American Scholarly Journal for Scientific Research (ASJSR)
https://asjsr.org/ Volume 2026, Issue 1003 ISSN: 3143-2999

Efficiency and Fragility as Joint Products: AI, Governance, and Financial Markets in the US and India

Author(s): Neil Sethia & Vihaan Rustagi
Institution(s): Shiv Nadar School Gurgaon and Sanskriti School, Delhi, India
Mentor: Geeti Gupta (Ascensions Careers)

Abstract

Artificial intelligence now makes most of the decisions in financial markets. Algorithms execute the majority of trading in large equity markets; JPMorgan has estimated that discretionary human traders account for only about a tenth of US equity trading volume, with quantitative and passive strategies making up the rest (Kolanovic, 2017), and the same tools price assets, construct portfolios, score borrowers, detect fraud, and interpret textual information such as news and sentiment. This development is conventionally framed as a trade-off, in which AI makes markets faster, cheaper, and more liquid at the cost of making them less stable and less fair. This paper argues that the trade-off framing is incomplete. Efficiency and risk are not two separate effects to be set against each other. They arise from a single change: the replacement of rules written by people with rules learned by machines from data, a change that produces the gains and the dangers together. Speed, scale, and pattern recognition narrow spreads and widen credit access, and the same three properties produce flash crashes, herding, opacity, and bias in lending. Because both sides share a source, the net effect of AI on markets is not determined by the technology. It is set by the rules built around it, which makes regulation the deciding variable. The argument is developed through formal models, secondary market data, and documented cases, using the United States as the mature reference case, where the long data series and the studied failures reside, and India as the live test, a younger and far more retail-heavy market in which adoption has outrun oversight. The central finding is that whether AI helps or harms markets is a question of governance rather than of technology.

Keywords: Artificial Intelligence, Algorithmic Trading, Systemic Risk, Financial Regulation, Algorithmic Fairness, India

Introduction

Most trading in financial markets is no longer carried out by people. By one JPMorgan estimate, human discretionary traders account for only about ten percent of US equity trading volume, with the remainder driven by quantitative and algorithmic strategies (Kolanovic, 2017), much of it priced and cancelled in fractions of a second with no person approving any single order. India presents the same development on a shorter timeline. Algorithmic trading became possible there only after the Securities and Exchange Board of India opened the exchanges to Direct Market Access in 2008 (Securities and Exchange Board of India, 2008), yet automated systems already account for more than 60 percent of activity (Soman, 2026). This is no longer the practice of a few specialist funds. Since the early 2010s, AI has spread out of trade execution and into portfolio management, credit scoring, fraud detection, and forecasting (Bahoo et al., 2024), reshaping not only how trades occur but who, or what, decides to make them (MacKenzie, 2018).

The appeal of these systems is straightforward. They process more data than a person can, react faster, and cost less to operate. They have narrowed the gap between buying and selling prices, reduced transaction costs, and brought services such as portfolio advice within reach of small investors who could never have afforded a human adviser (Dou, Goldstein, & Ji, 2025). In credit, models that draw on alternative data can lend to borrowers whom manual underwriting never reached (Fuster, Plosser, Schnabl, & Vickery, 2019), an advance of particular consequence in a country such as India, where much of the population remains poorly served by formal finance. By most standard measures, markets have become more efficient, and access to them has widened.

The same properties generate difficulties. Models trained on similar data behave in similar ways, and when many of them respond to one signal at once, prices can move violently before anyone intervenes. The 2010 Flash Crash is the defining instance, when a single large sell order set off a cascade that drove the Dow down by close to a thousand points in minutes (Kirilenko, Kyle, Samadi, & Tuzun, 2017), though smaller episodes occur far more frequently than is generally recognised (Borch, 2016). A model that determines who receives a loan can reproduce the bias embedded in its training data across millions of decisions, and with a consistency no human committee could match (Bartlett, Morse, Stanton, & Wallace, 2022). Because these systems are often difficult to interpret, even for the firms that build them, the reasons for their failures, and

the question of who should answer for them, are correspondingly difficult to establish. The shift that improved markets is the same shift that rendered them fragile, opaque, and in unfair places.

This paper advances a single claim. In financial markets, the properties that make AI efficient are the same properties that make those markets more fragile and less fair. Speed, scale, and pattern recognition are not the favourable half of the account, with the risks attached separately elsewhere. They are the source of both. Efficiency and risk are joint products of one mechanism, the replacement of legible rules written by people with opaque rules learned from data, and the gains and the harms are the same development observed from two sides.

A significant consequence follows, and it extends the argument beyond finance. If the benefits and the risks share a source, the net effect of AI on markets cannot be inferred from the technology itself. A faster and more pattern-sensitive model is neither safer nor more dangerous in the abstract. The outcome is determined by the structure built around it: the disclosure rules, the audit requirements, the allocation of liability, the circuit breakers. Regulation is therefore the deciding variable rather than a secondary consideration. The same instrument can stabilise a market or break it according to how it is governed, a proposition regulators themselves have begun to act upon, from the risk tiers of the EU AI Act (Regulation (EU) 2024/1689) to the rules now being written in India (Securities and Exchange Board of India, 2025a; Reserve Bank of India, 2025).

