International Journal of Management and Sustainability Practices
Vol. 1, Issue 3 (July–Sept 2026): Pages 82-89
e-ISSN (Online): 3139-5449
Website: https://ijmsp.com/

Behavioral Biases and Investment Decision-Making: The Mediating Role of Trust in AI and the Moderating Role of Financial and Digital Financial Literacy: A Narrative Literature Review

Rahul Negi*[1], Kaushal Kumar[2]

[1]Hemwati Nandan Bahuguna Garhwal University (HNBGU), Srinagar, Uttarakhand, India
[2]D.A.V. (P.G.) College, Dehradun, Uttarakhand, India
*Corresponding Author Email: rahulnegi844@gmail.com

Article Information

Submission History:

Received: 08 August, 2026
Revised: 12 August, 2026
Published: 16 August, 2026

Keywords:

Behavioral biases
Investment decision-making
Trust in artificial intelligence
Digital financial literacy
Explainable AI
Behavioral finance

Abstract

Narrative review of the research concerning the link between behavioral biases, investment decision making, and the rise of AI-powered financial technology is presented in this paper. In particular, attention is paid to the impact that the key concepts of behavioral finance such as overconfidence, herding, loss aversion, anchoring, mental accounting, and regret aversion have on the behavior of the investors. Then, the paper proceeds to the analysis of the rising popularity of robo-advisors and other AI-powered financial solutions, underlining the importance of investor's trust to these technologies for their adoption and effectiveness. Algorithmic transparency and explainability of AI become particularly notable factors influencing confidence and adoption of AI-generated financial recommendations. The reactions of the investors to both behavioral biases and technology-enabled financial services are considered from the perspective of financial literacy of the investors. The paper suggests that these literacies might help investors cope with the digital nature of finances and make more effective decisions based on data from earlier researches. There is also a mention of the differences between demographic groups, like age or gender. The paper analyzes several researches to reveal common understanding, debates, and gaps in this field not to test any hypothesis. It should be noted that the impact of behavioral biases, investor's trust, financial literacy, and the development of artificial intelligence in finance is underlined in the conclusion of the paper.

1. Introduction

One of the most frequent assumptions in finance which is constantly being challenged is that investors are rational when investing. With the development of Prospect Theory by Kahneman and Tversky (1979), showing that people estimate results relative to a reference point and value the losses more than the gains of the equal amount, there appeared a great number of studies about the impact of cognitive biases on investor's behavior (Barberis & Thaler, 2003; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). The purpose of this review paper is not to conduct a new experiment based on this hypothesis, but to show how the literature has changed over time – how the discussion has shifted from the description of cognitive biases that influence investor's decision making to the question of how artificial intelligence affects that behavior and how financial literacy plays its role in the process (Baker & Ricciardi, 2014; Ricciardi & Simon, 2000).

Timing, thus, constitutes the rationale for the revival of this literature. Algorithmically generated robo-advisers and financial planners have, within the past fifteen years or so, gone from being a novelty to becoming a common feature in retail investment practices, and industry predictions suggest that more retail investors would get some kind of AI-based financial advice in the future (Jung et al., 2018; Belanche et al., 2019). At the same time, another body of literature regarding users' trust in AI-generated financial advice has evolved to stress that such trust and acceptance on the part of the users of algorithmic recommendations require, above all else, reliability and competence of these recommendations (Glikson & Woolley, 2020; Logg et al., 2019). Finally, recent developments in the literature on transparency of algorithms and Explainable Artificial Intelligence (XAI) address the growing need for the transparency of "black box" recommendations for trust-based AI usage (Arrieta et al., 2020; Rai, 2020). On the other hand, research on financial literacy has begun increasingly to recognize the presence of a particular digital aspect of that financial literacy, given that many financial choices have become digitized and do not involve a human counterpart anymore but are taken through a digital interface (Morgan et al., 2019; OECD, 2023). In general, these four streams of research—behavioral biases, trust in AI, algorithmic transparency, and financial literacy—have been developed independently of each other without a meaningful dialogue between the strands. The review of the literature brings these streams together, tracing the developments of the questions and debates as they evolved. This paper takes an analytical approach by discussing the existing literature thematically, but not chronologically in each theme. Firstly, it discusses the already existing literature regarding behavioral biases and their relation to investment decisions (Aren & Hamamci, 2020; Goyal et al., 2022). Then, it discusses how these biases can differ across different generations (Glikson & Woolley, 2020; Rai, 2020). After that, it looks at the current research on the role of trust in AI-powered finance tools, as well as how transparency contributes to this kind of trust. Finally, it discusses how financial and digital financial literacy affect these relationships (Lusardi & Mitchell, 2014; Morgan et al., 2019).

