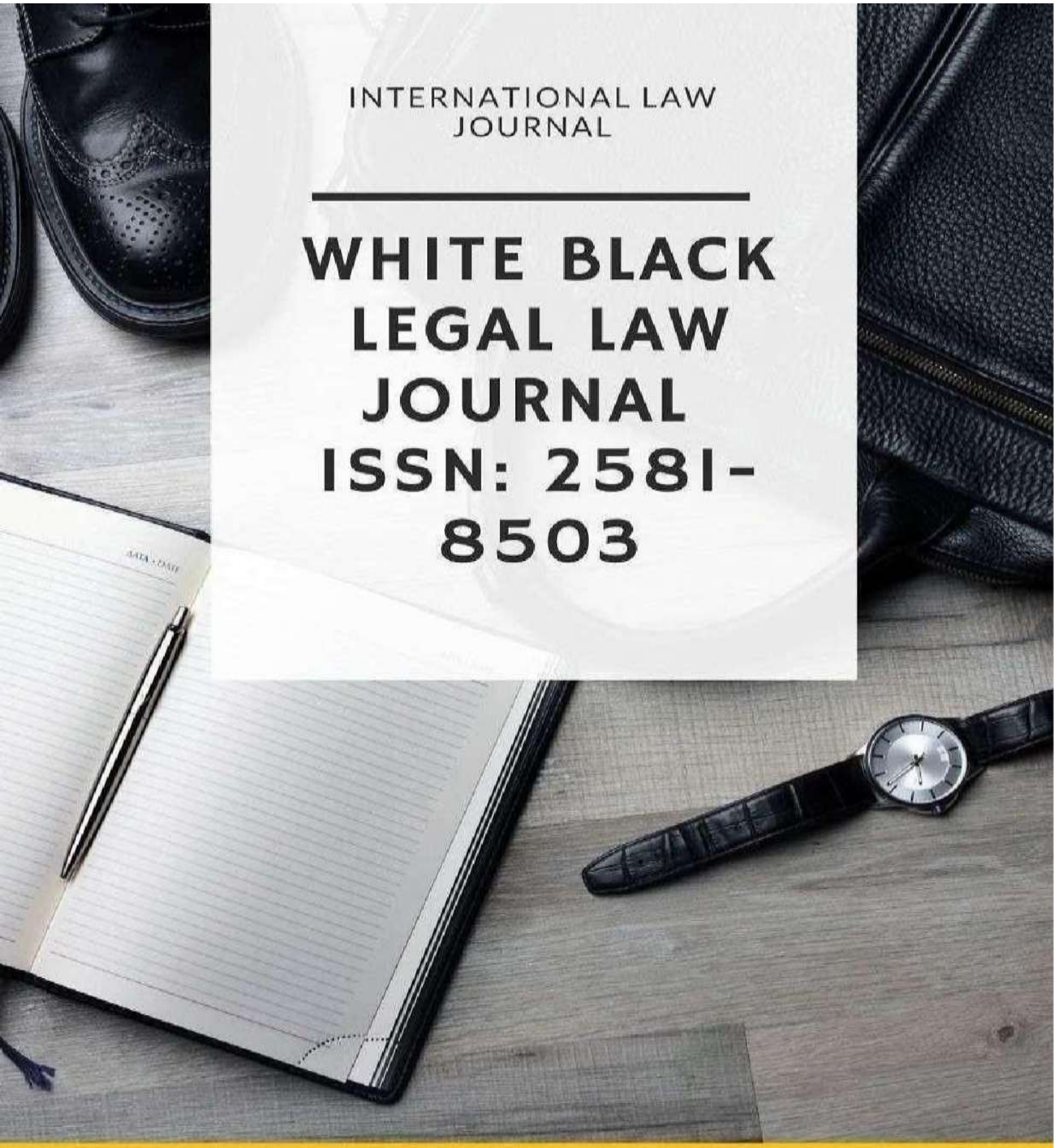

INTERNATIONAL LAW JOURNAL

WHITE BLACK LEGAL LAW JOURNAL

ISSN: 2581-8503

Peer - Reviewed & Refereed Journal

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

WWW.WHITEBLACKLEGAL.CO.IN

DISCLAIMER

The views, opinions, interpretations, and conclusions expressed in the articles are exclusively those of the respective authors. They do not represent or reflect the views of the Editorial Board, Editors, Reviewers, Advisors, Publisher, or Management of *White Black Legal*.

