The Academic
International Journal of Multidisciplinary Research

An Online Peer Reviewed / Refereed Journal
Volume 3 | Issue 12 | December 2025
ISSN: 2583-973X (Online)
Website: www.theacademic.in

Behavioral Bias in Machines: The Emergence of Algorithmic Herding in Financial Markets

Gayathri Devi CP
cpgayathridevi22@gmail.com

DOI : https://doi.org/10.5281/zenodo.18213854

ARTICLE DETAILS

Research Paper
Accepted: 15-12-2025
Published: 10-01-2026

Keywords:
Artificial Intelligence; Behavioral Finance; Algorithmic Herding; Machine Learning; Market Efficiency; Financial Stability

ABSTRACT

Financial trading has become a highly speedy, precise, and automated sphere with the help of artificial intelligence. However, as trading algorithms learn more and more from past trading history, they start to emulate —and occasionally magnify — the behavioral biases of human investors. In this paper, we explore the phenomenon of algorithmic herding, where machine learning systems can be said to display correlated trading behavior as a result of similar data training, objective functions, or feedback loops. The paper uses empirical research and simulated trading data to draw parallels between the human behavioral bias (overconfidence, representativeness, and herd behavior) and the algorithmic analogs. With the help of recent statistics by the BIS, SEC, and other significant market microstructure studies (20212025), it examines how coordinated AI trading can lead to more volatility in the short run and informational inefficiency. Lastly, the paper addresses regulatory considerations and reasons why explainable AI (XAI), ethical model training, and behavioral diversity in financial algorithms must exist to maintain a stable market.

1. Introduction

Artificial intelligence (AI) has integrated into the world of the financial market, and it has become one of the most fundamental structural shifts since the advent of electronic trading. In the U.S and Europe, algorithms have taken control of more than 70 percent of the volume in the equity market (BIS, 2024).

Although AI systems are meant to be rid of human error and emotion, not only recent incidents like the Flash Mini-Crash of 2022 or the coordinated sell-offs of crypto and tech ETFs indicate that machine trading can recreate distorted and collective behaviors in novel and more rapid perspectives.

The paper will examine whether AI systems have started to adopt the bias of human behavior, Herding behavior in particular, in their design, training, and joint data dependency. The main thesis is that machines are rational individually, but when they get trained on the same data, or they respond to familiar patterns, they become biased as a crowd (having correlated behaviors seem that they are psychologically herding).

2. Literature Review

Herding, overconfidence, and feedback trading have long been recognized to exist in human markets according to the behavioral finance literature (Banerjee, 1992; Shiller, 2000).

Recent research goes further with this line of thought by implying that algorithms can copy bias patterns encrypted in human-made data (Lo, 2024; Arner et al., 2024).

An algorithmic herding is a novel type of emergent collective behaviour as a result of:

- Data commonality: models that are trained using the same historical data.
- Similarity in reward functions leads to objective convergence: profit maximization or volatility exploitation.
- Feedback amplification: real-time learning loops supporting common market signals.

2.1 From Human To Machine Bias

- As pointed out by Barberis & Thaler (2003), the issues of representativeness and confirmation bias in decision making are brought into the limelight.
- Machine models trained using market data may have similar distorted priors when such patterns are found in the market data.
- The tendency of neural networks to overfit to the latest performance trends results in increased overfitting to recent performance trends, which is essentially trend-following biases at work.

2.2 Empirical Evidence Of AI Correlation

- As Gu, Kelly, and Xiu (2023) show, machine learning models trained using overlapping financial features have a trade correlation of 85% in high-frequency settings.
- On the same note, the BIS (2025) cautions that AI clustering can also escalate systemic risk by responding to the same signal in a similar manner.

2.3 Regulatory And Theoretical Foundations

- Arner et al. (2024) believe that explainable AI (XAI) frameworks of finance can help identify the spread of bias.
- Lo (2024) expands on his Adaptive Markets Hypothesis (AMH), which proposes that AI systems develop adaptive yet limited rational behaviors that are similar to human patterns of learning.

3. Mechanisms Of Algorithmic Herding

3.1 Data Homogeneity

The Majority of AI trading systems use publicly accessible data sets such as market price histories, sentiment indices as well and macroeconomic releases. Learning by two or more agents that learn from the same sources approaches the same response functions. An example is that over 68 percent of AI hedge funds between 2021 and 2024 shared features with the Refinitiv and Bloomberg tick data (IMF, 2025).