2. Review Approach

It should be noted that this is not a systematic review but a discursive synthesis of the literature, which explores how various theories and ideas emerged and intersect or differ from each other (Baumeister & Leary, 1997; Green et al., 2006). The majority of sources for this paper are peer-reviewed articles from the areas of behavioral finance, financial technology, human-computer interaction, and financial literacy studies. Additionally, if there were any reviews in the areas mentioned above, they have been used as well since systematic reviews and meta-analysis are useful for understanding the present state and evolution of any research area (Baumeister & Leary, 1997; Snyder, 2019). The scope of research was restricted to those studies that examined individual (retail) investors' behavior, artificial intelligence-powered or algorithmic investment advice, algorithmic transparency and explainability, and financial or digital financial literacy. References to literature on institutional investors and market microstructure were made only in cases where this literature provided some insights into the problem of retail investor decision making. Due to the nature of the underlying literature which belongs to the fields of finance, information systems and psychology, there may be some inconsistencies in the use of terminology.

3. Behavioral Biases and Investment Decision-Making

3.1. The Core Biases: Overconfidence and Herding

Overconfidence and herding emerge as the two most prominent biases that receive attention in the behavioral finance literature. An empirical investigation into more than 40 years of studies revealed that institutional investors herd more compared to individual investors, whereas individual investors are overconfident on the whole compared to institutional investors irrespective of shifts in measurement tools (Goyal et al., 2022; Kumar & Goyal, 2015). Overconfidence has been shown to cause excessive trading and increased risk taking. A classic piece of literature that is highly cited is that of Barber and Odean (2000), who revealed the relationship between overconfidence and excessive trading and showed that overconfident traders trade too often and earn less, an idea succinctly captured in the phrase "trading is hazardous to your wealth." More recent studies have added to this body of literature by demonstrating that overconfidence can be enhanced or undermined through feedback effects and the information environment in which the traders find themselves (Biais et al., 2005; Glaser et al., 2007).

On the other hand, the topic of herding is notable for its relevance to market-level phenomena like bubbles and crashes, and not just outcomes at the level of an individual’s portfolio. While overconfidence is a type of cognitive bias related to one's own judgment, herding is a type of cognitive bias related to the informational value of the behavior of others. Individuals can behave herd-like both rationally due to the belief that others know more or irrationally as a consequence of psychological satisfaction that results from conforming to others’ behavior (Bikhchandani & Sharma, 2001; Devenow & Welch, 1996). The scientific literature fails to reach a consensus about the relative importance of each mechanism and several surveys of the existing literature have pointed out that different ways of measuring herding, based on self-report surveys, correlations of volumes traded or similarities of portfolios held, often produce difficult-to-compare results (Goyal et al., 2022; Kumar & Goyal, 2015).

