiction

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

TABLE III
TIERED REGULATORY COMPLIANCE FRAMEWORK (TRCF)

Class	System Type	Risk Level	Transparency Requirement	Monitoring Requirement	Intervention
Class I	Deterministic rule-based systems with fully auditable logic.	Low	Full logic disclosure to broker and exchange.	Standard pre-trade risk controls; post-trade reporting.	None beyond standard registration.
Class II	Supervised ML models using structured, exchange-sourced data.	Moderate	Explainability documentation; feature importance reports.	Periodic audit; out-of-sample performance disclosure.	Regulatory review on anomalous performance signals.
Class III	Deep Learning / ensemble systems using alternative or unstructured data.	High	Model card; training data provenance; inference audit trail.	Real-time order-to-trade ratio monitoring; kill-switch mandate.	Mandatory notification within 30 minutes of anomalous behaviour.
Class IV	Autonomous RL systems capable of self-modification; multi-agent deployments.	Critical	Full architecture disclosure to regulator; third-party audit.	Regulator access to live telemetry; adversarial multi-agent sandbox pre-deployment.	Deployment suspension on any anomaly; human re-authorisation required.

to reflect local policy priorities, provided that any deviation from a standard weighting schedule is publicly disclosed.

An S_e score below 0.5 constitutes a presumptive compliance failure requiring regulatory intervention. A score between 0.5 and 0.74 triggers enhanced monitoring. A score of 0.75 or above represents baseline compliance. These thresholds are calibrated to the severity scale implicit in the TRCF tiers and are intended to provide regulators with a continuous, auditable signal that is more informative than the binary pass/fail outcomes of current pre-approval regimes.

C. Human-in-the-Loop Accountability Matrix

The HLAM addresses the fiduciary duty gap by specifying, for each decision category in an AI advisory context, the minimum level of human oversight necessary for the deploying firm to discharge its fiduciary obligations. Three oversight intensity levels are defined. Level 1 (ex-post review) applies to routine execution decisions—order routing, intraday rebalancing—where the decision is individually immaterial and human review within 24 hours is sufficient. Level 2 (pre-authorisation) applies to portfolio-level decisions—strategic allocation shifts, entry into or exit from asset classes—where a licensed human advisor must review and approve the AI's recommendation before execution. Level 3 (independent regulatory clearance) applies to the deployment of novel strategies not previously validated in production, where pre-deployment sandbox testing with regulator access to test results is required.

D. Regulatory Harmonisation Index

The RHI is computed as the cosine similarity between pairs of jurisdiction-level regulatory requirement vectors. Each vector has five dimensions corresponding to the five core regulatory requirements of the TRCF: pre-trade risk controls (w_{ptc}), explainability mandates (w_{xai}), real-time monitoring obligations (w_{rtm}), sandbox requirements (w_{sbx}), and fiduciary standards (w_{fid}). For jurisdictions J_1 and J_2:

$$RHI(J_1, J_2) = \frac{v(J_1) \cdot v(J_2)}{\|v(J_1)\| \times \|v(J_2)\|} \tag{2}$$

An RHI below 0.7 between any two jurisdictions in which a firm operates is treated as a presumptive arbitrage risk. Firms operating across low-RHI jurisdiction pairs are subject to enhanced reporting requirements and must demonstrate that they apply the stricter of the two regulatory standards to their AI systems regardless of where those systems are domiciled. IOSCO is proposed as the coordinating body for bilateral RHI reviews, consistent with its existing mandate for cross-border market oversight [11].

VI. RESULTS AND EVALUATION

The proposed framework is evaluated through two complementary methods: a structured case study assessment that tests whether the TRCF and extended S_e score would have detected or mitigated four documented market disruptions, and a gap-validation analysis that examines the extent to which recent regulatory and judicial developments corroborate the significance of the five identified research gaps.

A. Case Study Evaluation

Table IV applies the TRCF classification and S_e components to four cases. The analysis is counterfactual: it asks what outcome the framework would have predicted or produced had it been in force at the time.

B. Gap Validation Analysis

The significance of the five identified research gaps is independently corroborated by recent regulatory and judicial developments. For Gap 1, the FCA's 2024 explicit flagging of adversarial inter-agent learning dynamics [25] confirms that collusion risk is now a live regulatory concern rather than a theoretical one. For Gap 2, the SEC's formal inclusion of AI-washing in its 2024 and 2025 examination priorities [9] validates the identification of disclosure misrepresentation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