While reasonable efforts are made to ensure academic quality and accuracy through editorial and peer-review processes, *White Black Legal* makes no representations or warranties, express or implied, regarding the completeness, accuracy, reliability, or suitability of the content published. The journal shall not be liable for any errors, omissions, inaccuracies, or consequences arising from the use, interpretation, or reliance upon the information contained in this publication.

The content published in this journal is intended solely for academic and informational purposes and shall not be construed as legal advice, professional advice, or legal opinion. *White Black Legal* expressly disclaims all liability for any loss, damage, claim, or legal consequence arising directly or indirectly from the use of any material published herein.

ABOUT WHITE BLACK LEGAL

White Black Legal – The Law Journal is an open-access, peer-reviewed, and refereed legal journal established to provide a scholarly platform for the examination and discussion of contemporary legal issues. The journal is dedicated to encouraging rigorous legal research, critical analysis, and informed academic discourse across diverse fields of law.

The journal invites contributions from law students, researchers, academicians, legal practitioners, and policy scholars. By facilitating engagement between emerging scholars and experienced legal professionals, *White Black Legal* seeks to bridge theoretical legal research with practical, institutional, and societal perspectives.

In a rapidly evolving social, economic, and technological environment, the journal endeavours to examine the changing role of law and its impact on governance, justice systems, and society. *White Black Legal* remains committed to academic integrity, ethical research practices, and the dissemination of accessible legal scholarship to a global readership.

AIM & SCOPE

The aim of *White Black Legal – The Law Journal* is to promote excellence in legal research and to provide a credible academic forum for the analysis, discussion, and advancement of contemporary legal issues. The journal encourages original, analytical, and well-researched contributions that add substantive value to legal scholarship.

The journal publishes scholarly works examining doctrinal, theoretical, empirical, and interdisciplinary perspectives of law. Submissions are welcomed from academicians, legal professionals, researchers, scholars, and students who demonstrate intellectual rigour, analytical clarity, and relevance to current legal and policy developments.

The scope of the journal includes, but is not limited to:

- Constitutional and Administrative Law
- Criminal Law and Criminal Justice
- Corporate, Commercial, and Business Laws
- Intellectual Property and Technology Law
- International Law and Human Rights
- Environmental and Sustainable Development Law
- Cyber Law, Artificial Intelligence, and Emerging Technologies
- Family Law, Labour Law, and Social Justice Studies

The journal accepts original research articles, case comments, legislative and policy analyses, book reviews, and interdisciplinary studies addressing legal issues at national and international levels. All submissions are subject to a rigorous double-blind peer-review process to ensure academic quality, originality, and relevance.

Through its publications, *White Black Legal – The Law Journal* seeks to foster critical legal thinking and contribute to the development of law as an instrument of justice, governance, and social progress, while expressly disclaiming responsibility for the application or misuse of published content.

THE "BLACK BOX" ACCOUNTABILITY GAP: DOES MERELY REGISTERING AS A REGISTERED ADVISER SOLVE THE EXPLAINABILITY PROBLEM IN AI-DRIVEN FLASH CRASHES?

AUTHORED BY - HRISHITA PAL & SAMPOORNA ROY
Program: BBA & Bcom LL. B
AMITY LAW SCHOOL KOLKATA

Abstract

Financial regulators increasingly require firms deploying autonomous AI systems to register as investment advisers (RAs) or equivalent entities, under the assumption that registration imposes sufficient accountability and oversight. Yet when an AI-driven "flash crash" occurs and the provider claims the model is too complex to explain, registration alone fails to close the "black box" accountability gap. This paper argues that RA-style registration does not, by itself, resolve the explainability problem; it formalizes responsibility without materially improving transparency, interpretability, or investor protection. The paper proposes a layered regulatory architecture that couples registration with explainability standards, technical safeguards, and recovery-oriented liability rules to meaningfully protect investors in the era of autonomous financial AI.

Keywords: artificial intelligence, flash crash, explainable AI, black box, financial regulation, registered adviser, investor protection.