3.2. Loss Aversion, Anchoring, Mental Accounting, and Regret Aversion

Aside from overconfidence and herding, numerous studies have analyzed anchoring, loss aversion, disposition effect, mental accounting, and regret aversion biases, yet there is no consensus about their importance compared to each other. A review combining evidence of multiple behavioral biases showed that even though both overconfidence and herding biases are rather stable in terms of impact on investor behavior, some biases like mental accounting and self-control bias seem to be highly context-dependent (Kumar & Goyal, 2015). A similar review concerning behavioral biases in the Egyptian stock market showed that loss aversion and regret aversion biases seem to have a great impact on investor decision-making, while mental accounting and self-control biases seem to be less important (Mousa et al., 2023).

The concept of anchoring, described as a situation in which one tends to overly rely on the first piece of information, such as the purchase price of a stock or market peak, when interpreting future data, has been found among both individual and institutional investors (Tversky & Kahneman, 1974). Anchoring is often mentioned in conjunction with the phenomenon called the disposition effect, which implies the behavior of investors to prematurely sell their successful investments and hold their unsuccessful investments for too long (Shefrin & Statman, 1985). Both these behaviors can be explained by Prospect Theory. Together, the story emerging from the literature above indicates a fairly stable body of biases, such as overconfidence and herding, surrounded by additional behaviors that depend on the type of market, the level of expertise of the investors, and the research approach employed (Goyal et al., 2022; Kumar & Goyal, 2015). What is missing from the literature until recently is a consideration of how these biases interact in an increasingly algorithmic and AI-influenced environment, and how these interactions differ for different types of investors. The topics will be further discussed below.

3.3. Demographic and Generational Variation in Behavioral Bias

There has been considerable amount of research conducted on behavioral biases according to gender, age, and experience, which is helpful in understanding who would be impacted more by the concept of biases, trust, and literacy in the following paragraphs of this review. Differences in gender were observed in a consistent manner, though not necessarily uniform, across studies. In studies involving individual investors, men have shown a tendency towards significantly higher self-attributions, illusions of control, and overconfidence, while women have shown higher tendencies of regret aversion and risk aversion (Barber & Odean, 2001; Aren & Hamamci, 2020). Likewise, findings among Indian investors reveal that women investors are less prone to overconfidence bias compared to their male counterparts (Bashir et al., 2013). It is important to mention that some researchers have proven the possibility for financial literacy to level out these gender-related disparities. People with higher financial literacy tend to demonstrate less behavioral biases in general, and there is not much difference in performance between men and women anymore, which means that financial literacy can act as an equalizer instead of just lowering biases (Lusardi & Mitchell, 2014; Goyal et al., 2022).

Another layer of complexity is brought by age and generation-related differences. In survey-based studies with financial professionals, younger investors, who had less experience and education than their older counterparts, were more likely to see the effect of cognitive and emotional biases in investment decision-making process, while older and more experienced professionals exhibited higher levels of confidence and optimism (Pompian, 2012). Studies into the impact of digital investing platforms revealed that younger investors exhibited a greater inclination to speculate and to trade in a gambling-like way, interact more with crypto-currency markets, and used social media and influencers in the field more often when making investment decisions (OECD, 2024; Xiao & Porto, 2022). Evidence is emerging about Generation Z investors, which states that exposure to digital investments content makes them more vulnerable to behavioral biases like herding and overconfidence under highly volatile investment environment (Goyal et al., 2022).

Adoption studies in fintech offer an alternative view on the matter. Research analyzing the use of digital banking and mobile financial services has revealed that although all age groups tend to become more engaged in using financial technologies once access to digitization becomes available to them, the magnitude of the increase will be significantly higher in the case of younger respondents compared to older respondents (Morgan et al., 2019; Venkatesh et al., 2012). This evidence confirms the hypothesis that the interplay of behavioral biases, trust in technology, and financial advice through AI is not likely to impact all investors equally.

