MSW MANAGEMENT **-Multidisciplinary, Scientific Work and Management Journal**
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 1720-1728

Behavioral Finance Through Time: Tracing the Past, Analyzing the Present, and Forecasting the Future in the Age of AI

Sanjay Kumar
Scholar at Amity Business School, Noida, and
Executive Director, in Indian Oil
Email ID: sanjay.kumar2@s.amity.edu
OCID ID : 0009-0007-4071-9828
Amity Rd, Sector 125, Noida, Uttar Pradesh 201301

Dr. Anshu Singh,
Associate Professor,
Amity Business School, Noida
Email: asingh42@amity.edu
OCID ID : 0000-0001-9447-6250
Amity Rd, Sector 125, Noida, Uttar Pradesh 201301

Dr. Gautam Negi,
Associate Professor,
Lal Bahadur Shastri Institute of Management, Delhi
Email: gautam.negi@lbsim.ac.in
11/07 Dwarka Sector 11, near Metro Station, New Delhi, Delhi 110075

Abstract
This paper traces the evolution of behavioral finance from classical economic assumptions of rational agents to the integration of psychological insights that highlight biases and heuristics in decision-making. It then examines how the rapid rise of artificial intelligence (AI) is reshaping the field, particularly in investment contexts. While AI promises precision and freedom from human emotion, the persistence of human oversight raises questions about whether biases will be eliminated or simply transformed.
Drawing on survey data from 379 Indian investors across three age groups, this study explores attitudes toward AI-driven financial decision-making. Using regression and nonparametric analyses, we find that investors demonstrate conditional trust in AI: they are willing to follow AI recommendations when backed by a proven track record, but remain reluctant to rely on AI exclusively. Notably, higher familiarity with AI reduces unconditional acceptance, suggesting that informed investors approach AI with greater caution.
The findings highlight the emergence of *meta-biases*—cognitive patterns that arise when humans evaluate or intervene in machine-generated decisions—signalling that behavioral finance will continue to evolve rather than disappear in the age of AI.
Key Words: Behavioral economics, bias, artificial intelligence, meta-biases, Trust, algorithm aversion

Introduction

Economics has long been concerned with human behavior, yet early financial theories often assumed perfectly rational actors. Behavioral finance emerged to challenge this view, integrating psychology to explain why individuals systematically deviate from rational predictions. Pioneering work by Kahneman, Tversky and Thaler showed that biases such as loss aversion, overconfidence, and heuristics significantly influence decision-making.
In recent years, however, the financial landscape has been reshaped by technological innovation. Artificial intelligence (AI) now plays a growing role in portfolio management, algorithmic trading, and risk assessment. Unlike humans, AI systems are not subject to emotions such as fear, greed, or euphoria. This raises a critical question: does the growing reliance on AI signal the end of behavioral finance, or will human tendencies continue to shape financial decisions in new ways?
Existing literature has examined behavioral biases extensively, yet little empirical work addresses how investors perceive AI-driven decision-making, especially in emerging markets. The integration of AI introduces both opportunities and challenges: it may reduce biases by automating processes, but investor mistrust and the need for human oversight suggest that biases may persist or even take new forms.
This paper addresses three central questions:
1. Are investors willing to adopt AI-generated financial decisions without overriding them?
2. Does greater AI literacy influence acceptance of AI recommendations?
3. Do attitudes toward AI differ across generational groups?
To answer these questions, we conduct a survey of 379 Indian investors segmented by age group and analyze their

responses using regression and non-parametric statistical methods. The results reveal a nuanced picture: while investors are inclined to trust AI with a strong success record, they hesitate to rely on it exclusively. More importantly, familiarity with AI is associated with greater scrutiny and skepticism rather than blind trust.
The contribution of this study is twofold. First, it provides empirical evidence on investor perceptions of AI adoption in an emerging market context. Second, it introduces the concept of *meta-biases*: the new psychological patterns that arise when humans oversee or correct AIgenerated recommendations. These findings suggest that rather than rendering behavioral finance obsolete, AI is transforming its scope by shifting the ways in which biases manifest.
The remainder of this paper is organized as follows: it first reviews the literature on behavioral finance and the role of AI in decision-making, then describes the methodology and data, presents the empirical results, discusses the findings in the context of existing theory, and concludes with implications and directions for future research.