TABLE IV
CASE STUDY EVALUATION OF THE PROPOSED FRAMEWORK

Case	Date	Primary Failure	TRCF Class	Key S_e Signal	Predicted Intervention
2010 Flash Crash	May 6, 2010	Spoofing algorithm triggered self-reinforcing liquidity withdrawal; ∼1,000-point DJIA decline in minutes.	Class III	Fairness score $F \approx 0.10$ (predatory order patterns); Systemic risk $R \approx 0.95$.	Real-time order-to-trade ratio monitoring would have flagged anomalous cancellation patterns before liquidity vacuum formed; kill-switch activation.
Knight Capital	Aug 1, 2012	Deployment of legacy test software caused $440M loss in 45 minutes through unintended aggressive order execution.	Class I	Accountability score $A \approx 0.20$ (absent deployment verification protocol).	Mandatory pre-trade controls and deployment verification under Class I registration would have blocked unvalidated code from accessing live markets.
SEC v. Athena Capital	Oct 16, 2014	HFT firm used "Gravy" algorithm to artificially mark closing prices of thousands of NASDAQ-listed stocks.	Class III	$F \approx 0.05$ (systematic end-of-day price distortion); $T \approx 0.30$.	S_e below 0.5 triggers compliance failure; real-time monitoring of closing-auction order patterns detects systematic distortion.
SEBI Orders 2022–23	2022–2023	Brokers made misleading claims about guaranteed returns from "algo" strategies; retail investors misled.	Class II / III	$D_f \approx 0.80$ (material misrepresentation of AI capabilities).	D_f penalty drives S_e to compliance-failure range; mandatory disclosure verification during registration prevents misrepresentation.

as an enforcement-priority issue. For Gap 3, ESMA's 2024 statement that financial institutions bear full responsibility for AI system outputs [10] represents the most authoritative judicial-adjacent confirmation that fiduciary duty extends to AI advisory systems. For Gap 4, the IMF's 2025 warning about regulatory arbitrage in AI-enabled trading [3] provides multilateral institutional confirmation. For Gap 5, the contradiction between ESG mandates and AI energy footprints has been independently identified in the sustainable finance literature [13], [24].

C. Robustness of the S_e Score

A sensitivity analysis of the S_e formula across plausible ranges of weight values (w_1 through w_6 varied uniformly between 0.10 and 0.25) confirms that the compliance-failure threshold at $S_e < 0.5$ is robust for systems with clear manipulative characteristics: for the 2010 Flash Crash scenario, S_e remains below 0.35 across all weight combinations tested. For borderline cases—systems with moderate transparency and modest systemic risk scores—the outcome is more sensitive to weighting, which provides the rationale for requiring public disclosure of the weighting schedule adopted by each regulator.

VII. ENTREPRENEURIAL AND INDUSTRY IMPLICATIONS

A. Compliance Cost and the Fintech Startup Ecosystem

The regulatory requirements contemplated by the TRCF place a disproportionate burden on early-stage fintech ventures relative to established institutional participants. SEBI and SEC compliance infrastructure—including explainability documentation, real-time monitoring systems, and regulatory sandbox participation—may consume up to 20% of an early-stage firm's operational budget [8]. The EU AI Act's compliance cost of approximately €29,277 per AI product [12] is manageable for a large financial institution but potentially prohibitive for a startup whose entire AI portfolio may consist of a single system. A tiered compliance cost structure—in which Class I requirements are minimally burdensome and Class IV requirements are graduated according to the scale of the firm's market presence—would preserve the regulatory intent of the TRCF while reducing the barrier to entry for legitimate innovation.

B. Investor Appetite for Explainable AI

The AI-washing enforcement trend has a secondary market effect: venture capital and institutional investors are increasingly applying enhanced due diligence to AI capability claims in fintech investment proposals. This is reflected in a growing preference for Explainable AI (XAI) architectures, which

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

allow the investment thesis to be verified independently of the developer's representations. AI startups that can demonstrate TRCF Class II or Class III compliance with strong T and A scores—transparency and accountability—are likely to command a material valuation premium over comparable firms operating "black box" strategies, analogous to the ESG premium documented in the sustainable investment literature [24].

C. Dispute Resolution in AI-Enabled Markets

The fiduciary duty gap identified in Section III-C has direct implications for how trading disputes are resolved. Where an AI-driven strategy causes client losses, the firm's ability to demonstrate HLAM-compliant oversight—documented evidence of human pre-authorisation for Level 2 decisions and regulator-cleared sandbox results for Level 3 deployments—will be the central evidentiary question in any subsequent regulatory investigation or civil claim. Firms that have not implemented the HLAM face the prospect of being unable to mount a compliance defence, because they will be unable to demonstrate that adequate human oversight was in place. In this sense, the HLAM is not merely a regulatory obligation but an instrument of litigation risk management.