4. Trust in AI-Driven Financial Advice

4.1. Algorithm Aversion and the Trust Deficit

The literature concerning robo advisors and financial advisory tools based on artificial intelligence has expanded significantly from the moment when the first robo-advisory services have come into existence after the 2008 financial crisis (Jung et al., 2018; Belanche et al., 2019). The one key factor that appears repeatedly in the existing research is that trust, not availability or cost, plays the crucial role in the adoption of robo advice (Glikson & Woolley, 2020; Belanche et al., 2019). Specifically, there is research indicating that people tend not to accept advice of an algorithm even if it performs better than humans do, which has been referred to as algorithm aversion (Dietvorst et al., 2015). In addition, people are less inclined to trust algorithms if they make only one observable mistake even though they allow such mistakes when made by human beings (Dietvorst et al., 2015). Such asymmetry, in which an error by the algorithm apparently undermines trust more deeply than a similar mistake by a human being, is among the most persistent insights into the field of trust in the interactions between humans and machines (Glikson & Woolley, 2020; Logg et al., 2019). In this regard, the initial bad experiences with financial instruments based on AI may have strong and persistent effects on future behavior.

4.2. What Builds Trust: Design, Neutrality, and Anthropomorphism

Under these conditions of doubt, there has simultaneously been a body of research that attempts to explore the elements that encourage trust in AI financial tools. It is believed that the perceived objectivity of algorithms in comparison to human advice, their constant availability, as well as their interactive and timely feedback may positively influence investor's confidence and prevent the manifestation of emotional reactions like panicking sell-offs in times of market volatility (Belanche et al., 2019; Jung et al., 2018). There has been evidence found that anthropomorphic design elements like names and avatars of AI along with conversational interfaces may increase perceived competence, trust, and dependence on these systems by increasing cognitive and affective trust, especially in less experienced users (Glikson & Woolley, 2020; Longoni et al., 2019) .An international survey of young retail investors also showed that initial levels of trust in robo-advisers were positively related to the tendency toward trust, albeit influenced by other financial, social, and behavioral aspects such as savings behavior and performance expectations from investments (Belanche et al., 2019). The results point to the interrelatedness between trust in AI and the psychological disposition of investors involved in the process.

Here we encounter a link between the literature on trust and the literature on behavioral biases discussed above. Overconfident investors may be more inclined to experience algorithm aversion if AI-generated recommendations do not coincide with their beliefs (Barber & Odean, 2000; Dietvorst et al., 2015), as they tend to have higher self-trust than trust in the outside opinion. At the same time, individuals with herding tendency may start trusting robo-advisors more due to growing social acceptance and usage of the technology (Bikhchandani & Sharma, 2001). In recent studies in the field of robo-advising and fintech, trust has been increasingly perceived as the mediator between antecedents and outcomes like risk mitigation, investment performance, financial inclusion, and technology adoption (Belanche et al., 2019; Glikson & Woolley, 2020). Such approach is aligned with the assumption that trust in AI is one of the channels through which psychological traits impact investment behavior.

4.3. Algorithmic Transparency and Explainable AI

One of the recent and emerging streams within the relevant literature pays attention not only to the trust of investors towards AI-supported financial advice but also whether such trust is well justified by the reliability of the technology. This issue becomes crucial as regulators and business entities call for increased accountability and transparency in algorithmic decision making processes (Arrieta et al., 2020; Rai, 2020). The empirical analysis of AI-based fintech platforms shows that algorithmic transparency is one of the most powerful factors determining the level of investor trust, which influences the attitudes and investment decisions of investors. Digital financial literacy and risk perception can be also considered as significant factors of trust building in the context of technology-enabled financial services (Belanche et al., 2019; Morgan et al., 2019). As a general synthesis of XAI literature shows, transparency and explainability may affect the level of trust and acceptance, but highlights the fact that empirical studies focus on robo-advisory and fintech frameworks, thus failing to address the question of how transparency may help to build proper trust calibration, which would mean trusting trustworthy recommendations and mistrusting untrustworthy ones, instead of increasing the usage rate regardless of the recommendations' credibility (Arrieta et al., 2020; Rai, 2020).