Literature Review

Classical and Behavioral Foundation

The intellectual roots of behavioral finance can be traced back to classical economics, where early scholars recognized that human psychology shapes economic decisions. Adam Smith (Smith, 1776, 1790) noted tendencies toward overconfidence and the greater psychological weight of losses relative to gains. Pigou (Pigou A. C., 1920) highlighted "present bias," or the tendency to overvalue immediate rewards compared to future benefits. However, as economics evolved into a mathematically formal discipline, psychological insights were largely sidelined in favor of models assuming perfectly rational and self-interested agents(Bernoulli, 1954; Neumann Von John & Morgenstern Oskar, 1944; Samuelson Anthony Paul, 1947) .

By the mid-twentieth century, critics of the rational-agent paradigm began to argue for models grounded in empirical behavior. Simon(Simon, 1955)introduced the notion of bounded rationality, suggesting that decision-making is limited by information constraints and cognitive complexity. Subsequent work challenged the assumption of predictive accuracy in economic models, emphasizing instead the importance of observed behavioral anomalies (Friedman Milton, 1966; Herbert A. Simon, 1972).

This intellectual shift culminated in the emergence of behavioral economics and behavioral finance. Kahneman and Tversky's (Kahneman & Tversky', 1979) Prospect Theory provided a systematic framework for understanding decision-making under risk, demonstrating that individuals are risk-averse when facing gains but risk-seeking when facing losses. Their work on heuristics (Tversky & Kahneman, 1974) showed that reliance on mental shortcuts often leads to systematic biases such as anchoring, availability, and representativeness. Subsequent research confirmed these patterns in financial contexts: Thaler (R. Thaler, 1980)on the endowment effect, Samuelson and Zeckhauser (Samuelson & Zeckhauser, 1988) on status quo bias, and Thaler and Sunstein's (R. H. Thaler & Sunstein, 2008)"nudges" as tools to counteract cognitive inertia.

Together, this body of work reshaped financial theory by acknowledging that decisions are influenced not only by data and rationality but also by emotions, heuristics, and bounded judgment.

Behavioral Finance in Modern Market

The recognition of psychological influences has important implications for understanding asset pricing, market efficiency, and financial stability. The Efficient Market Hypothesis (Fama, 1970) posits that prices fully reflect available information, but evidence of anomalies has repeatedly challenged this view. Shiller (Robert J. Shiller, 2003)demonstrated that equity markets exhibit levels of volatility inconsistent with fundamentals. Thaler (R. H. Thaler, 2016)catalogued persistent anomalies, including closed-end fund discounts and investor overreactions to political events such as the Cuba fund surge. Historical cases, such as Newton's losses in the South Sea Bubble (Odlyzko, 2020), illustrate the enduring role of human emotion in driving financial excess.

Scholars such as Taleb (Taleb, 2007) and Minsky (Minsky, 1977) have also emphasized systemic risks arising from speculative behavior and the underestimation of rare, extreme events. Akerlof's (Akerlof, 1970) seminal work on asymmetric information further illustrates how mistrust and informational imbalances can erode market stability, generating environments in which conventional rational-choice models lose explanatory power. Complementing these perspectives, Odean and Gervais (Gervais & Odean, 2001) demonstrate that overconfidence-induced overtrading exacerbates volatility, thereby amplifying market instability.

Empirical work in behavioral finance has since extended beyond identifying anomalies to examining interventions that shape financial outcomes. Thaler and Benartzi's (R. H. Thaler & Benartzi, 2004) "Save More Tomorrow" program leveraged behavioral insights to improve retirement savings by addressing loss aversion and inertia. These contributions underscore the growing recognition that financial systems are both socially and psychologically constructed.

Technology and the Future of Behavioral Finance

While behavioral finance has expanded our understanding of decision-making, recent technological advancements pose new challenges. The rise of artificial intelligence (AI), big data, and algorithmic trading introduces a new dimension: the possibility that biases traditionally observed in humans may be reduced—or reshaped—by automation.

AI applications in finance are already widespread. Predictive analytics and machine learning models can detect subtle behavioral signals in large datasets, such as social media sentiment and trading patterns, enabling forecasts at both individual and aggregate levels (Huang et al., 2025; Zulaikha et al., 2025). Neuro-symbolic reasoning, which integrates neural networks with symbolic logic, offers improved interpretability in evaluating financial risk (Nawaz et al., 2025; Vijayrao Chaudhari & Ashokrao Charate, 2025). These developments suggest that decision-making no longer depends solely on human heuristics, but also on algorithmic outputs.