The claim is tested across two markets. The United States supplies the mature evidence: the long data series, the established regulators, and the failures that have already been studied closely, including the Flash Crash and the work on bias in algorithmic lending (Kirilenko et al., 2017; Bartlett et al., 2022). India supplies the live test. It is a younger market, growing faster, with a far larger share of inexperienced retail money exposed to automation, and with regulators writing the rules in real time. If the mechanism is genuine, it should appear in both, and it should appear more sharply in India, where the same properties are producing the same outcomes with thinner protection around them.

The paper proceeds in order. It first establishes how finance moved from human rules to machine rules, and where AI is now deployed. It then reviews the improvements AI has delivered and the risks it has created, together with the responses of regulators in the United States, Europe, and India. It then turns to its own analysis, examining the cases that test the claim, the policy that follows from it, and the direction in which the technology is moving. The conclusion returns to

its opening proposition: whether AI makes markets better or worse is, finally, a choice about how they are governed.

Methodology

This paper is an analytical study rather than an empirical one. It does not conduct experiments or assemble a new data set. It constructs its argument by drawing together existing work and reading the whole of it through a single idea, that the gains and the risks of AI in finance arise from one mechanism. The aim is to connect findings ordinarily kept apart, not to produce a new measurement.

The argument rests on three kinds of material. The first is a set of standard models from finance, used to make each mechanism precise rather than to estimate any quantity. Price movement is framed through geometric Brownian motion, the process underlying option-pricing theory (Black & Scholes, 1973). Loss is framed through Value at Risk and its extension Conditional Value at Risk, which corrects for the insensitivity of VaR to the size of losses beyond its threshold (Rockafellar & Uryasev, 2000). Credit decisions are framed through logistic regression, the classical anchor that gradient-boosted trees and neural networks generalise at the cost of interpretability (Berg, Burg, Gombović, & Puri, 2020). These are not the paper's own models. They are the established baselines of the field, and they are introduced because they make clear what a learned model adds and what it relinquishes. The second kind of material is published quantitative evidence on market behaviour, including spreads, liquidity, trading volume, and loss figures, drawn from regulators, banks, and academic studies. The third is a set of documented cases, among them the 2010 Flash Crash (Kirilenko et al., 2017), the bias identified in algorithmic lending (Bartlett et al., 2022), and recent regulatory actions in India (Securities and Exchange Board of India, 2025a; Reserve Bank of India, 2025), used to assess whether the claim holds against real events and not only in theory. The cases are treated as tests: if the mechanism is correct, these are the outcomes it should predict.

The scope is deliberate, and its limits warrant statement at the outset. The paper examines two markets, the United States and India, and assigns them different functions. The United States serves as the reference case, possessing the deepest data, the longest record of AI in markets, and the failures that have been most closely studied. India serves as the test case, being younger, growing faster, far more driven by retail participation, and still formulating its rules, which

makes it the setting in which the mechanism should appear most clearly if it is real. Other markets, including foreign exchange, cryptocurrency, and the wider emerging world, are raised where useful but not examined in depth. The focus falls principally on equities and credit, where both the evidence and the regulatory activity are richest. The study relies on secondary sources in English and on regulatory documents that are recent and in places still in draft, so the Indian material in particular should be read as a developing picture rather than a settled one. None of this weakens the central argument, which concerns a mechanism rather than a single market, but it does establish the boundary within which the specific claims should be held.

Literature Review

The literature on artificial intelligence in finance is frequently presented as two opposing camps, one documenting efficiency and one documenting danger, but this division does not withstand close examination. The properties that every efficiency study credits for sharper price discovery and wider credit access, namely speed, scalability, pattern recognition, and the capacity to learn, are the properties that every risk study identifies as the source of instability and unfairness (Bahoo et al., 2024; Sidley Austin LLP, 2024). These are not two findings to be weighed against each other but one finding observed from two directions. The benefit and the hazard are produced by a single underlying move, the replacement of legible, human-authored rules with opaque, machine-learned ones. The sections below trace that move through the history of the technology, its applications, its measured gains, and its risks, drawing throughout on the Indian market, where the tension is sharpest because adoption has outrun oversight.