Such a distinction proves to be especially crucial to the discussion presented in this paper. Indeed, research conducted in the field of explainable artificial intelligence suggests that people may demonstrate algorithm appreciation by embracing explanations produced by AI when the latter confirms their initial beliefs and neglecting the explanations that contradict the existing presumptions (Logg et al., 2019). Thus, explainability can prove to be ineffective in overcoming the biases, and it can give another avenue for biases like confirmation bias to act. Experimental research related to the interface of explainable AI has confirmed that not only the quality of explanations but also users' initial beliefs and predispositions can influence the effect of explanations (Gregor & Benbasat, 1999; Rai, 2020). Therefore, the approaches of transparency and explainability cannot be considered merely technical methods for resolving the problem described in the literature on algorithm aversion. They can interact with the psychological tendencies already inherent in the behavior of investors in different ways (Arrieta et al., 2020; Rai, 2020).

Indeed, the connection between behavioral bias, transparency and trust in the context of AI is more theory-driven than empirically proven. Indeed, a large portion of both robo-advisory and AI trust literature has developed separately from the classical behavioral finance literature, being primarily based on theories of technology acceptance, human-computer interactions and relational marketing-based approaches to understanding the concept of trust (Venkatesh et al., 2012; Glikson & Woolley, 2020), while the behavioral finance approach is always based on Prospect Theory and heuristics-and-biases models (Kahneman & Tversky, 1979; Tversky & Kahneman, 1974). Thus, what we have in the end is a literature that reflects two streams of research that are related to each other but not entirely integrated with each other and that explore the way in which biases influence decision-making and transparency and trust affect the adoption of AI respectively.

5. Regulatory and Ethical Considerations

5.1. Fiduciary Duty Applied to Algorithmic Advice

These dynamics of trust and transparency are not operating in a regulatory void, and an ever-increasing amount of practitioner-focused as well as legal literature has looked at the applicability of existing fiduciary and suitability regimes in the context of AI-based financial advice (Baker & Dellaert, 2018; SEC, 2017). Specifically, in the United States, the Securities and Exchange Commission (SEC) has taken the stance that the concept of a fiduciary is related to the nature of the financial advice offered, and not the process through which it is given, thereby meaning that robos must follow the same standards of care and loyalty as their human counterparts and that automation has not created a lesser form of regulation (SEC, 2017). Recent examination priorities of the SEC have also addressed the increased usage of artificial intelligence by advisory firms, especially in the context of portfolio management, trading, and marketing (SEC, 2024). Similar developments have been seen on an international scale. Regulatory authorities like the UK Financial Conduct Authority (FCA), European Securities and Markets Authority (ESMA), and Monetary Authority of Singapore (MAS) have mostly followed the principle of substance over form, whereas the Artificial Intelligence Act and the Digital Operational Resilience Act (DORA) of the European Union have added new requirements for transparency, accountability, operational resilience, and risk management during the AI lifecycle (European Parliament & Council of the European Union, 2024; ESMA, 2023; FCA, 2022).

This regulatory convergence is directly applicable to the literature on trust presented in Section 4. If the fiduciary duty mandates that AI-powered advisory operates in the best interests of the client no matter what delivery mechanism is used, then the investor trust in AI can also be viewed not only as trust in technology but also as trust in regulation (SEC, 2017). It is worth mentioning that enforcement activities concerning misleading statements regarding the application of artificial intelligence, which are called “AI-washing,” demonstrate that the differences between the capabilities of algorithms advertised and their true capabilities have become an issue for regulators (SEC, 2024). Such phenomena can be related to the literature on algorithm aversion because the frequent misleading of the investors concerning capabilities of AI might negatively affect the baseline willingness to trust the algorithm as suggested by various models of AI adoption (Dietvorst et al., 2015; Glikson & Woolley, 2020).