Yet, the presence of AI does not eliminate behavioral dynamics. Scholars argue that individuals who show concern about algorithms themselves may embed biases from training data or reflect the preferences of their designers (Aracri et al., 2023). Concerns over manipulation, transparency, and fairness persist, influencing whether investors are willing to trust algorithmic decisions(Orbán & Stefkovics, 2025). Onyenahazi and Antwi (Onyenahazi & Antwi, 2024) highlight both the opportunities and risks that AI poses for financial institutions, noting that efficiency gains are often tempered by skepticism regarding ethical robustness.

Importantly, research on investor perceptions of AI remains limited, particularly in emerging markets. Public opinion studies suggest that while individuals recognize the benefits of AI, they remain hesitant to cede full control to algorithms(Luca Liehner et al., 2023). This tension creates a new frontier for behavioral finance: instead of asking how human biases affect markets directly, scholars must now consider *meta-biases*—the cognitive patterns that arise when humans evaluate, supervise, or override machine-generated recommendations.

Although behavioral finance has illuminated the influence of biases on financial decision- making, the rapid integration of AI raises new questions. Existing studies have primarily examined how biases distort traditional market behavior, but little is known about how investors perceive and adapt to AI-driven decisions. The emerging literature highlights conditional trust in algorithms but has yet to systematically investigate differences across demographics or levels of AI literacy, particularly in the context of developing economies.

This study seeks to address that gap by analyzing Indian investors' perceptions of AI in financial decision-making. By examining the interplay between trust, perceived AI superiority, and familiarity, the research provides insight into how traditional biases may persist, evolve, or transform in the age of automation.

Methodology

Research Design This study adopts a survey-based design to investigate investor perceptions of artificial intelligence (AI) in financial decision-making. The survey was structured to capture attitudes toward trust in AI, perceived accuracy, willingness to follow AI advice, and concerns regarding manipulation. The approach is exploratory, aiming to provide initial evidence from an emerging market context.

Sampling and data collection

Data were obtained from 379 respondents through a combination of structured questionnaires online (350) and targeted interviews (29). A snowball sampling strategy was employed; although non-probabilistic in nature, this approach is particularly appropriate for exploratory research contexts where the target population is not readily accessible through systematic sampling procedures. Care was taken to ensure that participants possessed adequate financial capacity to engage in market investments and/or demonstrated literacy in artificial intelligence, thereby aligning the sample with the study's objectives. Furthermore, to enhance heterogeneity and mitigate sampling bias, deliberate efforts were made to incorporate respondents representing diverse professional backgrounds and life stages The sample was segmented into three age groups to explore generational differences: below 30 years (33.7%), 31–45 years (37.4%), and above 45 years (28.9%). These groups broadly align with Generations Z, Millennials, and Generation X, though the analysis focuses on age- based differences rather than generational identities per se. For the purposes of the study, respondents' AI literacy was categorized into two groups: those with little to no prior experience with AI (54.57%) and those possessing moderate to extensive knowledge of AI (45.43%).

Measurement Instruments

Survey items were measured on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree). The instrument included six core dimensions:

1. Trust in AI-generated investment suggestions
2. AI can accuracy for complex decisions
3. Belief that AI outperforms human experts
4. Willingness to follow AI when success rates are high
5. Willingness to rely exclusively on AI without human consultation

Reliability was satisfactory, with Cronbach's $\alpha = 0.81$, indicating acceptable internal consistency.

Schilke and Reimann (Schilke & Reimann, 2025) identify a noteworthy paradox in their study, demonstrating that transparency regarding the use of AI, rather than enhancing trust, may paradoxically diminish it. Building on this insight, the present research included a question on perceptions of manipulation to examine whether concerns about manipulative intent contribute to mistrust in AI.

6. Perceived manipulability of AI by third parties

Hypotheses

The study addressed three main hypotheses:

H_0: Investors are not ready to adopt AI-generated decisions.

H_1: Investors are ready to adopt AI-generated decisions.

H_0: Higher AI literacy does not increase acceptance of AI recommendations.

H_1: Higher AI literacy increases acceptance of AI recommendations.

H_0: AI perceptions do not differ across age groups or AI-knowledge levels.

H_1: AI perceptions differ across age groups and AI-knowledge levels.