Background and History

The entry of AI into finance was incremental, and the early literature treated it purely as an instrument of efficiency. The first systems were rule-based execution algorithms that exploited latency differences between venues, understood as tools for saving time rather than as sources of risk (Hendershott & Riordan, 2013). As machine learning matured and the "Fintech" subfield emerged, AI extended into credit analysis, customer service, and compliance (Arner, Barberis, & Buckley, 2016), a widening that bibliometric work dates to a sharp rise in output from the early 2000s (Bahoo, Cucculelli, Goga, & Mondolo, 2024). The decisive shift, however, was one of character rather than scope. AI ceased to be a fast calculator and became a general-purpose

technology that restructures competition itself, and the present wave of largely autonomous, agentic systems now dominates the governance debate (Charafeddine, 2026). Sociological accounts sharpen the point most relevant to this paper: although these systems are built by people, the decisions to buy and sell are increasingly taken by the algorithms themselves and executed automatically across the infrastructure linking firms to exchanges (MacKenzie, 2018; Min & Borch, 2022). India's trajectory compresses this entire arc into a far shorter window. Algorithmic trading became possible only after the Securities and Exchange Board of India introduced Direct Market Access in 2008, yet automated systems now account for the majority of activity on Indian exchanges, with one analysis placing algorithmic trading above sixty percent of market activity. The handover of authorship that required three decades in American markets has occurred in India in roughly half that time, and with a far larger share of inexperienced retail participants exposed to it.

1. Core Applications

High-frequency and algorithmic trading remain the most studied application, with algorithms executing in fractions of a second (Brogaard, Hendershott, & Riordan, 2014) and reinforcement-learning agents managing inventory and spread in market making (Gašperov & Kostanjčar, 2021). Where classical models represent price movement as geometric Brownian motion, a continuous random walk with constant drift and volatility (Black & Scholes, 1973), machine-learning systems earn their edge precisely by trading on the deviations from that baseline, the microstructure patterns and short-lived mispricings a simple stochastic process is built to ignore. A second cluster is forecasting, in which deep learning and long short-term memory networks are applied to prediction, though the literature is candid that out-of-sample accuracy remains contested (Li & Bastos, 2020). A third is credit scoring, in which models draw on both conventional records and alternative data such as transaction histories and digital footprints to assess borrowers whom manual underwriting cannot reach (Berg, Burg, Gombović, & Puri, 2020). The interpretable baseline here is logistic regression, whose coefficients state plainly how each input moves the decision; gradient-boosted trees and neural networks generalise it to sharper predictions, but at the cost of that legibility, which is the same trade the paper tracks everywhere else. Around these sit sentiment analysis through natural language processing, fraud and anti-money-laundering detection built on anomaly methods, and robo-advisory portfolio automation (Bahoo, Cucculelli, Goga, & Mondolo, 2024). What unifies these otherwise

unrelated applications is structural, and it is central to the present argument: in each, a task once resting on explicit human reasoning is delegated to a model that infers its own rule from data, gaining reach and speed precisely as it becomes harder to interrogate. India illustrates the stakes of this delegation with unusual clarity. Its derivatives market is now the largest in the world by contract volume, and retail investors form a large share of that flow, much of it increasingly automated through broker APIs, strategy builders, and third-party algorithm vendors. The application here is neither theoretical nor confined to institutions; it is a mass-market retail phenomenon, which renders the loss of legibility a question of consumer protection and not only of systemic plumbing.

2. Efficiencies and Benefits

The most consistent benefit identified across the literature concerns the speed and completeness with which information becomes priced. The data-processing capacity of AI is associated with faster price discovery, deeper liquidity, and lower transaction costs (Dou, Goldstein, & Ji, 2025). In credit, scoring models built on alternative data are reported to advance financial inclusion by lending to borrowers whom manual underwriting overlooks (Fuster et al., 2019), and a further argument holds that automation removes idiosyncratic human bias, since a model applies identical criteria to every applicant rather than varying with the disposition of an individual loan officer (Howell et al., 2022). At the broadest level, AI is characterised as a general-purpose technology that enables faster responses to demand shocks across banking, insurance, and asset management (Bahoo et al., 2024). The inclusion argument carries particular weight in India, where a large population remains underserved by formal credit and where AI-driven lending is positioned as an instrument for extending access. Yet the structure of the literature is itself revealing, and it is the reason this review treats efficiency and risk as one mechanism rather than two: the inclusion claim and the discrimination claim rest on the same capability. The precision that allows a model to price a thin-file borrower accurately is the precision that allows it to encode the bias contained in that borrower's historical data. The same authors who present AI as a general-purpose efficiency engine catalogue its systemic risks a few pages later (Bahoo et al., 2024). This is not an inconsistency on their part. It is the honest shape of the subject.

3. Risks and Challenges

Here the central tension of the field becomes explicit, and it resolves in favour of this paper's

hypothesis: the properties that make AI efficient, namely shared optimisation logic, common training data, and high-speed autonomy, are the same properties that make markets fragile, and, along a separate axis, less fair.