5.2. Ethical Considerations: Bias, Fairness, and Investor Vulnerability

In addition to this, there is a body of ethical literature that focuses on other aspects besides disclosure and fiduciary responsibilities regarding fairness and the design of AI advisory systems themselves. The scholars who are concerned with the ethical issues regarding AI-based financial advisory suggest that in addition to the competency of these systems, their trustworthiness, accountability, cybersecurity, and autonomy are the other factors influencing the adoption of such systems (Arrieta et al., 2020; Mittelstadt et al., 2016). Some authors even suggested a system of certification and ratings which would give investors an opportunity to assess these qualities (Floridi et al., 2018).

This ethical theory relates to the results obtained from the study of the demographic factors mentioned in Section 3.3. Considering that young, inexperienced, and financially illiterate people are more prone to certain types of cognitive errors and more likely to place blind faith in online finance services, they might be considered a target of increased investor protection efforts (Lusardi & Mitchell, 2014; OECD, 2023). This problem is also reflected in the regulatory statements on the need for special attention to new financial technologies and the way they relate to retail investors (FCA, 2022; SEC, 2024).

In combination, the two bodies of literature provide an institutional element to the psychological approach that was introduced in Sections 3 and 4. Not only is trust in investment advice provided via AI based on the characteristics of the algorithm and transparency and the biases of the investor but also on the integrity of the regulatory and disclosure environment within which it is set (Arrieta et al., 2020; SEC, 2017). In future research, an investigation of the connections between biases, trust, and financial literacy could thus be strengthened by including factors related to investor awareness of regulations and fiduciary obligations.

6. Financial Literacy as a Conditioning Factor

In addition, the other literature stream focuses on financial literacy as another factor which has an impact, and usually mitigates, the impact of behavioral bias on investment. Financial literacy, which is a broader term used to refer to the knowledge and skills required to make sound financial decisions, affects financial decision making, financial well-being, and risk-taking behaviors in many different settings (Lusardi & Mitchell, 2014; OECD, 2023). Financial literacy has generally been assessed through standardized tools assessing understanding of financial concepts like compound interest, inflation, and diversification of risk (Lusardi & Mitchell, 2014). In the special case of behavioral bias, review literature suggests that financial literacy is a moderating variable that mitigates the negative impact of behavioral biases on investment-related risk perception and decision making. Investigations undertaken in developing markets have also explored the same two variables, financial literacy and investor experience, as moderators of the effect of behaviorial biases, including heuristics and herding on investment decisions, treating financial literacy as an asset that helps investors process information in spite of their inherent cognitive biases (Bashir et al., 2013; Goyal et al., 2022).

In this context, the gender-related literature mentioned above brings in an interesting dimension. It does not seem as though financial literacy just eliminates behavioral biases across all investors; rather, it helps reduce the difference between various sub-groups. Research reveals that the difference between genders in terms of behavioral biases like self-attribution, confirmation bias, and over-confidence is significantly lesser in investors who are highly financially literate than in those who are less financially literate (Lusardi & Mitchell, 2014; Goyal et al., 2022). In summary, the academic literature is fairly unanimous regarding a conclusion that financially savvy investors are prone to exhibit biases as well, yet they seem to be better prepared to acknowledge their effects and reduce them when making investment decisions (Lusardi & Mitchell, 2014). Nevertheless, there is scarce evidence on the role of financial literacy in AI-based investment frameworks, since previous moderation analyses have mostly concentrated on conventional investment options.

7. Digital Financial Literacy and the Digital Turn

With the digitization of financial services, studies have suggested that just financial literacy is not adequate for the functioning of individuals in digital financial environments. Thus, the construct of digital financial literacy has been developed by researchers to define the abilities and competencies that are required to safely and effectively use digital financial products and services by making informed decisions in digitally-mediated environments (Morgan et al., 2019; OECD, 2023). As compared to financial literacy, whose measurement has been in practice for a long time, digital financial literacy is a relatively new construct about which debates regarding its exact definition and measurements are still ongoing (Morgan et al., 2019). Dimensions of digital financial literacy commonly used in studies include knowledge of digital financial products, understanding of digital financial risks and cybersecurity issues, digital financial technology competency, and digital consumer protections (Morgan et al., 2019; OECD, 2023).