Analytical Approach

Given the deviations from normality observed in the item distributions, the study employed a multi-method analytical strategy combining descriptive statistics, regression modeling, and non-parametric testing (Khang Trinh et al., 2025). To examine predictors of reliance on AI, regression analysis was conducted with willingness to rely exclusively on AI without human consultation as the dependent variable, and trust in AI, perceived AI superiority, and familiarity with AI as key independent variables. In addition, Mann–Whitney U tests were employed to examine generational differences, providing a non-parametric assessment of variations across age groups and differing levels of AI-related knowledge(Schreibelmayr et al., 2023). This integrative approach balances methodological rigor with the exploratory nature of the dataset, allowing for both descriptive interpretation and robust hypothesis testing.

Findings

Descriptive Statistics

Descriptive analysis of the six survey dimensions (Table 1) highlights patterns of conditional trust in AI. The highest mean was recorded for "Willingness to follow AI when success rates are high" (M = 3.60, CI95% [3.51 - 3.69]), indicating strong agreement among respondents. In contrast, "rely exclusively on AI without human consultation" had the lowest mean (M = 2.47, CI95% [2.37 - 2.57]), reflecting clear reluctance to fully delegate decision-making. Perceived manipulability of AI scored relatively high (M = 3.70), suggesting that concerns about external interference remain significant.

Descriptive Data (Table 1)

	Trust in AI-generated investment suggestions	AI can accuracy for complex decisions	Belief that AI outperforms human experts	Willingness to follow AI when success rates are high	Willingness to rely exclusively on AI without human consultation
Mean	2.96	3.04	3.16	3.6	2.47
Std. Deviation	0.87	0.85	0.96	0.92	1.01
Skew	-0.36	-0.26	-0.34	-0.7	0.31
Kurtosis	0.51	-0.2	-0.53	0.38	-0.62
95% Confidence interval for mean	2.87 - 3.05	2.96 - 3.13	3.06 - 3.26	3.51 - 3.69	2.37 - 2.57

Regression Analysis

Regression results (Table 2) tested the relationships between trust, perceived AI superiority, AI familiarity, and willingness to rely exclusively on AI.

- Trust in AI was a significant positive predictor ($\beta = 0.36$, $p < 0.001$).
- Perceived AI superiority over human experts was also significant ($\beta = -0.19$, $p < .001$).
- AI familiarity, however, showed a significant negative effect ($\beta = -0.22$, $p < .001$), indicating that greater knowledge of AI reduces unconditional reliance.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 1720-1728

- Other variables, including perceived accuracy and historical success rates, were not statistically significant predictors. Overall, the model explained 32% of the variance ($R^2 = 0.32$), underscoring the importance of trust and perceived capability, alongside skepticism arising from familiarity.

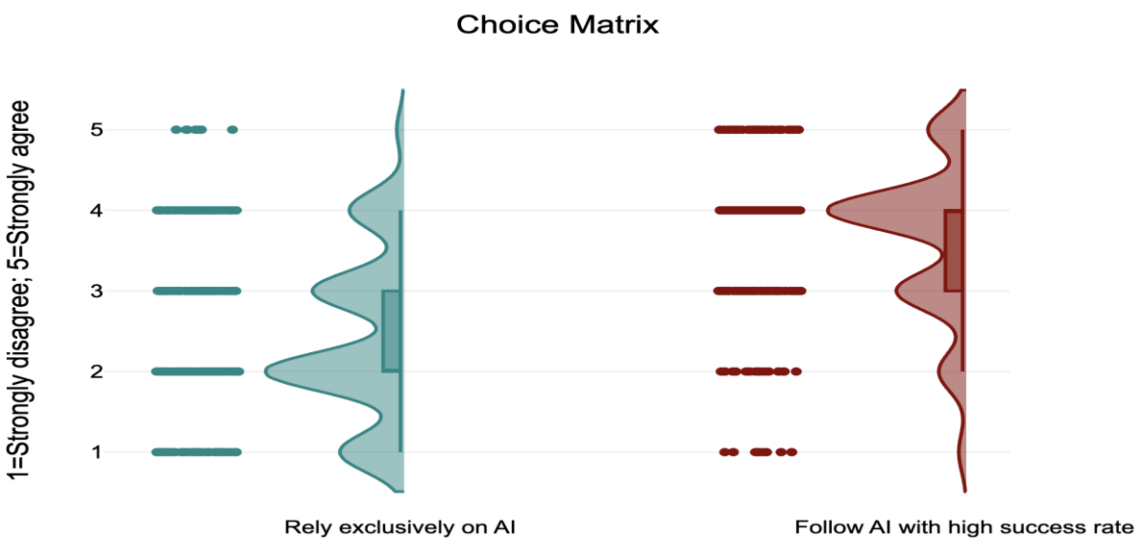
Regression Analysis (Table 2)

