Vol. 19 (2026): Proceedings of 2026 International Conference on Financial Technology, Mathematical Statistics, and Business Analytics (ICFMB 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/icfmb.202619516522

AI-Driven Detection, Prediction and Reduction of Investor Behavioural Biases in Financial Markets

Peishan Zhao*

The University of Auckland Business School, Auckland, New Zealand

**Corresponding author: Peishan Zhao.*

Abstract

Investor behavioural biases play an important role in financial markets, but their identification and prediction are difficult. While traditional methods including surveys, experiments and econometric proxies provide valuable insights, they are limited in terms of timeliness, scalability and predictive power. This paper reviews recent research on the use of artificial intelligence (AI) methods (machine learning, deep learning and reinforcement learning) in identifying investor behavioural biases. It examines how AI approaches are applied to identify and forecast overconfidence, loss aversion, anchoring, herding, the disposition effect, momentum and overreaction or underreaction in markets. The review concludes that AI methods are a better foundation for real-time and predictive behavioural analyses than conventional methods because they can detect nonlinear signals from trading records, financial text, social media and portfolio information. Random Forests, LSTMs, CNNs, graph-based learning and reinforcement learning are particularly helpful when investor behaviour is dynamic, diverse and hard to measure. At the same time, the review shows that AI does not solve behavioural-finance issues by itself: explainability, generalisation, privacy and the abuse of predicting investor biases are major challenges. So future research should integrate multiple data sources with explainable AI and additional ethical protections to ensure that AI-based tools help investors avoid their biases rather than predictably exploit them.

Keywords

artificial intelligence, behavioural finance, investor bias, machine learning, explainable AI

1. Introduction

The use of artificial intelligence (AI) in financial markets has grown rapidly, as AI has the capacity to capture complex nonlinear and dynamic effects that more traditional approaches can not [1]. Behavioural finance research often cites anomalies, biases and effects such as overconfidence bias, loss aversion, anchoring bias, herding behaviour, the disposition effect, momentum effects, overreaction and under-reaction as significant to understanding investor behaviour [2]. These bias effects are difficult to interpret, predict and measure. The use of AI offers a way forward in detecting behavioural signals from trading, financial and social media data at scale. AI-based detection and prediction have the potential to transform behavioural finance from a descriptive to a predictive science.

While empirical research is increasingly using AI in behavioural finance, the field is scattered. Some explore specific biases, others assess particular models, and many rely on small data sets. The application of various AI methods to study the same bias adds to this issue. This makes it difficult to draw comparisons across findings and develop general conclusions.

This article offers a structured review and conceptual integration of the use of AI in behavioural finance. It integrates information from several empirical studies and identifies a common ground, relates AI techniques to specific behavioural biases and data sources, and discusses limitations related to interpretability, generalisability and privacy. Rather than presenting individual empirical research questions, this review addresses the following four analytical goals: the comparison between AI-based detection and prediction and the status quo; the data sources used in the literature; the impact of AI-aware decision tools on costly errors from the effects of behavioural biases; and the challenges of interpretability, privacy and generalisability. The purpose of this study is to shed light on the role of AI in behavioural finance and to provide insights for future research.

2. Literature Review

2.1 Investor Behavioural Biases in Financial Markets

Previous behavioural finance studies primarily focused on the impact of cognitive and emotional biases on investor behaviour, trading volumes and prices. Rather than assuming that investors make fully rational decisions, this research demonstrates that investors' beliefs and decisions are influenced by reference points, recent investment performance, imitation and attention skewness [3]. A standard finding is that behavioural biases are not errors in the sense that they are systematic and may affect individual and aggregate outcomes.

Overconfidence studies typically conclude that investors who overvalue the information they have or their trading ability tend to trade too often, underestimate downside risk and are less likely to diversify their portfolios [4]. Other studies of loss aversion and the disposition effect demonstrate that investors tend to hold away gains too soon and hold onto losses too long, possibly at the expense of long-term portfolio returns [5]. These studies indicate that investor bias can be detected in expressed attitudes and in trading patterns.

Studies on anchoring, herding and momentum phenomena broaden the focus from individual investors to market dynamics. Anchoring research indicates that investors may be too influenced by previous prices, buy prices or analyst forecast targets in valuations [6]. Herding studies show investors may follow the actions of others in the face of uncertainty, leading to correlated order flows and prices that do not reflect fundamentals [7]. Momentum and overreaction studies also suggest that prices may be too slow or too fast to incorporate information, particularly when investors extrapolate recent price trends or overreact to news [8].

