International Journal of Information Technology Research and Development (IJITRD)
Volume 6 Issue 3, May – June (2025), pp. 13-18.

DEVELOPMENT OF AN AI-POWERED ENGINE FOR BEHAVIORAL BIAS DETECTION IN INVESTMENT ADVISORY

Agatha Christie William,
Behavioral Finance Analyst, Japan.

Abstract

Investment advisors often fall prey to behavioral biases that negatively impact financial decision-making. This study proposes and develops an AI-powered engine to detect behavioral biases—such as overconfidence, anchoring, and loss aversion—based on advisors' communications and portfolio decisions. Using machine learning models trained on historical advisory and market performance data, the engine flags bias-consistent patterns and provides corrective feedback. Our preliminary results indicate strong predictive validity and real-time detection capacity. This innovation holds promise for improving advisory outcomes and ensuring regulatory compliance.

Keywords: *Behavioral Finance, AI in Investment Advisory, Bias Detection, Overconfidence Bias, Anchoring, NLP in Finance.*

Cite this Article: William, A. C. (2025). Development of an AI-Powered Engine for Behavioral Bias Detection in Investment Advisory. International Journal of Information Technology Research and Development (IJITRD), 6(3), 13–18.

1. Introduction

Behavioral biases significantly distort investment decision-making, often resulting in suboptimal portfolio performance and investor dissatisfaction. Despite the proliferation of financial advisory services, few frameworks exist for systematically identifying and mitigating advisor-side biases. These cognitive and emotional deviations, such as confirmation bias, availability heuristics, and overconfidence, skew perceptions of risk and value. The problem is magnified in fast-paced market environments where advisors rely on heuristics under pressure.

In response to this challenge, artificial intelligence (AI), particularly in the domains of natural language processing (NLP) and behavioral modeling, offers new opportunities. This study aims to build an AI-powered system that evaluates advisors' decision-making patterns and discourse for signals of behavioral bias. The engine incorporates real-time text analytics and historical decision analysis to identify the presence of cognitive distortions. We envision this tool serving as a diagnostic and training resource for advisory firms, regulators, and behavioral researchers.

2. Objective and Scope

The primary goal is to develop and validate an AI-driven system that detects key behavioral biases in investment advisors' behavior using structured and unstructured data. The system should analyze client communications, portfolio adjustments, and market commentary, providing timely bias flags and actionable feedback. The research focuses on high-prevalence biases like overconfidence, anchoring, recency effect, and loss aversion.

The engine is expected to perform in real-time, integrating with customer relationship management (CRM) tools and investment platforms. It will be trained using historical advisory data and validated against ground-truth labels derived from expert judgment and market deviation patterns. The scope excludes direct consumer advice applications and focuses solely on institutional advisory contexts.

3. Literature Review

Research into behavioral biases in financial contexts is well-established. **Kahneman and Tversky (1979)** laid the groundwork with *Prospect Theory*, emphasizing loss aversion and framing effects. **Barberis and Thaler (2003)** synthesized developments in behavioral finance, identifying systematic irrationalities in investor behavior. **Shefrin (2000)** categorized common advisory errors, arguing for structured interventions to reduce them.

In the AI domain, **Lo (2004)** proposed the Adaptive Markets Hypothesis, suggesting that market actors, including advisors, adapt heuristics in response to environmental cues—sometimes maladaptively. **Chen et al. (2016)** used sentiment analysis in analyst reports to identify overconfidence, while **Das and Chen (2007)** explored NLP techniques to extract investor emotions from financial forums. Recently, **Baker et al. (2020)** analyzed tweets and financial texts to build predictive models for market mood. These contributions underpin the rationale for integrating AI to detect behavioral distortions in investment advice.

4. Methodology

4.1 Data Collection and Preprocessing

Data sources include anonymized advisor-client email logs, investment recommendations, and corresponding market data from 2017–2023. Natural language preprocessing involves tokenization, lemmatization, sentiment tagging, and syntactic parsing.

4.2 Modeling Approach

We apply supervised machine learning (Random Forest, XGBoost) and transformer-based NLP models (e.g., BERT fine-tuned on financial text) to classify advisory content for bias indicators. Labeling uses an expert-verified taxonomy of bias triggers.

Bias Type	NLP Feature Example	Labeling Cue
Overconfidence	"I'm certain this stock will rise"	High certainty modal verbs
Anchoring	"This is a fair price, like last Q"	Reference to outdated benchmarks
Loss Aversion	"We can't afford to sell now"	Emotionally loaded loss terms

5. Results and Evaluation

The proposed AI-powered behavioral bias detection engine was empirically validated using a labeled dataset comprising 1,200 unique investment advisor decision instances. These instances included both structured investment actions (e.g., buy/sell orders) and unstructured textual communication (e.g., advisor-client emails, investment memos). Each instance was annotated for one or more behavioral biases by a panel of finance and behavioral psychology experts, forming the ground truth labels for supervised learning.

Three models were evaluated: a **Logistic Regression (baseline classifier)**, a **Random Forest model**, and a **fine-tuned FinBERT model**—the latter being a domain-adapted variant of BERT pretrained on financial texts. Model performance was assessed using standard classification metrics: **precision**, **recall**, and **F1-score**, defined as follows:

- *Precision* indicates the proportion of predicted bias instances that were correctly classified.
- *Recall* measures the model's ability to identify all actual bias instances.
- *F1-score* is the harmonic mean of precision and recall, offering a balanced assessment.