Coefficients					
Model	Un-standard. Coef. B	Standard. Coef. Beta	Std. Error	t	p
Constant	1.6		0.27	5.96	<.001
Familiarity with AI	-0.28	-0.22	0.06	-5.08	<.001
Trust in AI-generated investment suggestions	0.42	0.36	0.06	7.13	<.001
Belief that AI outperforms human experts	0.17	0.16	0.05	3.18	0.002
Willingness to follow AI when success rates are high	0.14	0.12	0.06	2.36	0.019
Perceived manipulability of AI by third parties	-0.19	-0.19	0.04	-4.39	<.001

- Indipendent variable : Rely exclusively on AI without human consultation.

The regression model was statistically significant, accounting for approximately 32% of the variance in the outcome. Among the predictors, trust in AI-generated investment suggestions emerged as the strongest positive determinant, while belief in AI's superiority over human experts and willingness to follow AI when success rates are high also contributed positively, though with smaller effects. Interestingly, familiarity with AI was a significant negative predictor. This somewhat counterintuitive result aligns with prior findings by Tully (Tully et al., 2025) and Schilke and Reimann (Schilke & Reimann, 2025), who observed that increased AI literacy or disclosure may reduce receptivity by heightening critical scrutiny. Similarly, perceived manipulability of AI by third parties exerted a negative influence on acceptance of AI-driven investment recommendations. Such concerns are not unfounded, as Tarsney (Tarsney, 2025) has demonstrated in his work on Deception and Manipulation in Generative AI, highlighting the risks of third-party influence and covert manipulation.

As illustrated in the raincloud plot (Figure 1), the findings reveal a nuanced stance toward AI-driven recommendations. While participants demonstrate a willingness to consider and follow the guidance provided by AI systems, they also exhibit a pronounced reluctance to accept such recommendations in the absence of human consultation. This suggests that trust in AI remains conditional and is mediated by the perceived necessity of human oversight, highlighting the ongoing importance of human judgment in decision-making processes that involve algorithmic outputs.



Taken together, these findings suggest that the acceptance of AI in investment decision-making is fostered by trust and confidence in AI systems, but simultaneously constrained by scepticism arising from overfamiliarity and fears of

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 35 Issue 2, 2025, Pages: 1720-1728

manipulation. The negative association between familiarity with AI and acceptance may be interpreted as reflecting the greater awareness among informed individuals of AI's limitations, biases, and potential vulnerabilities, leading to heightened caution rather than uncritical adoption. These dynamic warrants further empirical exploration to better understand the conditions under which familiarity fosters scepticism versus when it contributes to more calibrated trust.

Familiarity and Generational Difference

The Mann–Whitney U test revealed that familiarity with AI significantly shapes trust and reliance on AI in investment contexts (Table-3). Participants with different levels of familiarity showed notable differences in their trust in AI-generated investment suggestions and their belief that AI can outperform human experts, as well as in their willingness to rely solely on AI advice without human consultation. By contrast, no significant differences were observed in perceptions of AI's accuracy in complex investment decisions or in the likelihood of following AI when supported by historical success rates. These results suggest that familiarity primarily influences trust and exclusivity of reliance, rather than evaluations of technical accuracy or evidence-based performance.

Test of Familiarity with AI (Table -3)

Variable	U statistic	p value
Trust in AI-generated investment suggestions	14625	0.001
AI can accuracy for complex decisions	17321.5	0.527
Belief that AI outperforms human experts	14714.5	0.001
Willingness to follow AI when success rates are high	16625.5	0.178
Willingness to rely exclusively on AI without human consultation	20116.5	0.034
Group 1: Individuals with limited or basic familiarity with AI		
Group 2: Individuals with moderate to expert familiarity with AI		

The Mann–Whitney U test revealed that age has a selective influence on perceptions of AI in investment contexts. Significant differences emerged between participants below 30 and those aged 31–45 in both their trust in AI-generated investment suggestions ($U = 10646$, $p = .03$) and their willingness to follow AI advice when supported by strong historical success ($U = 10745.5$, $p = .03$). By contrast, no significant age-related differences were observed in perceptions of AI's accuracy, beliefs about AI outperforming human experts, or the willingness to rely exclusively on AI advice. These findings indicate that younger and middle-aged investors differ mainly in their trust and conditional reliance on AI, whereas broader evaluations of AI's technical competence remain consistent across age groups. More study needs to be done to find out fault lines amongst the different age groups.