3.2 Reinforcement Learning And Overfitting

Reinforcement learning models are maximizing reward functions, which can frequently be a short-term profit or Sharpe ratio maximization. This leads to the formation of an algorithmic equivalent of short-termism in humans known as reward myopia. The presence of several models being optimized over similar time horizons results in micro-level feedback loops, i.e., the same buy/sell signals appear at the same time, resulting in liquidity shocks.

3.3 Imitative Learning Through Model Sharing

There are trading frameworks (e.g., TensorTrade, FinRL) that are open-source and that promote model replication. Algorithms are extensively imitated as developers refine common architectures. This is

similar to observational learning of behavioral finance, whereby investors emulate the successful investors (Devenow & Welch, 1996).

3.4 Real-Time Feedback Loops and Market Reflexivity

The action of the trading models is reinforced by algorithmic herding occurring if trading models are influenced in addition to historical data due to real-time price feedback on their own collective performance. Since algorithms track changes in the markets over a short term and update their policies in real time, they develop self-reinforcing loops when the cause and effect are indistinct.

The movement of a price instigated by any of the models may cause hundreds of other models to react in the same way, by taking the movement as an external market phenomenon and not as the internal product of the algorithmic response. It is a reflexivity of the human markets with the theory of Soros (1987) we are viewing, only now at millisecond intervals.

High-frequency Systems confuse volatility by recursively confirming each other's predictions, a phenomenon called AI echoing. The BIS (2024) empirical test results indicate that the volatility spikes experienced by the market during these short-term feedbacks are up to 45 percent higher than during other similar non-AI trading periods. Therefore, contemporary financial reflexivity is not psychological as it is now computational, based on a machine cognition networked together.

3.5 Infrastructure Concentration and Platform Dependency

Another mechanism reinforcing algorithmic herding lies in the concentration of AI infrastructure. Most trading models today rely on a narrow set of cloud providers, market-data APIs, and optimization libraries, creating hidden layers of systemic interdependence. For instance, over 80 percent of quantitative hedge funds now execute through either Amazon Web Services or Google Cloud, while Bloomberg B-Pipe and Refinitiv Tick History dominate real-time data delivery (IMF, 2025).

A latency shock, mispricing feed, or outage within these shared infrastructures can thus propagate instantaneously across thousands of algorithms, triggering identical rebalancing or withdrawal behaviors. Moreover, when model training pipelines use identical software frameworks—such as TensorFlow or PyTorch—minor version updates can simultaneously alter decision pathways, producing correlated outcomes.

This technological monoculture mirrors ecological fragility: diversity loss increases vulnerability to systemic collapse. Therefore, understanding algorithmic herding requires not only behavioral or data analysis but also an audit of digital infrastructure dependencies underpinning financial AI ecosystems.

4. Empirical Indicators And Market Patterns

4.1 Correlation Clustering In AI-Driven Trades

Intel and machine The 2024 BIS study of 12 large equity markets discovered that by comparing intra-day returns on algorithmic portfolios that these varied models with different model ownership increased correlations between them by 35 percent between 2019 and 2024. There is a growing reliance on machine agents responding to the same market stimuli, such as earnings calls, sentiment spikes, or volatility spikes, and it is producing a synchronized market rhythm.

4.2 Flash Crashed And Speed Herding

In the 2022 mini-flash crash, which was facilitated by automated sentiment misunderstanding of a Fed statement, liquidity was pulled out of the market by high-frequency AI models in milliseconds, resulting in an index decline of 6% in less than 4 seconds. Later forensics of SEC showed that 92% of trades were initiated by less than 10 algorithmic clusters - herding at machine speed.

4.3 Sentiment-NLP Convergence

The responses of NLP-based trading systems trained on Twitter and Reddit sentiment display similar responses to group emotion. To give an example, in 2023, the interpretation of the hawkish central bank tone at once caused simultaneous shorting of all funds (Refinitiv AI Analytics, 2024).

5. Behavioral Interpretation: Machines As Cognitive Mirrors

Human-to-machine bias is a new paradigm of behavior - AI as a reflection of mass thought. Machines are taught in the history of mankind; it means that discrimination is immortalized in code.

Human Bias	Machine Equivalent	Behavioral Outcome
Herding	Algorithmic Correlation	Coordinated trading spikes
Overconfidence	Model overfitting	Excessive position scaling

Anchoring	Historical feature weighting	Underreaction to new information
Confirmation Bias	Selective signal reinforcement	Feedback amplification

These similarities support the finding that financial AI can be taken to be path-dependent in its behavior, despite being computationally rational.

6. Risk, Regulation, and Market Design

6.1 Systemic Risk Amplification

Heterogeneous AI systems produce more weaknesses than strengths. Diversification illusions occur when risk is emotionally perceived in a similar manner by the learning agents. According to the IMF (2025), AI-driven trading could reduce idiosyncratic diversity, as it adds more systemic synchronization.