On fragility, the 2010 Flash Crash remains the defining case. A single large automated sell order triggered a cascade among high-frequency traders that drove the Dow Jones Industrial Average down by close to a thousand points within minutes before it recovered almost as quickly (Kirilenko, Kyle, Samadi, & Tuzun, 2017; Min & Borch, 2022). The most durable interpretation describes the episode as a "normal accident" rooted in the tight coupling of automated markets, in which safeguards entirely rational for one firm, such as stop-loss triggers, interact across firms to produce a spiral no participant intended (Min & Borch, 2022). Drawing on documentary evidence and 189 interviews, this work demonstrates that algorithmic interaction is frequently nonlinear and unpredictable, giving markets the signature of complex systems prone to precisely this failure (Min & Borch, 2022). That such events are routine rather than exceptional is underscored by estimates that small flash crashes occur many times on an average trading day (Golub et al., 2012, in Borch, 2016). Agent-based modelling adds the important qualification that contagion speed depends on portfolio crowding and network structure, so that greater diversification does not reliably reduce systemic exposure and can worsen it (Paulin, Calinescu, & Wooldridge, 2019). The mechanism beneath these findings is convergence. Former SEC Chair Gary Gensler has argued that herding among algorithms built by similarly trained developers, an "apprentice effect," has contributed to past crashes (OMFIF, 2024), and because deep learning depends on vast datasets and heavy compute, firms tend to concentrate on a few dominant data and model providers, producing a financial-system "monoculture" in which many institutions hold correlated positions and react identically (Sidley Austin LLP, 2024). Recent theoretical work formalises this as a risk multiplier that compounds once adoption, model similarity, and prediction-outcome feedback rise together, while conceding that AI need not raise systemic risk where homogeneity remains low (Meng & Chen, 2026). A further strand documents algorithmic collusion, in which independent pricing agents learn to sustain supra-competitive outcomes with no explicit agreement, defeating an antitrust framework built on the proof of intent (Dou et al., 2025).

These dynamics also expose a weakness in the tools used to measure the risk. The standard

measure of potential loss is Value at Risk, which reports the loss that will not be exceeded at a given confidence level α:

$$P(L > VaR_\alpha) = 1 - \alpha$$

Its appeal is that it compresses risk into a single number, but that number is silent about the tail. VaR reports the threshold and says nothing about the size of the losses beyond it, which is the region where correlated, high-speed selling does its damage. Conditional Value at Risk, also called Expected Shortfall, corrects for this by measuring the average loss once the threshold is breached:

$$CVaR_\alpha = E[L \mid L > VaR_\alpha]$$

The distinction matters for the argument. The monoculture and herding described above produce fat-tailed, jointly distributed losses, the kind VaR is built to overlook and that only a tail-sensitive measure such as CVaR can capture (Rockafellar & Uryasev, 2000). The efficiency-generating mechanism does not merely create tail risk. It creates tail risk of the specific shape that the industry's default measure understates.

The Indian evidence renders these abstractions concrete and recent. More than ninety percent of retail futures-and-options traders have consistently lost money, and the regulator's own research found net losses for individual traders widening by forty-one percent to ₹1.05 lakh crore in a single financial year (Securities and Exchange Board of India, 2024), in a market where unregulated black-box algorithms promising guaranteed returns proliferated without transparency or recourse. In July 2025 the regulator barred the global trading firm Jane Street over allegedly manipulative algorithmic strategies, a vivid instance of high-speed autonomy outpacing oversight. The fairness axis is equally well evidenced. AI credit scoring reproduces discrimination embedded in historical data, and the bias is rarely a coding error: it originates in training data, is reinforced through proxy variables correlated with protected characteristics even when those characteristics are formally excluded and persists because institutions face little penalty in the absence of regulation. The empirical findings are pointed. Examining the United States mortgage market, Bartlett, Morse, Stanton, and Wallace (2022) find that algorithmic lenders do discriminate, charging minority borrowers higher rates, yet roughly forty percent less

than face-to-face lenders, a result that captures the double edge precisely: automation removes some human bias while entrenching a subtler and harder-to-detect form. Opacity compounds the problem, since the disparate impact of a neural network is far more difficult to identify and contest than that of older scoring methods. India's regulator corroborates the danger from the supply side: a Reserve Bank survey found that among entities using AI, only about a third validated for bias and fairness, only a tenth maintained bias-mitigation protocols, and fewer still kept audit logs, even as the bias is amplified across high-volume transactions.

A newer hazard has arrived with the generative and agentic systems now entering finance, and it expresses the legibility problem in its most acute form. Large language models are increasingly deployed for research, compliance, customer service, and the drafting of financial analysis, and they are prone to hallucination, the confident production of fluent output that is fabricated or false (Remolina, 2024). The difficulty is not simply that such systems err, since every model errs, but that the error cannot be detected from the output itself, which carries the same authority whether it is accurate or invented. A classical trading model that misprices an asset can be checked against the market, and a credit model's decision can in principle be audited against its inputs, but a generative model that fabricates a figure, a citation, or a regulatory provision presents the fabrication in the same register as the truth. This is opacity of a deeper kind than the black box of a scoring model, because frequently there is no stable rule to inspect at all, only a probabilistic output that shifts from one query to the next. The hazard compounds as these systems acquire autonomy: a fabrication a human reviewer would catch may propagate unchecked through an agentic pipeline that acts on its own outputs, and large US banks have already begun extending their model-risk frameworks specifically to test for hallucination before deployment (Remolina, 2024). India's regulator has registered the concern directly, observing that the non-deterministic outcomes of such systems make both the detection of error and the allocation of liability substantially harder (Reserve Bank of India, 2025). The same generative capability that allows a model to read a thousand filings and summarise them in seconds is the capability that allows it to invent the contents of a filing it never read. Power and unreliability are, once again, the same property.