Discussion

Overview of Findings

This study investigated investor attitudes toward AI-driven financial decision-making. The results highlight a pattern of conditional trust: investors are willing to follow AI when it demonstrates a record of success but remain reluctant to rely on it exclusively. Trust in AI and perceived superiority over human experts were significant positive predictors of reliance, while familiarity with AI had a negative effect, reducing unconditional acceptance.

Familiarity with AI plays a critical role in shaping trust in the outcomes produced by algorithmic systems. Concerns about potential misrepresentation or distortion of information by AI give rise to a novel form of bias—one that extends beyond the traditional biases and heuristics previously documented in the literature.

Generational differences were minimal, with only younger respondents showing slightly higher openness to exclusive reliance on AI.

Interpreting conditional trust dynamic

The results suggest that investors approach AI not as an autonomous authority but as a tool whose credibility must be earned. Consistent with technology adoption theory, trust and demonstrated effectiveness increase willingness to delegate decisions(Venkatesh et al., 2003) . At the same time, the reluctance to remove human oversight reflects broader concerns about fairness, transparency, and manipulability in algorithmic systems(Aracri et al., 2023; Orbán & Stefkovics, 2025).

The role of AI familiarity: Towards Meta-Biases

Perhaps the most significant contribution is the finding that greater familiarity with AI decreases willingness to rely on it without interference. Contrary to the assumption that knowledge breeds confidence, familiarity appears to foster critical scrutiny. Informed investors may recognize issues such as training bias, model drift, and susceptibility to manipulation, leading them to preserve human oversight.

This finding introduces the concept of meta-biases—cognitive patterns that emerge when humans evaluate or second-guess machine-generated outputs. Examples include:

- Control preference bias (algorithm-aversion): the tendency to retain final authority even when delegating to an algorithm. (Dietvorst et al., 2018).
- Safeguard heuristic (Human-in-the-Loop): the belief that human intervention reduces the risk of error (Sele & Chugunova, 2024).
- Transparency bias (Black-box): scepticism toward opaque or "black-box" models, regardless of their accuracy (Hassan et al., 2025).

These meta-biases indicate that rather than disappearing, behavioral tendencies are adapting to the presence of AI. Behavioral finance must therefore expand its scope to study not only human decision-making but also human–machine interaction.

Implications

Biases don't disappear with AI; they simply move from the decision itself to how investors evaluate and trust AI systems. Behavioral finance must therefore study *when and why* investors trust AI, alongside traditional biases.

Implications for Theory and Theory Development

This study contributes to the advancement of behavioral finance by introducing the concept of meta-biases—cognitive patterns that emerge when humans evaluate, supervise, or override AI-generated decisions. The findings demonstrate that biases are not eliminated in an AI-mediated environment; rather, they shift toward the assessment of algorithmic processes. This theoretical reframing requires scholars to extend behavioral finance beyond traditional human decision-making anomalies to encompass human–AI interaction dynamics. By empirically showing that greater AI familiarity reduces unconditional acceptance, the research highlights the paradox that knowledge fosters scrutiny instead of blind trust. Future theoretical work must therefore examine how varying levels of AI literacy shape decision-making, and how meta-biases interact with classical biases to influence financial outcomes.

Implications for Readers, Business, and Management Practice

For practitioners, the findings underscore the importance of building trust in AI systems through transparency, explainability, and evidence of consistent performance. Financial institutions must recognize that while strong success records encourage adoption, investors remain reluctant to rely exclusively on AI without human oversight. Interestingly, higher AI literacy may increase skepticism, suggesting that education campaigns should strike a balance—informing users about AI's strengths while addressing limitations openly.

For AI developers, this implies the need to design hybrid human–AI systems where human oversight is retained to enhance acceptance. Policymakers and regulators should prioritize ethical safeguards and disclosure standards to mitigate fears of manipulation. Finally, for individual readers and investors, the results encourage reflection on how meta-biases such as algorithm aversion, safeguard heuristics, and transparency concerns shape financial decisions. By acknowledging these dynamics, investors and institutions can adopt more calibrated, trust-enhancing approaches to AI integration in financial practice.

Limitation and future research

This study has several limitations. The reliance on snowball sampling limits generalizability, and the survey data are restricted to Indian investors. Replication with larger, representative, and cross-country samples would enhance external validity. Additionally, the study measured perceptions rather than actual behavioral outcomes. Future research could employ experiments or longitudinal designs to examine how investors act when confronted with real AI-generated recommendations.