I. Introduction

Autonomous AI systems now execute millions of trading decisions daily, often with minimal human oversight. When these systems malfunction at scale—such as in a "flash crash" style event—losses can be systemic yet attribution becomes opaque. Despite growing regulatory attention to "black box" AI, many jurisdictions still lean on classic prudential tools like registration as a registered adviser (RA) or equivalent gatekeeping license as the primary accountability mechanism. [2][8]

The central question is whether this registration-based regime is sufficient to protect investors when an autonomous AI system causes a flash-style crash and the provider asserts that the model's internal logic is, in practice, inexplicable. This paper contends that merely registering an AI-driven firm as an RA does not solve the explainability problem and therefore leaves a critical accountability gap. [4][8]

II. The "Black Box" Accountability Gap

1. What is the "black box" problem?

In finance, the "black box" refers to AI or algorithmic systems whose internal decision-making processes are either intentionally opaque (proprietary) or cognitively opaque (too complex for humans to follow). [2][4] Deep-learning-based trading agents, reinforcement learning "bots," and large-scale AI-driven execution platforms often fall into this category.

Opacity creates an accountability gap because:

- It is difficult for regulators, auditors, or investors to reconstruct *why* a specific loss occurred.
- Firms can credibly claim that some decisions are "emergent" or "non-deterministic," complicating fault-based liability. [6][8]

2. Flash crashes as a case study

The 2010 flash crash, caused by interacting high-frequency trading algorithms, illustrated how speed and scale can amplify small errors into trillion-dollar temporary wealth destruction. [3][9] Modern AI-driven systems add another layer: decisions may stem not from hand-coded rules but from learned patterns in high-dimensional, dynamic data.

When an AI-enabled system participates in or triggers a flash-style event, the damage can be:

- Rapidly disseminated across multiple venues and instruments. [5]
- Mediated by opaque "feedback loops" between agents, liquidity providers, and market-wide risk controls. [3][5]

Yet the system's complexity may prevent even the provider from offering a clear,

human-comprehensible explanation of the causal chain—only high-level statistics or post-event diagnostics. This is precisely the "black box" accountability gap: financial harm occurs, but explanation and responsibility remain blurred. [4][8]

III. The Research Gap: Registration as Registered Adviser

1. The regulatory intuition behind registration

Most regulatory regimes presume that registration as an RA or equivalent entity creates accountability through:

- Auditable governance structures (board oversight, compliance, risk functions).
- Disclosure obligations (fees, conflicts, material risks).
- Fiduciary or supervisory duties enforced by regulators. [2][8]

In the AI context, registration is often framed as a "gateway" that ensures only "responsible" entities can deploy complex systems. [8] However, this framework assumes that human-managed oversight is sufficient even when the underlying model is effectively opaque.

2. Why registration alone does not solve explainability

Registration does four things well:

- It identifies the responsible legal person.
- It subjects the firm to prudential and conduct-of-business rules.
- It requires certain disclosures about the *existence* and *broad design* of AI usage.
- It creates a supervisory relationship with regulators. [2][8]

But registration does *not* inherently:

- Require that AI decisions be interpretable or explainable in human-understandable terms.
- Mandate that post-event diagnostics must reconstruct *why* a flash crash occurred at the transactional level.
- Ensure that investors can challenge decisions or obtain redress when the system "acts alone." [6][8]

Thus, a firm can register as an RA, run a fully opaque AI trading engine, suffer or contribute to a flash-style crash, and then legitimately claim that the model's causal structure is too complex to unpack. The *formal* accountability line runs to the registered firm, but the *substantive* explainability line remains absent. This is the research gap: we lack robust empirical and

doctrinal analysis of whether registration-based regimes actually constrain opaque AI-driven harm, or merely paper-over it. [4][8]

IV. The Flash Crash Scenario: AI, Registration, and Liability

1. A stylized case

Imagine a hedge fund that:

- Registers as an RA and discloses that it uses an autonomous AI system for execution and risk-management.
- Trains the system on multi-venue data, reinforcement-learning-style objectives, and live market-making signals.
- After a liquidity shock, the system unleashes a cascade of market-sweeping orders, contributing to a five-minute flash-style crash.