4. Regulatory and Ethical Context

Regulation has advanced along two tracks that are only now converging. After the Flash Crash, United States regulators strengthened circuit breakers and introduced Regulation SCI in 2014,

while the European Union adopted the Markets in Financial Instruments Directive II (MiFID II), which took effect in 2018, though both were criticised for leaving firms substantial compliance latitude (Min & Borch, 2022). The EU AI Act (Regulation 2024/1689) subsequently introduced a risk-based regime classifying credit scoring, underwriting, and anti-money-laundering systems as "high risk," with obligations for human oversight and incident reporting (Regulation (EU) 2024/1689); a 2025 European Banking Authority mapping found it broadly compatible with banking rules but flagged an unresolved tension between stability-oriented prudential regulation and rights-oriented AI law (European Banking Authority, 2025). The Indian response is where the paper's thesis becomes visible in the regulation itself. SEBI's February 2025 framework on retail algorithmic trading classifies algorithms into "white box" systems, whose logic is disclosed and replicable, and "black box" systems, whose logic is proprietary and non-replicable, and imposes substantially heavier requirements on the latter, including registration of providers as research analysts, audit trails, unique order tagging, and mandatory kill switches. This is the legibility-for-power trade written directly into law: the regulator has effectively conceded that opacity is the variable that must be controlled, which is precisely the claim this paper advances. Critics nonetheless argue that the regime still falls short of MiFID II and IOSCO standards and shifts risk onto digitally unsophisticated retail investors (Soman, 2026). On the credit side, the Reserve Bank's FREE-AI Committee report of August 2025 set out seven principles, among them fairness, accountability, and explainability, and named the same hazards the academic literature does, including model convergence undermining market diversity, herding, collusion, and the difficulty of allocating liability (Reserve Bank of India, 2025). The conclusion recurring across this entire body of work, American, European, and Indian alike, is that stability, efficiency, and fairness cannot be regulated separately, because a single set of technological properties drives all three at once. This is the gap the paper addresses: the literature has established the mechanism in mature markets but has not yet tested it in a young, retail-dominated market where, as India demonstrates, the same properties are producing the same outcomes faster and with less protection.

Results and Analysis

The preceding review established the state of existing knowledge. This section advances the paper's own argument. If the hypothesis is correct, and efficiency and risk are genuinely joint

products of one mechanism, two conditions should hold in the record. The trait that delivers the gain should be the trait that produces the harm, observable within the same event. And the outcome, whether the harm was contained, should turn on the rules around the technology rather than on the technology itself. The cases below are selected to test exactly these conditions. They span the three faces of the risk, fragility, autonomy, and fairness, and they extend from the mature US market to the live Indian one.

Case Studies

The 2010 Flash Crash is the clearest test of fragility. On 6 May 2010 a single large automated sell order set off a chain reaction among high-frequency traders, and the Dow Jones Industrial Average fell by close to a thousand points in minutes before recovering almost as quickly (Kirilenko, Kyle, Samadi, & Tuzun, 2017). The significance lies in the ordinariness of the failure: nothing exotic broke. The execution speed that narrows spreads on a normal day is what allowed the cascade to move faster than any human could follow, and the risk controls that protect a single firm, the stop-loss and the risk limit, are what transmitted the selling from one algorithm to the next. Min and Borch (2022) describe this as a "normal accident," a failure originating not in a faulty component but in the tight coupling of the system as a whole. Efficiency and fragility are not two features but one design, read on a calm day and on a turbulent one. What altered the outcome afterwards was regulatory: the introduction of market-wide circuit breakers and single-stock limit-up limit-down rules, which made the algorithms no slower but placed around them a structure capable of arresting a spiral. The technology remained constant. The governance is what changed.

Knight Capital, two years later, tests the same proposition from the side of autonomy. On 1 August 2012 the firm deployed new code to its trading systems, but the update reached only seven of eight servers, and the eighth revived a dormant routine unused since 2003 (U.S. Securities and Exchange Commission, 2013). Within approximately forty-five minutes the system sent more than four million orders into the market, traded close to 397 million shares across 154 stocks, and accumulated billions in unwanted positions, producing a loss of roughly $440 million, exceeding the firm's annual earnings (U.S. Securities and Exchange Commission, 2013). The automation that enabled Knight to operate as one of the largest market makers in the

country, executing thousands of orders a second across dozens of venues, is what rendered the failure impossible to halt by hand once it had begun. Speed was both the product and the disaster. Here too the regulatory response is decisive for the argument. The SEC brought its first action under the Market Access Rule, Rule 15c3-5, finding that Knight maintained no adequate pre-trade controls or deployment procedures, and the firm paid a penalty and was shortly absorbed by a competitor (U.S. Securities and Exchange Commission, 2013). The conclusion the regulator drew was not that automated market-making is too dangerous to permit, but that the controls around it were absent. The deciding variable was governance, consistent with the hypothesis.