Conclusion

This study examined investor perceptions of artificial intelligence (AI) in financial decision- making, with a focus on whether automation reduces or reshapes traditional behavioral biases. Survey of Indian investors reveals a pattern of conditional trust: investors are willing to follow AI recommendations when supported by proven success, but remain hesitant to rely on AI exclusively. Trust and perceived superiority over human experts increase willingness to adopt AI, while familiarity with AI reduces unconditional reliance, highlighting the persistence of skepticism even among informed users.

Contributions

The findings make three contributions.

- Theoretical: The study introduces the concept of meta-biases—cognitive patterns that emerge when humans oversee or correct algorithmic decisions. This extends behavioral finance by showing that biases do not disappear in the age of AI but instead shift toward the evaluation of machine intelligence.
- Empirical: By analyzing survey data from Indian investors, the study provides novel evidence from an emerging

market context, addressing a gap in the literature that has been dominated by Western perspectives.

- Practical: For practitioners and policymakers, the results highlight the importance of transparency, explainability, and balanced communication in fostering trust in AI systems. While performance matters, investor acceptance depends equally on perceived fairness and control.

Implications and future research

The results suggest that AI will not render behavioral finance obsolete but will redefine its focus. Human oversight ensures that biases remain embedded in financial decision-making, even as algorithms mitigate some errors. For financial institutions, this underscores the need to design AI tools that are both effective and trustworthy, integrating explainability features to address investor concerns.

Future studies could expand on this work by employing experimental or longitudinal designs to observe actual behavior when investors are presented with AI-generated advice. More emphasis should be given to study the three meta biases algorithm-aversion, safeguard heuristic or Human-in-the-Loop, and Transparency bias or Black-box.

Comparative research across countries would also provide insights into cultural and institutional factors shaping trust in AI. Finally, investigating domain-specific applications— such as robo-advisory, credit scoring, or algorithmic trading—could further clarify how acceptance varies across financial contexts.

In sum, rather than erasing behavioral finance, the rise of AI is likely to open a new chapter. Biases will persist, not in spite of automation but through the ways humans judge, adapt, and intervene in machine-driven decision.

Reference:

Akerlof, G. A. (1970). The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. In *The Quarterly Journal of Economics* (Vol. 84, Issue 3).

Aracri, S., Pisoni, G., Faiz, M., Faiz, I., & Brauner, P. (2023). What does the public think about artificial intelligence?—A criticality map to understand bias in the public perception of AI. *Frontiers in Computer Science*, *5*.

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. In *Econometrica* (Vol. 22, Issue 1).

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, *25*(2), 383–417.

Friedman Milton. (1966). The Methodology of Positive Economics : In Essays In Positive Economics. *University of Chicago Press*, 3 to 16–30 43.

Gervais, S., & Odean, T. (2001). Learning To Be Overconfident. *The Review of Financial Studies*, *14*(1), 1–27.

Hassan, R., Nguyen, N., Finserås, S. R., Adde, L., Strümke, I., & Støen, R. (2025). Unlocking the black box: Enhancing human-AI collaboration in high-stakes healthcare scenarios through explainable AI. *Technological Forecasting and Social Change*, *219*. https://doi.org/10.1016/j.techfore.2025.124265

Herbert A. Simon. (1972). Theories of Bounded Rationality. In C. B. McGuire & R. Radner (Eds.), *Decision and Organization* (pp. 161–176). North Holland Publishing Company.

Huang, X., Li, H., & Wei, N. (2025). Big Data-Dr iven Behavior al Economics: Analysis of Decision Biases in Financial Mar kets. *International Journal of Global Perspective in Academic Research*, *2–1*, 38–48. https://ijgpar.org/index.php/journal/index

Kahneman, D., & Tversky’, A. (1979). PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK. *ECNOMETRICA*, *47*(2), 263–291.

Khang Trinh, T., Jia, G., Cheng, C., & Ni, C. (2025). *Behavioral Responses to AI Financial Advisors: Trust Dynamics and Decision Quality Among Retail Investors*. https://doi.org/10.54254/2755-2721/144/2025.21859

Luca Liehner, G., Hick, A., Biermann, H., Brauner, P., & Ziefle, M. (2023). Perceptions, attitudes and trust toward artificial intelligence — An assessment of the public opinion. *Artificial Intelligence and Social Computing*, *72*, 32–41. https://doi.org/10.54941/ahfe1003271

Minsky, H. P. (1977). The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to “Standard” Theory. *Challenge*, *20*(No.1 (MARCH/APRIL 1977),), 20–27.