Post-crash, the provider states that:

- The model's internal state is high-dimensional and stochastic.
- No single variable or "rule" can be isolated as the primary cause.
- Explanations are limited to statistical diagnostics (e.g., feature-importance scores, anomaly-detection alerts). [1][6]

Under current regimes, regulators may compel the firm to:

- Halt the system,
- Submit a post-mortem report, and
- Pay fines for breaches of conduct rules. [2][8]

However, without a clear causal narration, investors' ability to:

- Understand their specific losses,
- Attribute fault in litigation, or
- Demand meaningful system redesign is materially constrained. Here registration delivers *formal* accountability but not *substantive* explainability.

2. The "cannot explain" defense

The "we cannot explain due to complexity" defense is plausible in deep-learning and reinforcement-learning systems. Such models often:

- Learn non-linear, non-additive interactions across thousands of features.
- Evolve behavior over time via online learning or feedback loops. [1][6]

Policy-level discussions of explainability increasingly recognize that:

- There is no universal, technically feasible definition of an "adequate" explanation across all AI systems. [8]
- Many regulations still adopt "coarse" requirements for "transparency" or "meaningful explanations" without specifying granularity, timing, or audience-specific explanations. [8][10]

Thus, when a flash-style event occurs, the provider can plausibly argue that any "explanation" is partial, approximate, or post-hoc, and that full interpretability is either technically impossible or commercially damaging. In that context, RA-style registration becomes a liability parking-space rather than a problem-solving mechanism. [4][8]

V. Does Registration Protect Investors? A Critical Assessment

1. Where registration helps

Registration provides several investor-protection benefits:

- Gatekeeping: Only firms with minimum capital, governance, and compliance capacity may operate. [2]
- Oversight: Regulators can inspect, request documentation, and impose sanctions.
- Redress channels: Investors can complain to regulators or sue the registered entity, even if they do not fully understand the model. [2][8]

In a flash-crash context, this formal structure can:

- Trigger incident-reporting and emergency measures.
- Support ex-post disciplinary action against the firm. [3][5]

2. Where registration falls short

Despite these advantages, registration fails on three key fronts:

First, transparency is mandatory. Registration regimes rarely require that AI models be interpretable or that decisions be explainable in human-intelligible terms. [8][10] Regulators may demand logs and high-level descriptions, but not causal reconstructions.

Second, diagnostic tools are insufficient for accountability. Post-crash diagnostics (e.g., anomaly-detection dashboards or feature-importance plots) may help engineers tune the model

but do not, by themselves, translate into clear fault-lines for investors or courts. [6][8]

Third, the burden of proof remains on the harmed investor. When an RA-registered firm deploys an opaque AI system and a flash crash occurs, investors must still prove:

- Breach of duty,
- Causation, and
- Damages.

Opacity makes proving causation especially hard, because the link between the AI's internal states and the harmful outcome is not reconstructable in a court-friendly form. [6][8]

Thus, while registration identifies *who* is liable, it does not clarify *why* liability arises or *how* the AI operated. This is insufficient to close the "black box" accountability gap. [4][8]

VI. Toward a Layered Regulatory Architecture

Registration should be one layer of a broader regulatory architecture, not the sole accountability mechanism. The following components could meaningfully enhance investor protection without stifling innovation.

1. Explainability standards tailored to AI use

Rather than treating "explainability" as a generic buzzword, regulators should define:

- Levels of explainability by risk tier (e.g., core execution vs. ancillary analytics).
- Target audiences (regulators, internal risk teams, retail investors) and corresponding explanation formats. [8][10]

For high-impact AI systems contributing to systemic risk (including flash-crash-capable agents), regulators could:

- Require "counterfactual" or "scenario-based" explanations (e.g., "if liquidity had been 10% higher, the system would have reduced order size by X").
- Mandate pre-approval of explanation methodologies as part of system design. [2][8]

2. Technical safeguards and "circuit breakers"

Registration-style regimes should be supplemented with technical requirements:

- Algorithmic circuit breakers that throttle, isolate, or pause AI agents when predefined volatility

or liquidity thresholds are breached. [3][5]