This literature makes three points that are relevant for AI. First, behavioural biases are reflected in trading data, portfolios, stock price reactions and investor chatter. Second, a given pattern can result from multiple behavioural biases, making it hard to pin down a single mechanism. Third, classic survey or econometric methods are helpful but can be too slow and not scalable for real-time bias detection. This opens the methodological door for AI models capable of analysing big, diverse and evolving data.

2.2 AI and Machine Learning in Bias Detection and Prediction

Recent research has shown how AI can be used to model behavioural biases. Gupta and Rao [9] used supervised and unsupervised machine learning algorithms to model trading data, sentiment from financial news and social media, finding that Random Forest classifiers were superior to SVM and Decision Trees in identifying overconfidence, confirmation bias and loss aversion. Their research is valuable because it highlights the association of different model types with different behavioural tasks; but their findings also highlight a crucial limitation of this field, which is that despite high classification accuracy, this does not necessarily explain the underlying reason for a bias or whether the pattern detected will be robust across markets.

Neural networks offer new possibilities because they are capable of handling sequential and unstructured data. Long Short-Term Memory (LSTM) networks can be used to detect loss-aversion patterns because of their ability to capture sequential portfolio revisions, whereas Convolutional Neural Networks (CNNs) can be applied to news and social media data to detect overreaction and anchoring biases [10]. Reinforcement learning has also been suggested to capture adaptive learning, and model investor behaviour in changing portfolios under risk [11]. These approaches extend behavioural finance as they model bias as an evolving process.

Predictions have been extended by profiling investors using unsupervised clustering methods, such as K-Means. For instance, Gupta and Rao [9] found rational-analytical, emotion-driven and trend-following investors, where the emotion-driven investors were more biased and more predictable. This adds to the behavioural finance literature on diverse investor types and the scalability of AI for investor profiling. On the other hand, the application of clustering methods needs to be interpreted with caution as group labels could represent researchers' prior beliefs rather than observed psychological states.

The research on AI-assisted decision-making also indicates that cognitive features may enhance predictions over financial features. Zhang et al. [12] demonstrate that AI-assisted decisions can be affected by human biases, suggesting that behavioural bias should be considered not only among investors but also in the interaction between investors and AI systems. That is an important consideration because AI may eliminate some human biases, but introduce new forms of bias, such as automation bias.

The network and multi-modal approaches also broaden the scope. Singh, Kaunert and Gautam [7] used AI tools to study herding and market overreaction, while Khuntia et al. [13] advocate for models that combine sentiment measures derived from NLP with market and consumer financial behaviours. The above studies indicate that bias detection should not be based on one source of data. What would be more beneficial is to combine order flow, portfolio data, news and social media sentiment. The problem with the current research is that many studies continue to test models in isolated environments and do not test whether the same model works for different investors, assets or institutions.

Liaudinskas [14] measured the disposition effect in algorithmic and human professional day traders, using millisecond-stamped trade-by-trade data from Copenhagen Stock Exchange. The results showed human traders had a strong disposition effect but algorithms did not. This is a useful distinction as it highlights that AI can help us discover what is human about the trading process. But it also suggests that if more trading is "outsourced" to algorithms, behavioural finance will need to consider not only human biases, but design incentives and failure mechanisms of algorithms.

Overall, across the papers reviewed, the relationship between bias, data and AI algorithm could be presented using the text rather than a table. Overconfidence is usually identified through excessive trading, overly positive sentiment in analyst reports or social media or forecast errors; Random Forests, ensemble learning and Natural Language Processing (NLP) models are often applied in this context [9]. Loss aversion or the disposition effect is typically studied using portfolios, buy-sell orders or brokerage data, where the LSTM model, decision trees and reinforcement learning models can detect the inclination not to sell losers or sell winners prematurely [10, 11, 14]. Anchoring typically relates to past prices, buying prices, IPO prices or target prices, and can be studied through anomaly detection, clustering or regression models [6]. Herding is usually detected in order flow, fund holdings and correlated trading networks, with logistic classifiers, unsupervised learning and graphs being helpful [13]. Momentum, overreaction and underreaction are usually analysed on return histories, volume, earnings announcements, macroeconomic news and press releases, with sequence and text-based learning providing the means to model delayed or excessive reactions to information [8, 10]. As this overview reveals, AI is most clearly applied when the data source is related to the behavioural process.