Nawaz, U., Anees-ur-Rahaman, M., & Saeed, Z. (2025). A review of neuro-symbolic AI integrating reasoning and learning for advanced cognitive systems. In *Intelligent Systems with Applications* (Vol. 26). Elsevier B.V. https://doi.org/10.1016/j.iswa.2025.200541

Neumann Von John, & Morgenstern Oskar. (1944). Theory of Games and Economic Behaviour. *Princeton University Press*, 498–504.

Odlyzko, A. (2020). Isaac Newton and the perils of the financial South Sea. In *Physics Today* (Vol. 73, Issue 7, pp. 30–36). American Institute of Physics Inc. https://doi.org/10.1063/PT.3.4521

Onyenahazi, O. B., & Antwi, B. O. (2024). The Role of Artificial Intelligence in Investment Decision-Making: Opportunities and Risks for Financial Institutions. *International Journal of Research Publication and Reviews*, *5*(10), 70–85. https://doi.org/10.55248/gengpi.5.1024.2701

Orbán, F., & Stefkovics, Á. (2025). Trust in artificial intelligence: a survey experiment to assess trust in algorithmic decision-making. *AI and Society*. https://doi.org/10.1007/s00146-025-02237-6

Pigou A. C. (1920). *THE ECONOMICS OF WELFARE* (Fourth Edition). MACMILLAN AND CO., LIMITED ST. MARTIN'S STREET, LONDON.

Robert J. Shiller. (2003). From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, *17*(Number I-Winter), 83–104.

Samuelson Anthony Paul. (1947). *Foundations Of Economic Analysis: Vol. Sixth Print* (Sixth Print). Oxford University Press.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*. https://doi.org/10.1016/j.obhdp.2025.104405

Schreibelmayr, S., Moradbakhti, L., & Mara, M. (2023). First impressions of a financial AI assistant: differences between high trust and low trust users. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1241290

Sele, D., & Chugunova, M. (2024). Putting a Human in the Loop: Increasing Uptake, but Decreasing Accuracy of Automated Decision-Making Putting a Human in the Loop: Increasing Uptake, but Decreasing Accuracy of Automated Decision-Making *. *PLOS ONE*, *19*(2). www.rationality-and-competition.de

Simon, H. A. (1955). A BEHAVIORAL MODEL OF RATIONAL CHOICE. *Quarterly Journal of Economics*, *69*, 99–118.

Smith, A. (1776). *THE WEALTH OF NATIONS* (Vol. 1, pp. 95–96).]. M. DENT & SONS LTD, E. P. DUTTON & CO. INC.

Smith, A. (1790). *The Theory of Moral Sentiments* (Sixth Edition). MetaLibri.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House. www.rh-newsletters.com

Tarsney, C. (2025). Deception and Manipulation in Generative AI. *Philosophical Studies*, *182*, 1865–1887. https://doi.org/10.1007/s11098-024-02259-8

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Thaler, R. H. (2016). Behavioral economics: Past, present, and future. In *American Economic Review* (Vol. 106, Issue 7, pp. 1577–1600). American Economic Association. https://doi.org/10.1257/aer.106.7.1577

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*(1). https://doi.org/10.1086/380085

Thaler, R. H., & Sunstein, C. R. (2008). *NUDGE*. Yale University Press.

Tully, S. M., Longoni, C., & Apple, G. (2025). Lower Artificial Intelligence Literacy Predicts Greater AI Receptivity. *Journal of Marketing*, *89*(5).

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *Source: MIS Quarterly*, *27*(3), 425–478.

Vijayrao Chaudhari, A., & Ashokrao Charate, P. (2025). Self-Evolving AI Agents for Financial Risk Prediction Using Continual Learning and Neuro-Symbolic Reasoning. *JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING*, *13*(2), 76–92. https://doi.org/10.70589/jrtcse.2025.13.2.8

Zulaikha, S., Mohamed, H., Kurniawati, M., Rusgianto, S., & Rusmita, S. A. (2025). CUSTOMER PREDICTIVE ANALYTICS USING ARTIFICIAL INTELLIGENCE. *Singapore Economic Review*, *70*(4), 1009–1020. https://doi.org/10.1142/S0217590820480021