D. Strategic IP and Data Governance

AI trading systems derive their competitive advantage partly from proprietary training data pipelines and partly from model architecture innovations. Both are increasingly recognised as forms of intellectual property with uncertain legal status—analogous to the AI-generated IP issues examined in the parallel literature [23]. Firms operating under the TRCF's Class III and IV requirements—which mandate training data provenance disclosure—face a potential tension between regulatory transparency obligations and the protection of commercially sensitive data assets. Legal frameworks for managing this tension, including regulatory sandboxes in which proprietary information is disclosed to regulators under confidentiality, represent an important area for further institutional development.

VIII. Legal Positioning and Policy Implications

The TRCF and extended S_e score are designed as evidentiary and analytical instruments that complement, rather than replace, existing securities law frameworks. They operate at the interface between technical system assessment and legal liability determination, providing the quantitative indicators that existing doctrine lacks. Verifiable records of S_e scores, HLAM compliance documentation, and RHI calculations constitute a form of regulatory evidence base that can support enforcement actions, due diligence processes, and dispute resolution proceedings.

A. Recommendations for SEBI (India)

SEBI should establish a centralised Algorithmic Trading Sandbox in which all Class III and Class IV systems must undergo pre-deployment stress-testing, including adversarial multi-agent scenarios designed to probe for emergent collusive behaviour. The sandbox should be supplemented by a mandatory disclosure regime for alternative data strategies that requires firms to characterise the nature and source of their data inputs without disclosing proprietary model details. SEBI's broker accountability framework should be extended to specify HLAM-equivalent oversight requirements for AI-mediated wealth management and advisory services, addressing the fiduciary duty gap in the domestic context.

B. Recommendations for the SEC (USA)

The SEC should formalise AI-washing as a sub-category of material misrepresentation under Rule 10b-5, with specific guidance on what constitutes an adequate and verifiable AI capability claim in offering documents and marketing materials. The existing MNPI framework should be supplemented with a disclosure requirement for alternative data strategies that relies on the regulatory concept of "analytical infrastructure access" rather than information possession. The SEC should also engage with IOSCO to develop the RHI framework as a multilateral instrument for identifying and addressing regulatory arbitrage risks across major trading jurisdictions.

C. Recommendations for ESMA and the EU

The EU AI Act's High-Risk classification for financial AI should be supplemented with specific technical standards for the energy-intensity disclosure requirement proposed in this paper, integrating the E_c component of the S_e score into the Act's existing conformity assessment obligations. ESMA's supervisory briefing on AI compliance should address the multi-agent collusion risk explicitly and specify that sandbox assessments for Class IV systems must include adversarial multi-agent testing. The EU should use its regulatory leadership position to promote RHI-based harmonisation with non-EU jurisdictions through the Commission's equivalence decision process.

D. Global Coordination

The most significant limitation of any single-jurisdiction regulatory initiative is that it creates the arbitrage incentives documented in Gap 4. A durable solution requires multilateral coordination at the level of IOSCO and the Financial Stability Board. The FSB's existing mandate to assess systemic financial risks provides a natural institutional home for RHI monitoring and for the development of minimum standards for cross-border AI trading oversight. The energy-intensity disclosure requirement has natural synergies with the FSB's climate-related financial disclosure agenda, providing an additional impetus for coordinated action.

IX. Conclusion

This paper has examined, in depth, the legal and ethical challenges that arise from the deployment of AI-driven financial prediction systems across modern markets. The central argument is that existing regulatory frameworks—while increasingly attentive to single-system manipulation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

and systemic risk—have not adequately addressed a set of structurally important threats: the emergence of autonomous collusive behaviour in multi-agent AI environments; the misrepresentation of AI capabilities in investment disclosures; the unresolved allocation of fiduciary responsibility in AI advisory relationships; the structural incentives for cross-border regulatory arbitrage; and the contradiction between the financial sector's ESG commitments and the energy footprint of its AI infrastructure.

The Tiered Regulatory Compliance Framework proposed here provides a risk-proportionate classification system that aligns regulatory burden with the degree of opacity and systemic potential of each AI trading system. The Extended Ethical Compliance Score integrates transparency, accountability, fairness, systemic risk, energy intensity, and disclosure fidelity into a single auditable indicator. The Human-in-the-Loop Accountability Matrix and the Regulatory Harmonisation Index provide operational instruments for the fiduciary duty and arbitrage gaps respectively. Evaluated against four documented market disruptions, the framework demonstrates consistent predictive and mitigative value.

The framework is proposed not as a substitute for regulatory judgment but as an evidentiary infrastructure that makes that judgment more defensible and more consistent across jurisdictions. The overarching policy goal—preserving market integrity while accommodating legitimate AI-driven innovation—is better served by a framework of this kind than by either of the dominant alternatives: reactive enforcement after the fact or prescriptive prohibition that forecloses innovation.

