Rebutting the 'Opacity Defence' in Algorithmic Trading: The Case for Explainable and Suitable AI in Robo-Advisories

– *Hrudya Ravi*

ABSTRACT

This essay criticises the approach to algorithmic opacity in India's securities markets. It argues that SEBI's current white-box-black-box differentiation creates a false dichotomy, as such static disclosures are insufficient to address dynamic model drifts and real-time operational transparency. The regulatory and fiduciary blind spots can be bridged by mandating counterfactual explanations, algorithmic nutrition labels, and utilising adversarial red teaming within regulatory sandboxes to ensure accountability.

INTRODUCTION

The past decade has seen a burgeoning usage of algorithms in legal and financial decision-making, for various purposes ranging from assessing the criminality of individuals to making decisions on their behalf.[1] For instance, one case that garnered criticism was when the State of Wisconsin used an algorithm named 'COMPAS' to generate high-risk scores to assess an individual's recidivism (likelihood of re-offending).[2] Such usage resulted in one individual being sentenced to six years in prison, wholly based on the score. There was no rationale behind the decision, as the inventors of COMPAS refused to disclose the formula or training data behind it. This rendered it a '*black box'* which was insulated from scrutiny regarding data validity, input bias, or improper heuristics.[3] Similarly, and perhaps on a much wider scale, financial corporations regularly engage algorithmic systems in complex investment platforms (*robo-advisories*) where the management hides behind a shield of opacity and corporate secrecy.[4] This presents a systemic risk to market integrity because, as of FY'25, algorithmic trading accounted for nearly 70% of all equity derivative volumes in

[1] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION, 5-8 & 101 (Harvard University Press, 2015).
[2] State v. Loomis, 881 N.W.2d 749 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017).
[3] Julia Angwin et al., Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks, ProPublica (May 23, 2016, 3:00 PM), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
[4] Gregory Scorpino, Do Automated Trading Systems Dream of Manipulating the Price of Futures Contracts? Policing Markets for Improper Trading Practices by Algorithmic Robots, 67 FLA. L. REV. 221, 222 (2015).

India.[5] Although Deep Neural Network (DNN) models work on multi-layered parameters, there cannot be blanket acceptance of their status as *black box* models for three reasons: firstly, opacity is a direct mask of algorithmic bias; secondly, it could be a tactic to evade detection and prevention of market manipulation; and lastly, due to the unpredictability of such models, it could lead to systemic risks like flash crashes. This is substantiated by the fact that in FY'25, individual retail trader losses widened by 41% to INR 1.05 lakh crores, partially blamed on unregulated 'black box' platforms.[6] While the SEBI has formulated new regulations addressing opacity, it continues to struggle to ensure that AI models make interpretable decisions.[7] This essay aims to explore the explainability and suitability gap created by the dichotomy of a white box-black box differentiation and the loopholes posed by recent regulations in the Indian Securities market. It also aims to suggest certain policy reforms such as enhancing algorithmic accountability.

CURRENT REGULATORY FRAMEWORK AND GAPS

The SEBI has categorised algorithms into *white box* (hereinafter referred to as the 'WB') algorithms, where the logic is both disclosed and replicable, and *black box* (hereinafter referred to as the 'BB') algorithms, where the logic is neither understandable nor replicable.[8] While WB algorithms are blanketly accepted, in the case of BB algorithms, the algo provider would have to register as a research analyst and maintain a report for each algorithm. In the case of any change to its logic, they would have to register it as a new algorithm. These developments face regulatory lag amidst rapid technological evolution. One research report made by such a 'research analyst' cannot be expected to capture the dynamic drifts of an AI model which relies on Machine Learning ('*ML*') to update its logic consistently.[9] There is a big gap between regulatory disclosure and operational transparency; a situation in which regulators, algo providers, and users are given information on why AI made a specific trade, even an erroneous one, in real-time. It is redundant for the SEBI to focus on *ex-ante* disclosure for models that evolve daily (in fact, their entire purpose is to evolve daily).

[5] Ashish Rukhaiyar, Algos are now the dominating force in Indian equities, MoneyControl (June 2, 2025, 4:03 PM) https://www.moneycontrol.com/news/business/markets/algos-are-now-the-dominating-force-in-indian-equities-f-o-share-in-derivatives-nearly-70-13086705.html

[6] Times of India, Jane Street 'market manipulation' impact? (Jul 8, 2025, 3:36 PM) https://timesofindia.indiatimes.com/business/india-business/jane-street-market-manipulation-impact-indian-retail-traders-suffered-heavy-losses-rs-1-05-lakh-crore-wiped-out-in-derivatives-trading-in-fy25/articleshow/122317829.cms.