The Apple Card episode tests the fairness axis, and its value lies in its refusal of the easy answer. In November 2019 the entrepreneur David Heinemeier Hansson reported that his Apple Card credit limit was twenty times his wife's, although she held the higher credit score and they filed jointly, and Steve Wozniak described the same pattern (New York State Department of Financial Services [NYDFS], 2021). Hansson's objection was that the bank had delegated the decision to a black box that no one could explain. The New York Department of Financial Services investigated, reviewed underwriting data for roughly 400,000 applicants in the state, and in March 2021 found no fair-lending violation, concluding that men and women with similar credit characteristics generally received similar outcomes (NYDFS, 2021). This complicates any lazy reading of the thesis, as it should. The harm in this case was not proven statistical bias. The harm was that the decisions could not be explained, by the customer, by the public, or initially by the bank, which is what permitted the perception of bias to take hold and persist for more than a year. The regulator's own conclusion established the structural point: it found the decisions lawful but held that credit scoring and the anti-discrimination laws surrounding it require modernisation for an age of algorithmic lending (NYDFS, 2021). Opacity is a harm even where bias is absent, because a decision no one can account for can be neither trusted nor contested. This is the legibility problem in its purest form.

India is the live test, and the mechanism appears there faster and with less protection around it. Automated systems already constitute more than sixty percent of activity, in a market opened to direct algorithmic access only in 2008 (Soman, 2026), and the country's derivatives market is now the largest in the world by contract volume, with retail investors forming a large share of the

flow. That retail base is where the exposure concentrates. The regulator's own research found that more than ninety percent of individual futures-and-options traders lose money, with net losses widening by forty-one percent to ₹1.05 lakh crore in a single financial year (Securities and Exchange Board of India, 2024). These aggregate losses are not attributable to algorithmic trading alone, since heavy retail losses in derivatives would occur with or without automation; they describe the scale and the vulnerability of the population now exposed to it. The AI-specific harm is sharper and more direct. Unregulated black-box services promising fixed returns proliferated among these traders with no transparency and no recourse on failure, and in July 2025 SEBI barred the global firm Jane Street over allegedly manipulative algorithmic strategies, a clear instance of high-speed autonomy outpacing the oversight meant to catch it (Securities and Exchange Board of India, 2025b). On the credit side, the Reserve Bank's survey found that among firms using AI, only about a third tested for bias and fairness and only a tenth operated any bias-mitigation protocol (Reserve Bank of India, 2025). The opacity and autonomy that drive the US cases are present here as well. What differs is the thinness of the structure around them and the size of the retail base left unprotected, which on the hypothesis is what determines how much harm passes through.

Read together, the four cases sustain the claim. In each, the efficiency-producing trait and the harm-producing trait are the same trait, not adjacent ones. And in each, the outcome turned on the rules: circuit breakers after 2010, the Market Access Rule after Knight, the regulatory gap exposed by Apple Card, and the thin protection surrounding the Indian retail market. The technology did not decide. The governance did.

Policy Recommendations

If the harms arise from the same source as the gains, policy cannot seek to remove the harmful traits without destroying the useful ones. It must instead build structure around them, and each measure below addresses a specific failure demonstrated in the cases.

The first is explainability and audit. The Apple Card episode shows that opacity is itself a harm, so deployed models that make consequential decisions, above all in credit, should carry a requirement that their decisions be explicable and their data and outcomes auditable for disparate impact. SEBI has taken the sharpest version of this step, sorting algorithms into "white box"

systems whose logic is disclosed and reproducible and "black box" systems whose logic is not and imposing heavier requirements on the latter (Securities and Exchange Board of India, 2025a). That distinction is the paper's thesis written into law, and it is the correct instinct.

The second is pre-trade controls and kill switches. Knight Capital failed because nothing could halt a runaway system from within once it had begun, and the SEC's Market Access Rule supplied the remedy (U.S. Securities and Exchange Commission, 2013). Every firm with automated market access should be required to maintain tested pre-trade limits and a functioning kill switch, a requirement SEBI's 2025 framework now imposes on the Indian retail market.

The third is structural protection against herding and monoculture. Because deep-learning systems converge on a few data and model providers, producing correlated positions across the market (Sidley Austin LLP, 2024), regulators should monitor concentration among model and data suppliers as they would in any other systemically important function, and calibrate circuit breakers to correlated selling rather than to single-stock movements alone.

The fourth is enforceable fairness standards. The evidence that bias persists through proxy variables even when protected characteristics are excluded (Bartlett, Morse, Stanton, & Wallace, 2022), together with the Reserve Bank's finding that few Indian firms test for it at all (Reserve Bank of India, 2025), establishes that bias testing cannot remain voluntary. It should be a condition of deployment, with the burden placed on the lender to demonstrate that the model has been examined.