2.3 AI and Machine Learning in Bias Reduction

A recent study explores the application of reinforcement learning in reducing biases. Li, Zhao and Huang [11] show that adaptive trading bots using reinforcement learning can adjust their trading strategies in response to human investors' loss aversion. This moves the emphasis from descriptive modelling of biases to prescriptive strategies that address or exploit biases.

Beyond reinforcement learning, other literature studies the role of AI in designing decision support systems to mitigate behavioural mistakes. Saltık et al. [15] show that machine learning models forecasting loss aversion can be integrated into decision-support tools to notify traders when they show signs of loss aversion. These systems act as digital nudge strategies, promoting more rational investment while allowing greater control autonomy.

Another potential application is AI-based robo-advisors. When the investor's trading pattern suggests overconfidence, the system can recommend diversification in response to excessive trading or portfolio concentration. But the advisers' effectiveness depends on user understanding and trust. If the output is not comprehensible, investors might ignore helpful warnings or blindly follow inappropriate recommendations.

Another body of research focuses on the bias reduction potential of explainable AI (XAI). LIME and SHAP, for instance, help users understand what variables such as past return, volatility or sentiment indices are used to make a recommendation [16, 17]. By improving the interpretability of decision-support systems, XAI may encourage investors to be more aware of their own biases. Interpretability thus serves a technical and behavioural role.

Evidence on cost reduction is preliminary. To date, research has shown that AI-aware tools can alleviate loss aversion, momentum chasing and overreaction in a back-tested or simulated setting [11, 14]. However, many findings are preliminary, as implementation in the real world is contingent on trust, design, regulation and market forces.

2.4 Interpretability and Privacy Concerns

A recurring theme in the literature is model interpretability. Many modern ML methods are black boxes, making them difficult to use for policy or advisory purposes. Methods like LIME and SHAP are used to explain the role of features such as past performance, volatility or sentiment in bias-driven trading [16, 17]. A number of authors warn that although AI can resolve biases, it can also create new biases. Liaudinskas [14] demonstrates that algorithmic traders did not suffer from the disposition effect, but the increasing adoption of automated trading systems may lead to automation bias when human traders misinterpret or misapply predictions.

Additionally, there are on-going ethical debates about whether firms might use such tools not to benefit investors, but to capitalise on behavioural biases. For instance, firms might encourage investors to invest in higher fee products by “correcting” bias. Gupta and Rao [9] stress that although ML models can be used to detect behavioural traits at large scale, privacy, ethics and the potential for “gaming” the prediction of bias must be at the forefront of research.

2.5 Generalisability

Multimarket research is starting to address the question of generalisability. Khuntia et al. [13] suggest that the application of AI in behavioural finance depends on the data, market and investor environment. A model built in one market may not work as well in another due to the impact of institutional environment, market microstructure, regulatory environment and investor culture on the types and magnitude of behavioural biases and their detection. This suggests caution when relying on AI-based bias detection for advisory and regulatory purposes and the importance of out-of-sample and cross-market testing.

3. Discussion

The reviewed papers consistently show that AI approaches significantly increase the detection and prediction of behavioural biases over the traditional econometric approaches. Detection is the classification of biases (such as overconfidence and loss aversion) in near real time, and prediction is anticipating the occurrence of biases. For example, sequence models can foreshadow the presence of momentum trading or overreaction before observing such patterns, while clustering can explain which groups of investors are more prone to biases during crises. This detection-prediction feature of AI positions AI in a diagnostic and preventative role, which takes us from a reactive to a proactive approach to managing biases in behavioural finance.

A key insight of this review is that AI is a useful framework, but using it can be problematic. Explainability is a recurrent theme, with black boxes being powerful predictors but not transparent. Transparency is also an issue, with models trained for one market not necessarily translating to another, especially when there are different investor and regulatory regimes. Finally, there are concerns about privacy and ethics, especially as bias detection can be used to negatively impact investors.

A key point is that the next step in research is beyond model development. Progress needs to move beyond simply testing models to build more sophisticated models which can incorporate data from many sources, including transaction and portfolio data, market microstructure factors, financial and social media news. It also needs interpretable AI models which can explain predictions to investors, advisers and regulators. So, the future research should be more accurate, safe, investor-friendly and publicly verifiable.

4. Conclusion

This review demonstrates the increasing use of artificial intelligence and machine learning in behavioural finance. It draws on a number of studies to demonstrate how AI algorithms can detect behavioural biases and in some cases even predict them from trading data, financial documents and social media. This is helpful as it helps transform research from retrospective to predictive and preventative. Anticipating bias can help investors, firms and regulators.