[7] Securities and Exchange Board of India ('*SEBI*'), Safer participation of retail investors in Algorithmic Trading, SEBI/HO/MIRSD/MIRSD-PoD/CIR/2025/0000013 (Issued on Feb 4, 2025).

[8] *Id.,* at 5.

[9] Surya Patchipala, Tackling Data and Model Drift in AI, 10 INT. J. SCI. R. ARCHIVE 1198-1209 (2023).

Further, as per the regulations,[10] if a trading frequency stays below the threshold of 10 orders per second ('OPS'), such trader would be considered a regular API[11] user and not an algo trader. This would exempt them from mandatory registration. Again, this creates a grey zone where sophisticated retail traders could run automated strategies which lack oversight as they do not meet the official OPS trigger. This may create a systemic risk where sub-threshold BB bots react similarly to market signals using the same technical indicators, and herd together to cause a flash crash. As such bots would not be subject to the SEBI's algo-tracking IDs, such market instability would not be traceable.

Lastly, despite WB-BB differentiation,[12] the suitability requirement is not specifically regulated. That is, there is no measure to check whether the algorithm's logic itself is unbiased or whether it nudges users towards high-churn products to the benefit of the platform. There is a clear conflict of interest here as the IA Regulations[13] require acting in the best interests of the client, but the Research Analyst Regulations[14] (intended for BB models) do not impose an equitable fiduciary standard.

UNDERSTANDING THE DICHOTOMY BETWEEN DISCLOSURE AND EXPLAINABILITY

The SEBI defines WB models as automated strategies of trading which execute orders based on 'transparent' algorithms.[15] Transparency, in this context, refers to algorithms where the logic and rules for decision-making are both accessible and understandable to users.[16] Such models are whitelisted for usage, without any further stipulations on how exactly such models are to be 'accessible' and 'understanding'. On the other hand, BB models are also permissible, provided that the algo provider registers as a research analyst ('RA') and maintains a report on the algorithm's functioning, and registers all changes to it.

Such differentiation between WB and BB models is a superficial solution, considering how static disclosure fails to address model drifts in DNN or satiate consumer curiosity about investment portfolios. A strategy may be disclosed to regulators, but still remain unexplainable in real-time execution. This leads to a gap where regulators possess the code but do not practically understand the AI's evolving logic, especially during situations of

[10] National Stock Exchange ('NSE'), Implementation Standards for Safer participation of retail investors in Algorithmic Trading, NSE/INVG/67858 (Issued on May 5, 2025).
[11] *Application Programming Interface*
[12] *Supra* note 8.
[13] SEBI (Investment Advisers) Regulations, 2013, Third Schedule, Code of Conduct, § 2.
[14] *Supra* notes 7 & 10.
[15] *Supra* note 7, at p. 5.
[16] *Id.*

market volatility. For instance, when a WB provider discloses a strategy, they might provide a snapshot of the code (or the factors weighing its decision) at T_0. However, in DNNs, the weights and biases of the model naturally shift autonomously as it processes diverse market data. By T_1, the 'disclosed' logic may no longer represent the actual logic in operation. By allowing BB providers to simply file a research report, there is a false equivalence of prediction and explanation. This means that a report stating that an algorithm uses momentum indicators, reports what the goal is, but fails to explain *why* exactly the model chose to liquidate a position during a specific liquidity crunch. The opacity defence has not been dismantled enough for regulators to challenge such non-reporting. This forms a veneer of transparency where the regulator has numerous reports but no actual visibility into model drift, a phenomenon where the AI's performance degrades or its logic evolves further away from its original training parameters. The classic case of Knight Capital comes into mind wherein the company lost a behemothic amount of $440 million in a mere 45 minutes because an old, undisclosed code remained in their system and interacted unexpectedly with new software.[17] In this case, even if the company had filed a research report with the SEC, that report would not have actually captured the emergent behaviour of the code interaction. This is proof enough that disclosure of intent is useless without the disclosure of system state (i.e., real-time logic). As highlighted by the Securities Appellate Tribunal in the case of Kohlhoff v. SEBI, technical glitches do not absolve a broker of their fiduciary duty to the client.[18]

Algorithmic opacity should legally be treated as a continuous technical glitch. If an investment adviser cannot explain the real-time logic of a trade made through AI or ML, they must be in a state of constructive non-compliance with the fiduciary duties enshrined under the Regulations.[19]