The fifth is clear liability. When an autonomous system causes loss, responsibility must be assignable, and the Reserve Bank's FREE-AI report rightly proposes a graded liability framework that holds the deploying institution accountable regardless of the model's degree of autonomy (Reserve Bank of India, 2025). The governing principle should be that delegation to a model never transfers responsibility away from the firm.

A single point underlies all five. Stability, efficiency, and fairness cannot be regulated in separate rooms, because one set of technical properties drives all three at once. The Indian regulators have grasped this in principle, in SEBI's framework and the Reserve Bank's seven principles, but principles are not yet binding rules, and the country's pro-innovation stance stands in real tension with the protection its retail market requires. Closing that gap, from principle to enforceable rule, is the central policy task.

Future Outlook

The trade between power and legibility is not settling but deepening. The newest systems, generative and agentic models, are moving from reading filings and drafting research toward acting with limited human supervision, growing more capable and less interpretable at each step (Charafeddine, 2026). Nothing in the recent direction of technology reverses the core problem this paper describes; it sharpens it. A market governed by models that can be inspected line by line presents a different governance problem from one governed by models that learn their own rules and, increasingly, take their own actions.

Two paths are open, and the difference between them is the whole of this paper's argument. On the first, capability continues to outrun oversight, opacity is accepted as the price of performance, and autonomy spreads without the structure to contain it, which is the Knight Capital morning repeated at the scale of an entire market. On the second, the instruments of legibility keep pace, through explainability requirements, model audits, regulatory sandboxes that test systems before they reach the public, and graded liability that holds a human institution answerable. India stands between these paths. Its India AI Mission and its regulators' embrace of innovation sandboxes pull toward rapid adoption, while the FREE-AI principles and the 2025 trading framework pull toward control (Reserve Bank of India, 2025; Securities and Exchange Board of India, 2025a). Which prevails is not a question the technology answers. It is the governance choice toward which the whole paper has pointed, and it is being made now, in real time, in precisely the young and fast-moving market this study set out to observe.

Conclusion

This paper began with a market in which most trading is no longer done by people, and it set out to challenge the way that shift is usually understood. The conventional account treats AI's efficiency and its dangers as a trade-off, two separate effects to be balanced against each other. The argument advanced here is that they are not separate at all. They are joint products of a single mechanism, the replacement of legible rules written by people with opaque rules learned from data. Speed, scale, and pattern recognition deliver tighter spreads, faster price discovery, and wider credit access, and the same three properties produce flash crashes, herding, opacity, bias, and now the fluent fabrications of generative systems. The gain and the harm are one development observed from two sides.

The analysis bears this out. Across the cases, the trait responsible for the efficiency was the trait responsible for the failure, visible within the same event. The execution speed that narrows spreads is what drove the 2010 cascade. The automation that made Knight Capital a leading market maker is what made its collapse impossible to halt by hand. The precision that lets a credit model price an overlooked borrower is the precision that lets it encode that borrower's inherited disadvantage, and the generative capability that summarises a thousand filings is the capability that invents the contents of one. In each instance the harm was not an accident attached to the technology from outside. It followed from the design.

From this the paper's central consequence follows. If efficiency and risk share a source, the net effect of AI cannot be read from technology, because technology is the same on a calm day and a catastrophic one. What separates the two is the structure built around it: the disclosure rules, the pre-trade controls, the audit requirements, the lines of liability. Regulation is therefore the deciding variable, and the outcome is a policy choice rather than a technical destiny. The comparison between the United States and India sharpens the point. The mechanism operates identically in both, but it produces greater harm in India, not because the technology differs but because the protection around it is thinner and the exposure, concentrated in an enormous and inexperienced retail base, is greater. The variable that changed was governance, exactly as the hypothesis predicts.

Several questions remain open, and they mark the ground for further work. Whether a regulatory distinction as clean as SEBI's white-box and black-box categories can survive models that are neither fully disclosed nor fully proprietary is uncertain, and the rise of generative and agentic systems strains it further. The mechanism this paper traces would also benefit from direct empirical measurement in the Indian market rather than inference from mature ones. And the deepest problem, whether the instruments of legibility can be made to keep pace with systems that grow more capable and less interpretable together, is not one that any single jurisdiction can answer alone. What the paper establishes is narrower but firm. The same machinery drives the promise and the peril, and which prevails is being decided now, by the rules societies choose to place around it.

References

Arner, D. W., Barberis, J., & Buckley, R. P. (2016). The evolution of Fintech: A new post-crisis paradigm? *Georgetown Journal of International Law, 47*(4), 1271–1319.

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in finance: A comprehensive review through bibliometric and content analysis. *SN Business & Economics, 4*(2), Article 23. https://doi.org/10.1007/s43546-023-00618-x

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2022). Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics, 143*(1), 30–56.

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of FinTechs: Credit scoring using digital footprints. *The Review of Financial Studies, 33*(7), 2845–2897.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637–654.

Borch, C. (2016). High-frequency trading, algorithmic finance and the flash crash: Reflections on eventalization. *Economy and Society, 45*(3–4), 350–378.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies, 27*(8), 2267–2306.