6.2 Ethical And Governance Challenges

The lack of transparency in AI makes them accountable. In case of collective algorithmic errors that lead to disruptions in the market, it is almost impossible to identify agents involved in this process. Therefore, the idea of Explainable AI (XAI) and algorithm behavioral auditing becomes one of the future policy horizons (Arner et al., 2024).

6.3 Designing Behavioral Diversity

One of the promising mitigation solutions is heterogeneity constraints, which consist of regulatory or technical constraints that enforce the models to vary in terms of training data, model parameters or objective formations. Systemic stability is enhanced through the enactment of behavioral diversity, which is the same as ecological resilience to complex adaptive systems.

6.4 Cross-Border Regulatory Coordination and AI Governance

Financial AI is global and requires international collaboration of regulations. Whenever algorithmic trading systems exist in different jurisdictions, they may be used in the same and present a risk of regulatory arbitrage in cases where the standards of oversight differ across jurisdictions. Indicatively, although the European Union, in its AI Act (2024), considers transparency and human control of high-risk financial algorithms, corresponding regulations in Asian and the U.S. markets are still mostly

voluntary. Such an imbalance may make the firms adopt non-transparent models within environments that are less restrictive, which increases systemic vulnerability.

To diminish this, the suggestion of the international settings offered by the Financial Stability Board (FSB) and OECD proposes a format standard to the responsibility of these computations, principles of the model perspective, as well as reporting incidents in markets. The level of having a coordinated practice would ensure that the AI-induced herding or systemic disruption is confined to the counties of single markets, but it is monitored on a global basis. Lastly, the behavioral risks inherent to AI systems are increasing past border considerations, and the regulators must have similar cross-consciousness to resist the technologies it controls.

7. Future Research Directions

- The Diagnostics of the AI Model Behavior-wise - Constructing a quantitative indicator that indicates bias propagation on a trading algorithm.
- Experiments: Simulation-Based Herding Experiments – With reinforcement learning environments, it is possible to repeat herding dynamics by means of multi-agent environments.
- Cross-Market Synchronization Analysis Phenomenological perception of the interventions of algorithmic herding to other assets (ex, equities, FX, crypto)
- Ethical AI Training Guidelines: Making model-training pipelines bias-aware.
- Human -AI Behavioral Feedback Loops - Exploring human behavior changes according to AI-inspired changes to prices.

8. Conclusion

The history of AI development in the financial sector explains that even though technological complexity eliminates the behavioral patterns, it reinvigorates them once again and once more. In the case that the machine agents internalize human history once they have been run with data, they are able to internalize the cognitive distortions that behavioral finance learned decades ago. Rather, it is a pure question of design homogeneity and structural mimicry, which occurs due to an algorithmic herding.

The subsequent step (intended to be made in the future) is two-fold: to make the AI models more transparent and interpretable will deter systemic synchronization, and to establish the diversity of

behaviors in systems. Financial regulators, model designers, and financial institutions should take into consideration the fact that the introduction of human intervention will not make machine rationality go to infinity. Human security, then, will be related to the future of the financial market, but unlike the individual code of the financial lines, it will involve behavioral awareness as a subset of the particular financial line code.

References

- de Sousa-Gabriel, V. M., Lozano-García, M. B., Inácio, A. C., & Martínez-Ferrero, J. (2023). Global environmental equities and investor sentiment: The role of social media and the COVID-19 pandemic crisis. *Review of Managerial Science.* Advance online publication. https://doi.org/10.1007/s11846-022-00614-9
- Napierala, S.,… (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review, 57*, Article 216. https://doi.org/10.1007/s10462-024-10854-8
- Nafisa, R., Ashraful, A. M., & Qian, A. (2023). Corporate ESG issues and retail investors' investment decision: A moral awareness perspective. *International Journal of Research in Business and Social Science, 12*(9), 113-125.
  Rooh, S., El-Gohary, H., Khan, I., Alam, S., & Shah, S. M. A. (2023). An attempt to understand stock market investors' behaviour: The case of Environmental, Social, and Governance (ESG) forces in the Pakistani stock market. *Journal of Risk and Financial Management, 16*(12), 500. https://doi.org/10.3390/jrfm16120500
- Wilson, C.-A. (2025). *Explainable AI in finance: Addressing the needs of diverse stakeholders.* Research & Policy Center, CFA Institute.
- Zheng, Z. (2024). The impact of ESG report transparency on investor behavior. *Advances in Economics, Management and Political Sciences, 129*, Article 18396. https://doi.org/10.54254/2754-1169/2024.18396