EFFICACY OF REGULATORY THRESHOLDS: QUANTITATIVE LIMITS V. QUALITATIVE RISKS

The SEBI circular regulating algorithmic trading and the broker's industry standards forum has specified that to constitute an algo, such trading platform must cross the threshold of 10 OPS.[20] However, it is not only possible but very common for algos to operate at less than 10 OPS and the majority of retail-level automation falls into such 'low-frequency' category.[21]

[17] In re Knight Cap. Gateway LLC, Exchange Act Release No. 70,694, 2013 WL 5630403 (Oct. 16, 2013).
[18] Kohlhoff v. Sec. & Exch. Bd. of India, Appeal No. 256 of 2021, [2023] SAT (India).
[19] SEBI (Investment Advisers) Regulations, 2013, Third Schedule, Code of Conduct, § 2.
[20] *Supra* note 14.
[21] Talal Al-Sulaiman, Review of Recent Research Directions and Practical Implementation of Low-Frequency Algorithmic Trading, 2(1) AMERICAN J. FIN. TECH. INNOV. (2024) https://doi.org/10.54536/ajfti.v2i1.2354.

For instance, there are trend-following, portfolio-rebalancing, arbitrage algos, etc. which only execute orders with a frequency ranging from days to months (let alone seconds).[22] There are also certain execution algos like VWAP and TWAP which take a large order and slice it to be executed over 6 hours to avoid moving the market price.[23] These may only fire one order every minute. Mathematically speaking, if one algo executes 100 OPS, the SEBI would take note of it immediately. However, if 50,000 retail traders make use of the same BB bot or different bots running on the same technology which fires 0.1 orders per second, the result is 5,000 OPS hitting the market at once. If considered cumulatively, such a small shoal of bots stands to be equally potent but remain legally invisible to algorithmic surveillance. For instance, in 2010, a lone trader (Navinder Sarao) in London used a simple spoofing algorithm which triggered a $1 trillion flash crash in the USA as his singular algo was qualitatively manipulative.[24] This is why the European Securities and Markets Authority ('ESMA') uses a much lower threshold of 2 messages per second to define an algo.[25] Even in India, after 16 years, a modern Navinder could single-handedly run a manipulative spoofing bot executing 9 OPS and cause systemic instability. Even a kill switch introduced strategically by the SEBI, would fail to deter such market volatility.[26]

A practical solution to this would be categorising algorithms by checking how correlated orders are, rather than checking if they were placed together or within 10 seconds. Using cluster tracking, SEBI could use real-time data to see if thousands of independent API keys are firing identical orders at the same millisecond. If such a cluster is detected, the 10 OPS exemption could be potentially revoked, and the group could be treated as a single algorithmic entity.

BRIDGING THE FIDUCIARY VACUUM OF ALGORITHMIC SUITABILITY AND THE IA-RA MISMATCH

The same fiduciary duties are levied on robo-advisers as are generally expected from investment advisers. However, opaque algorithms cannot be compared with human fund managers. It may be validly pointed out that human traders cannot explain the trillions of neural firings which lead them to predict (or not predict) the coming of a market crisis. The

[22] Marko Kolanovic & Zhen Wei, Momentum Strategies across Asset Classes, J.P. Morgan (2015) https://www.cmegroup.com/content/dam/cmegroup/education/files/jpm-momentum-strategies-2015-04-15-1681565.pdf.

[23] Motilal Oswal, VWAP and TWAP: Choosing the Right Trading Strategy (July 30, 2025) https://www.motilaloswal.com/learning-centre/2025/7/vwap-vs-twap-key-differences-in-trading-strategies.

[24] United States v. Sarao, No. 15-CR-00075, 2016 WL 6832442 (N.D. Ill. Nov. 9, 2016).

[25] ESMA, Supervisory Briefing on Algorithmic Trading in the EU, ESMA74-1505669079-10311 (Feb. 26, 2026).