Charafeddine, J. (2026). Agentic AI: A comprehensive survey of architectures, applications, and future directions. *Artificial Intelligence Review, 59*(1), Article 11. https://doi.org/10.1007/s10462-025-11422-4

Cheng, E. (2017, June 13). *Just 10% of trading is regular stock picking, JPMorgan estimates.* CNBC. https://www.cnbc.com/2017/06/13/death-of-the-human-investor-just-10-percent-of-trading-is-regular-stock-picking-jpmorgan-estimates.html

Dou, W. W., Goldstein, I., & Ji, Y. (2025). *AI-powered trading, algorithmic collusion, and price efficiency* (NBER Working Paper No. 34054). National Bureau of Economic Research. https://doi.org/10.3386/w34054

Fuster, A., Plosser, M., Schnabl, P., & Vickery, J. (2019). The role of technology in mortgage lending. *The Review of Financial Studies, 32*(5), 1854–1899.

Gašperov, B., & Kostanjčar, Z. (2021). Market making with signals through deep reinforcement learning. *IEEE Access, 9*, 61611–61622. https://doi.org/10.1109/ACCESS.2021.3074782

Hendershott, T., & Riordan, R. (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis, 48*(4), 1001–1024.

Howell, S. T., Kuchler, T., Snitkof, D., Stroebel, J., & Wong, J. (2024). Lender automation and racial disparities in credit access. *The Journal of Finance, 79*(2), 1259–1311.

Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance, 72*(3), 967–998.

Li, A. W., & Bastos, G. S. (2020). Stock market forecasting using deep learning and technical analysis: A systematic review. *IEEE Access, 8*, 185232–185242. https://doi.org/10.1109/ACCESS.2020.3030226

MacKenzie, D. (2018). Material signals: A historical sociology of high-frequency trading. *American Journal of Sociology, 123*(6), 1635–1683.

Meng, S., & Chen, X. (2026). *Artificial intelligence and systemic risk: A unified model of performative prediction, algorithmic herding, and cognitive dependency in financial markets* (arXiv:2604.03272). arXiv. https://arxiv.org/abs/2604.03272

Min, B. H., & Borch, C. (2022). Systemic failures and organizational risk management in algorithmic trading: Normal accidents and high reliability in financial markets. *Social Studies of Science, 52*(2), 277–302.
New York State Department of Financial Services. (2021). *Report on the Apple Card investigation.* https://www.dfs.ny.gov/
Official Monetary and Financial Institutions Forum. (2024, January 29). *Risks around AI and algorithmic convergence are causing "regulatory gaps."* OMFIF. https://www.omfif.org/2024/01/risks-around-ai-and-algorithmic-convergence-are-causing-regulatory-gaps/
Paulin, J., Calinescu, A., & Wooldridge, M. (2019). Understanding flash crash contagion and systemic risk: A micro–macro agent-based approach. *Journal of Economic Dynamics and Control, 100*, 200–229.
Remolina, N. (2024). *Generative AI in finance: Risks and potential solutions* (SMU Centre for AI & Data Governance Research Paper). Singapore Management University.
Reserve Bank of India. (2025). *Report of the committee on the framework for responsible and ethical enablement of artificial intelligence (FREE-AI) in the financial sector.* https://www.rbi.org.in/
Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk, 2*(3), 21–41.
Securities and Exchange Board of India. (2008). *Direct market access* (Circular MRD/DoP/SE/Cir-7/2008). https://www.sebi.gov.in/
Securities and Exchange Board of India. (2024). *Study on the analysis of profit and loss of individual traders dealing in the equity F&O segment.* https://www.sebi.gov.in/
Securities and Exchange Board of India. (2025a, February 4). *Safer participation of retail investors in algorithmic trading* (Circular SEBI/HO/MIRSD/MIRSD-PoD/P/CIR/2025/0000013). https://www.sebi.gov.in/
Securities and Exchange Board of India. (2025b, July 3). *Interim order in the matter of Jane Street.* https://www.sebi.gov.in/
Sidley Austin LLP. (2024, December). *Artificial intelligence in financial markets: Systemic risk and market abuse concerns.* Sidley Austin. https://www.sidley.com/en/insights/newsupdates/2024/12/artificial-intelligence-in-financial-markets-systemic-risk-and-market-abuse-concerns
Soman, A. M. (2026). SEBI's 2025 framework for safer retail participation in algorithmic trading: An evaluation of investor protections, regulatory gaps, and international comparisons. *Indian Journal of Law and Legal Research.* https://www.ijllr.com/
U.S. Securities and Exchange Commission. (2013). *In the matter of Knight Capital Americas LLC* (Exchange Act Release No. 70694; Admin. Proc. File No. 3-15570). https://www.sec.gov/
Legislation and regulatory instruments
Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments (MiFID II). (2014). *Official Journal of the European Union,* L 173/349.
Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). *Official Journal of the European Union,* L 2024/1689.
U.S. Securities and Exchange Commission. (2014). *Regulation Systems Compliance and Integrity* (Release No. 34-73639).