X. Limitations and Future Work

Several limitations of the present analysis warrant acknowledgment. The S_e formula's weight parameters are normatively determined and require empirical calibration; the sensitivity analysis reported in Section VI-C indicates robustness for extreme cases but not for borderline ones. The RHI is proposed as a conceptual instrument and has not been computed against live regulatory texts; a follow-on empirical study mapping SEBI, SEC, FCA, and ESMA requirements onto the five-dimensional vector space is needed to validate the 0.7 threshold. The HLAM's three-tier oversight structure has not been tested in a live regulatory context and may require refinement based on implementation experience.

Future research should pursue three directions. First, the development of Regulator-AI systems—AI agents designed specifically to monitor and audit other AI trading systems in real time—represents the most promising long-term approach to the speed-of-market challenge that human oversight alone cannot address. Second, the energy-intensity disclosure framework requires collaboration between financial regulators and standards bodies to develop a standardised E_c measurement protocol. Third, the multi-agent collusion risk identified in Gap 1 warrants dedicated experimental research using market microstructure simulations to characterise the conditions under which independently-operating AI agents converge on collusive equilibria in realistic trading environments.

References

[1] U.S. SEC, "Staff Report on Algorithmic Trading in U.S. Capital Markets," Washington, DC, 2020.
[2] IDC, "Worldwide Spending on Artificial Intelligence Systems: A Forecast Through 2027," International Data Corporation, 2023.
[3] IMF, "Regulatory Considerations Regarding Accelerated Use of AI in Securities Markets," Technical Notes and Manuals, vol. 2025, no. 016, Dec. 2025.
[4] J. Armour and H. Eidenmüller, "Self-Driving Corporations?" Harvard Business Law Review, vol. 10, no. 1, pp. 87–139, 2020.
[5] EU AI Act, "Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence," Official Journal of the European Union, 2024.
[6] SEC v. Athena Capital Research, LLC, Securities Exchange Act Release No. 34-73369, Oct. 2014.
[7] CFTC and SEC, "Findings Regarding the Market Events of May 6, 2010," Joint Report, Sep. 2010.
[8] McKinsey & Company, "The State of AI in 2024: Generative AI's Breakout Year," McKinsey Global Institute, May 2024.
[9] SEC Division of Examinations, "2024 Examination Priorities," Washington, DC, 2024; SEC, "2025 Examination Priorities," Washington, DC, 2025.
[10] Venable LLP, "Artificial Intelligence in Investment Management: Regulatory Challenges and Fiduciary Implications," Dec. 2025. [Online]. Available: https://www.venable.com
[11] IOSCO, "The Use of Artificial Intelligence and Machine Learning by Market Intermediaries and Asset Managers," Consultation Report CR/01/2025, Mar. 2025.
[12] European Commission, "Impact Assessment of the EU AI Act," 2024; H. Laurer, A. Renda, and K. Yeung, Compliance Cost Estimation, 2021.
[13] INTL Sustainable Development Observatory, "Sustainable Finance and AI: ESG Risk Assessment with Machine Learning," ISDO, Dec. 2025.
[14] SEC v. Visionary Private Equity Group, Press Release No. 2010-131, 2010.
[15] Knight Capital Americas LLC, SEC Administrative Proceeding No. 3-15570, Oct. 2013.
[16] SEBI, "Master Circular for Stock Brokers," Mumbai, 2023; SEBI Orders on Algo Trading Manipulation, 2022–23.
[17] E. Calvano, G. Calzolari, V. Denicolò, and S. Pastorello, "Artificial Intelligence, Algorithmic Pricing, and Collusion," American Economic Review, vol. 110, no. 10, pp. 3267–3297, Oct. 2020.
[18] Oxford Journal of Financial Regulation, "Fintech and the Future of Securities Law," vol. 9, 2023.
[19] SSRN, "Algorithmic Trading and Market Manipulation: A Legal Analysis," Working Paper, 2024.
[20] IEEE Xplore, "Deep Learning for Financial Market Prediction: Ethical Considerations," Proc. IEEE Conf. on AI and Finance, 2025.
[21] Bank for International Settlements, "The Impact of Artificial Intelligence on the Financial Sector," BIS Working Papers, 2024.
[22] Financial Stability Board, "Artificial Intelligence and Machine Learning in Financial Services," FSB Report, Nov. 2017.
[23] Frontiers in Artificial Intelligence, "Ethical Theories, Governance Models, and Strategic Frameworks for Responsible AI Adoption," vol. 8, 2025.
[24] Future Business Journal, "AI-Driven Sustainable Finance: Computational Tools, ESG Metrics, and Global Implementation," Springer Open, Aug. 2025.
[25] Sidley Austin LLP, "Artificial Intelligence in Financial Markets: Systemic Risk and Market Abuse Concerns," Client Advisory, Nov. 2025.