[26] *Supra* note 7, at p. 4. ("*kill switch is an emergency function and the last level of defence…*")

issue then arises whether an algorithm can be held to a higher standard of "explainability" than a human brain. Would it be a double standard if a human trader losing INR 1.05 lakh crores is termed 'market risk', and a similar AI effect is called 'algorithmic failure'? To answer this, a false equivalence in fiduciary law is to compare a neural black box of a human trader with a digital black box of a DNN. Human cognitive processes are biologically opaque but are bound by a traceable logic of professional experience and standardised risk-metrics. In contrast, an unauditable algorithm lacks such a standard and creates a liability vacuum for what would otherwise be termed professional negligence for human investment advisers.[27]

There is no concept of algorithmic accountability in India, and thereby, no binding requirement for explainable AI models or algorithmic audits. While SEBI requires risk profiling and assessment, enforcement still focuses on whether the particular recommendations were appropriate, not the process of how the algorithm arrived at it for the best interests of the client. India's tech-neutral approach lacks the kind of algorithmic specificity found in other jurisdictions. The USA, for instance, treats fiduciary duties as the highest priority and in one case,[28] the SEC fined a robo-adviser for failing to maintain proper compliance for its algorithm. The official stance was that algorithms cannot be used as any sort of shield against fiduciary failures. The EU, through the MiFID II[29] and AI Act,[30] mandates an algorithmic suitability audit in which advisors must prove that the objective function of the AI was aligned with the risk tolerance of the client. This can be contrasted with the situation in India, where only the output of the trade may be audited, but not the logic of the recommendation itself. However, SEBI has made headways by acknowledging that digital platforms must not hide behind 'execution-only' status if using automated tools to influence user choice.[31] This establishes, to some extent, that the medium itself does not exempt the advice from IA Regulations.

An improved solution would be to move away from auditing the RAs, and instead auditing the objective function of the code. By mandating explainable suitability protocols for any platform managing retail capital, robo-advisories would have to provide a suitability audit trail indicating which client parameters triggered which investment weightings. Further,

[27] SEBI (Investment Advisers) Regulations, 2013, Third Schedule, Code of Conduct, § 2.
[28] *In re* Wahed Invest LLC, Investment Advisers Act Release No. 5959, 2022 WL 414781 (Feb. 10, 2022).
[29] Directive 2014/65/EU, of the European Parliament and of the Council of 15 May 2014 on Markets in Financial Instruments and Amending Directive 2002/92/EC and Directive 2011/61/EU, 2014 O.J. (L 173) 349.
[30] Regulation 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 O.J. (L 1689) 1.
[31] Paytm Money Ltd., SEBI Informal Guidance, SEBI/HO/IMD/DF1/OW/P/2021/8743/1 (Apr. 9, 2021) (India).

adopting a stricter standard for algorithmic fiduciary duties would better regulate the IA-RA gap.

SUGGESTIONS AND SOLUTIONS TO REGULATORY CHALLENGES

These multifaceted concerns can be resolved through three major steps. Firstly, counterfactual explanations must be mandated in periodic algorithm-audits.[32] For every major recommendation, the AI must be able to answer what change in the input data would have resulted in a different recommendation. This creates a localised explanation that would enable laypersons to understand algorithmic decisions. Secondly, borrowing from the EU's AI Act,[33] every registered algo could implement a standardised 'nutrition' label quantifying model uncertainty and data freshness. Providers must display a score of live sensitivity which discloses how much the algo relies on volatile variables like social media sentiment and hard-earnings data. A third initiative would be to strengthen the SEBI's regulatory sandbox by employing its own adversarial algorithms (commonly called red teaming) to stress-test BB models in a simulated environment before allowing them to manage retail capital. If an algo exhibits biased or high-churn behaviour, it may be denied an Algo-ID regardless of its OPS.

Another element to be considered here is that if explainability is mandated, it might have a negative impact on the intellectual property of certain hedge funds or fintech startups and their proprietary algorithms. If a developer is compelled to disclose their trade secrets, they would likely take their capital to more opaque-tolerant markets like Dubai or Singapore. This phenomenon is commonly called regulatory arbitrage,[34] and can be tackled robustly if securities organisations collaborate with regulators such as the IFSCA to standardise foundational algorithmic rules and prevent domestic capital flight.

CONCLUSION

The exemptions here could act as regulatory blind spots, ultimately to the detriment of retail investors and broader market stability. It is of utmost importance that the SEBI move towards mandating cluster-based surveillance and Real-time Explainability (XAI) protocols for any model which manages public capital. Administrative ease cannot sacrifice market integrity, and the law must absolutely ensure that if an algorithm is sophisticated enough to trade, it is transparent enough to be judged.

[32] Sandra Wachter et al., Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR, 31 HARV. J. L. & TECH. 841 (2018)/

[33] Regulation 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 O.J. (L 1689) 1.

[34] Guido Perboli et al., Navigating the AI regulatory landscape, 13 TAYLOR & FRANCIS J. 367-397 (Dec. 23, 2025).