5.1 Model Performance Metrics

Model	Precision	Recall	F1-Score
Logistic Regression	0.71	0.68	0.69
Random Forest	0.78	0.74	0.76
FinBERT (fine-tuned)	**0.89**	**0.87**	**0.88**

The Logistic Regression model served as a baseline, demonstrating moderate accuracy but limited capacity for capturing the nuanced semantic patterns often associated with behavioral biases, particularly in complex sentences. Random Forest improved classification by leveraging feature interactions more effectively, particularly in capturing keyword co-occurrences and syntactic structures indicative of cognitive distortions (e.g., hedging, excessive certainty).

5.2 Confusion Matrix and Error Analysis

An error analysis of the FinBERT model revealed that most false positives were attributable to ambiguous phrasing that mimicked bias indicators but lacked confirmatory context. For example, expressions like *"I believe this will go up"* were sometimes flagged as overconfidence, despite being cautious forecasts. Conversely, false negatives often arose when biases were implied through tone or implication rather than explicit phrasing, highlighting potential areas for future improvement in affective and pragmatic language modeling.

5.3 Practical Implications

The results suggest that transformer-based models such as FinBERT can significantly enhance the accuracy of behavioral bias detection in financial advisory contexts. The high recall rate (87%) is especially important for practical deployment, as it ensures that most true bias instances are identified and flagged for advisor review. Furthermore, the interpretability module—integrated with attention-based visualization of bias-laden phrases—received positive qualitative feedback during usability testing with 15 advisory professionals. Over **85% of participants** agreed that the tool's output was "intuitively understandable and professionally actionable."

6. System Integration and Ethical Considerations

6.1 System Architecture and Integration

The AI-powered engine for behavioral bias detection is architected for seamless integration with existing financial advisory workflows. It is built using a modular, microservices-based structure, allowing it to function as a plug-in or standalone component in enterprise environments. Specifically, the engine exposes RESTful APIs and supports common integration protocols (e.g., OAuth 2.0, JSON, XML), enabling interoperability with Customer Relationship Management (CRM) systems, Portfolio Management Systems (PMS), and client interaction dashboards. This allows real-time ingestion of advisor communications—emails, chat logs, and investment notes—without disrupting existing infrastructure.

6.2 Ethical Compliance and Data Privacy Protocols

The deployment of AI in financial advisory raises significant ethical and regulatory considerations, particularly around **data privacy**, **algorithmic fairness**, and **decision transparency**. The system has been developed in accordance with the **General Data Protection Regulation (GDPR)** and sector-specific regulatory standards such as the **Financial Conduct Authority (FCA)** guidelines and **U.S. SEC AI oversight frameworks**.

7. Conclusion

This study presents a novel approach to detecting behavioral biases in investment advisory using AI. By leveraging machine learning and NLP on textual and decision data, the engine provides real-time feedback to improve decision quality and client outcomes. Future research will extend the model to multilingual advisory contexts and explore reinforcement learning for adaptive bias correction.

The proposed system demonstrates not only technological feasibility but also strong alignment with behavioral finance theory. It offers a scalable solution to a persistent industry challenge, combining academic rigor with practical applicability.

References

1. Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." Econometrica, vol. 47, no. 2, 1979, pp. 263–291.

2. Biru, S. (2025). Transforming Investment Banking Middle Office: A Framework for Advanced Security and Data Management. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11(1), 608–616. https://doi.org/10.32628/CSEIT25111268

3. Barberis, Nicholas, and Richard Thaler. "A Survey of Behavioral Finance." Handbook of the Economics of Finance, edited by George M. Constantinides, Milton Harris, and René M. Stulz, vol. 1, Elsevier, 2003, pp. 1053–1128.

4. Shefrin, Hersh. Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing. Oxford UP, 2000.

5. Lo, Andrew W. "The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective." Journal of Portfolio Management, vol. 30, no. 5, 2004, pp. 15–29.

6. Das, Sanjiv R., and Mike Y. Chen. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web." Management Science, vol. 53, no. 9, 2007, pp. 1375–1388.

7. Biru, S. (2025). Intelligent Automation in Banking Operations: Impact Analysis on Renewable Energy Investment Assessment. International Journal of Computer Engineering and Technology (IJCET), 16(1), 673–687. https://doi.org/10.34218/IJCET_16_01_056

8. Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media." Review of Financial Studies, vol. 29, no. 9, 2016, pp. 2061–2085.

9. Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan. "Global, Local, and Contagious Investor Sentiment." Journal of Financial Economics, vol. 104, no. 2, 2020, pp. 272–287.

10. Tetlock, Paul C. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." Journal of Finance, vol. 62, no. 3, 2007, pp. 1139–1168.

11. Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. "Salience Theory of Choice under Risk." Quarterly Journal of Economics, vol. 127, no. 3, 2012, pp. 1243–1285.

12. Biru, S. (2025). AI-Powered Deduplication in Investment Banking Middle Office. International Journal of Research in Computer Applications and Information Technology (IJRCAIT), 8(1), 1713–1723. https://doi.org/10.34218/IJRCAIT_08_01_125

13. Kogan, Shimon, et al. "Predicting Risk from Financial Reports with Regression." Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, 2010, pp. 272–280.

14. Bholat, David, et al. "Machine Learning and Big Data in Finance: The Case of Risk Management." Bank of England Quarterly Bulletin, vol. 57, no. 3, 2017, pp. 1–12.

15. Barber, Brad M., and Terrance Odean. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." Journal of Finance, vol. 55, no. 2, 2000, pp. 773–806.

16. Shiller, Robert J. Irrational Exuberance. 3rd ed., Princeton UP, 2015.

17. Feng, Lei, and Mark S. Seasholes. "Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Finance Markets?" Review of Finance, vol. 9, no. 3, 2005, pp. 305–351.

18. Luss, Ronny, and Alexandre d'Aspremont. "Predicting Abnormal Returns from News Using Text Classification." *Quantitative Finance*, vol. 15, no. 6, 2015, pp. 999–1012.