- Red-team and stress-testing regimes that simulate AI-driven crashes and verify that the system can be halted or overridden. [1][5]

These safeguards would not make AI fully explainable, but they would reduce the probability and severity of flash-style events, thereby indirectly protecting investors. [3][5]

3. Recovery-oriented liability and insurance

Given that full explainability may be technically infeasible, liability regimes should shift toward:

- Strict or risk-based liability for AI-driven flash-crash events, conditioned on the degree of risk posed by the system.
- Mandatory insurance or capital buffers for firms deploying high-velocity, high-impact AI trading systems. [6][8]

This approach focuses less on reconstructing every decision and more on:

- Ensuring that firms internalize the systemic costs of opaque automation.
- Providing investors with compensation even when the why of the crash remains partially opaque. [6][8]

4. Public-interest auditing and model registries Regulators could require:

- Model registries that catalog AI systems used by registered firms, including high-level descriptions, risk classifications, and update logs. [2][8]
- Public-interest audits conducted by independent experts, with the power to reconstruct a sample of flash-crash-relevant episodes without full proprietary disclosure. [8][10]

Such registries would help regulators and researchers assess whether registration-coupled oversight is actually containing AI-related risks, rather than merely formalizing an opaque status quo. [8]

VII. Policy Implications

Several implications follow for policymakers and standard-setting bodies:

1. Registration is necessary but not sufficient. RA-style regimes should be treated as a baseline layer of accountability, not a panacea for AI-driven opacity. [2][8]
2. Explainability must be operationalized. Regulators should move beyond vague "transparency" requirements and define granularity, format, and audience-specific explanations for different AI use cases. [8][10]
3. Systemic-risk-aware design is critical. AI systems capable of triggering or amplifying flash-style events must be built with circuit breakers, kill-switches, and human-override capabilities. [3][5]

4. Investor-centric redress mechanisms are essential.Where causation is partially opaque, compensation schemes and insurance-style mechanisms can protect investors even when full explanation is unattainable. [6][8]

VIII. Conclusion

The "black box" accountability gap in AI-driven finance reflects a mismatch between traditional regulatory tools—such as registration as a registered adviser—and the structural opacity of modern autonomous systems. Registration identifies the responsible party but does not ensure that investors can understand, challenge, or recover from an AI-driven flash-crash-style event when the provider claims the model is too complex to explain. [4][8]

To close this gap, regulators must supplement registration with explainability standards, technical safeguards, and recovery-oriented liability rules. Only then can AI-driven finance evolve toward a regime where accountability is not merely formal but substantively protective of investors in the face of opaque, high-speed automation.

Citations:

[1] Detecting flash crash events using deep reinforcement https://developers.lseg.com/en/article-catalog/article/modelling-and-evaluation-deep-reinforcement-learning-flash-crashes

[2] The AI black box problem in finance https://www.trustpath.ai/blog/how-financial-organizations-can-ensure-ai-explainability-and-transparency

[3] Algorithmic Circuit Breakers: Preventing "Flash Crashes ... https://www.arionresearch.com/blog/algorithmic-circuit-breakers-preventing-flash-crashes-of-logic-in-autonomous-workflows

[4] 1 https://www.finance-watch.org/wp-content/uploads/2025/03/Artificial_intelligence_in_finance_report_final.pdf

[5] AI Agents and Market Volatility: Flash Crash Risks Unpacked https://www.aicerts.ai/news/ai-agents-and-market-volatility-flash-crash-risks-unpacked/

[6] Expert Analysis https://www.arias-us.org/wp-content/uploads/2019/04/Tobey-Explainability-Where-AI-and-Liability-Meet.pdf

[7] When Algorithms Act Alone (Flash Crash & AI Incidents) https://www.youtube.com/shorts/P7PAQ237L88

[8] Explainability in AI Policies: A Critical Review of Communications, Reports, Regulations, and Standards in the EU, US, and UK https://dl.acm.org/doi/fullHtml/10.1145/3593013.3594074

[9] 2010 Flash Crash https://en.wikipedia.org/wiki/2010_flash_crash

[10] Decoding Black Box AI: The Global Push for Explainability and ... https://trendsresearch.org/insight/decoding-black-box-ai-the-global-push-for-explainability-and-transparency/