This study is useful as it brings order to a diverse field, identifies what techniques work for what biases, and highlights research opportunities. Theoretically, the study contributes to behavioural finance by showing how psychology can be dynamically captured from different sources of financial data. Technically, it stresses the trade-off between accuracy and interpretability. Pragmatically, it suggests that bias-sensitive tools could prevent costly mistakes and foster responsible financial practices.

This research should prioritise more explainable AI techniques, multi-source data integration and empirical analysis with real data. Addressing these issues will move us from diagnosis to treatment, and help investors and regulators to use AI to guide them, not confuse them.

References

[1] Ahmed, S., Alshater, M.M., Ammari, A.E. and Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. Research in International Business and Finance, 61, 101646. https://doi.org/10.1016/j.ribaf.2022.101646.

[2] Shalini Srivastav and Gupta, A. (2022). AI Application and Psychological Biases in Behavior Finance. ICCAKM 2022, pp. 1-7. https://doi.org/10.1109/ICCAKM54721.2022.9990393.

[3] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), pp. 1124-1131. https://doi.org/10.1126/science.185.4157.1124.

[4] Nkukpornu, E., Gyimah, P. and Sakyiwaa, L. (2020). Behavioural Finance and Investment Decisions: Does Behavioral Bias Matter? International Business Research, 13(11), p. 65. https://doi.org/10.5539/ibr.v13n11p65.

[5] Barberis, N. and Xiong, W. (2009). What drives the disposition effect? An analysis of a long-standing preference-based explanation. Journal of Finance, 64(2), pp. 751-784. https://doi.org/10.1111/j.1540-6261.2009.01448.x.

[6] Hilary, G., Moeller, T. and Zhang, W. (2020). Anchoring and financial decision-making: Evidence from markets. Bayes Business School Working Paper. Available at: https://bayes.citystgeorges.ac.uk/__data/assets/pdf_file/0011/79472/Hilary-et-al.pdf.

[7] Singh, B., Kaunert, C. and Gautam, R. (2024). Artificial Intelligence in Detecting Herding and Market Overreaction. In Advances in Marketing, Customer Relationship Management, and E-Services, pp. 1-22. https://doi.org/10.4018/979-8-3693-7827-4.ch001.

[8] Ansh, R. (2023). Behavioral finance: Why does momentum work? A behavioural anomaly. Wright Research. Available at: https://www.wrightresearch.in/blog/why-does-momentum-work-a-behavioural-anomaly.

[9] Gupta, S. and Rao, V.S. (2025). AI in Behavioral Finance: Understanding Investor Bias Through Machine Learning. Journal of Informatics Education and Research, 5(2), pp. 5870-5880. https://doi.org/10.52783/jier.v5i2.3114.

[10] Zhou, P. and Wu, L. (2024). Investor overreaction detection via convolutional neural networks. Neurocomputing, 549, pp. 151-168. https://doi.org/10.1016/j.neucom.2024.03.012.

[11] Li, H., Zhao, Y. and Huang, X. (2023). Reinforcement learning for investment strategy adjusting to behavioral biases. Artificial Intelligence in Finance, 12(3), pp. 45-60.

[12] Zhang, Z., Wei, D., Varshney, K.R., Dhurandhar, A. and Tomsett, R. (2022). Deciding fast and slow: The role of cognitive biases in AI-assisted decision-making. arXiv. Available at: https://arxiv.org/pdf/2010.07938.

[13] Khuntia, R., Ingle, A.S., Manna, R., Srivastava, A.K. and Dwivedi, K. (2025). Decoding Consumer Financial Behavior: AI Applications in Behavioral Finance. Advances in Consumer Research, 2(3), pp. 725-732.

[14]Liaudinskas, K. (2022). Human vs. machine: Disposition effect among algorithmic and human day traders. Barcelona School of Economics Working Paper No. 1133. Available at: https://bse.eu/sites/default/files/working_paper_pdfs/1133.pdf.

[15] Saltık, Ö., Rehman, W.U., Söyü, R. and Şengönül, A. (2023). Predicting loss aversion behavior with machine-learning methods. Humanities and Social Sciences Communications, 10, 161. https://doi.org/10.1057/s41599-023-01620-2.

[16] Ribeiro, M.T., Singh, S. and Guestrin, C. (2016). 'Why should I trust you?': Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144. https://doi.org/10.1145/2939672.2939778.

[17] Lundberg, S.M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, pp. 4765-4774.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

Acknowledgment
This paper is an output of the science project.