Empirical research usually indicates that digital financial literacy is a crucial predictor of financial decision-making in digital settings. For example, Indian studies reveal that digital financial literacy serves as an independent and mediating variable of financial decision-making and financial well-being in conjunction with financial capability and financial autonomy (Rastogi et al., 2022). Additionally, research related to women investors has indicated that digital financial literacy and financial attitudes in combination with perceived behavioral control are crucial determinants of financial decision-making and investment intentions (Ajzen, 1991; Rastogi et al., 2022). In connection with the AI-trust literature reviewed above, research on trust in fintech platforms has revealed that digital financial literacy is a crucial predictor of trust in fintech, which means that it not only increases digital financial literacy but also contributes to the ability of investors to trust AI-based financial systems (Belanche et al., 2019; Morgan et al., 2019).

From this body of literature, one recurring implication is that digital financial literacy operates somewhat differently than traditional financial literacy. Digital financial literacy becomes particularly relevant in contexts where decision-making is dependent on digital interfaces, something that is a defining feature of AI-driven investment platforms (Morgan et al., 2019). Nonetheless, an important research gap remains. While there is more empirical evidence in favor of the hypothesis that digital financial literacy is a determinant of trust compared to that of its role in moderating the link between trust and investments, it is not known if digital financial literacy just affects trust but does not impact investments based on trust.

7.1. Summary of Key Studies

Table 1 summarizes a selection of the studies discussed above, organized by focus area, to give an at-a-glance view of how the literature's various strands relate to one another before turning to the full synthesis.

Table 1. Summary of Selected Studies across the Reviewed Literature

Focus Area	Representative Study	Key Finding	Context
Overconfidence & trading	Barber & Odean (2000)	Overconfident investors trade more and earn lower net returns	U.S. retail brokerage accounts
Systematic review of biases	Goyal et al. (2022)	Institutional investors herd more; individuals show more overconfidence	Systematic literature review
Gender & bias	Lusardi & Mitchell (2014)	High financial literacy narrows gender gaps in bias	Cross-study synthesis
Algorithm aversion	Dietvorst, Simmons & Massey (2015)	People avoid algorithms after seeing them err, even when accurate	Experimental / forecasting
Trust in robo-advisors	Belanche et al. (2019)	Initial trust tied to propensity-to-trust, moderated by behavioral factors	Cross-country, young investors
Algorithmic transparency	Arrieta et al. (2020)	Transparency and explainability shape trust and acceptance of AI systems	Broader XAI literature synthesis
Financial literacy moderation	Bashir et al. (2013)	Financial literacy and experience moderate the bias-to-decision relationship	Pakistan, individual investors
Digital financial literacy	Rastogi et al. (2022)	DFL is a direct and indirect predictor of financial decision-making and well-being	India, emerging economy

Source: Author’s Own Work

8. Synthesis: Toward an Integrated Understanding

Taken as a whole, however, these four bodies of research paint a complex picture rather than a single, coherent one. First, the literature on behavioral biases demonstrates quite convincingly that such psychological elements as overconfidence and herding play a significant role in investment decision-making, irrespective of the technology being used for such decision-making and irrespective of the investor's gender, age, and experience. Second, the literature on trust and trust mechanisms proves, to some extent more theoretical than empirical, that trust is a gatekeeping factor, deciding whether an AI-based recommendation is acted upon or not, and that such trust is determined not only by the design and transparency but also by the investor's psychological makeup.

Indeed, the growing field of explainability literature offers a helpful addition to this story in arguing that increased transparency does not necessarily imply properly calibrated trust due to the presence of individual biases among investors, namely confirmation bias. The financial literacy literature suggests that financial literacy helps mitigate the influence of bias on decision-making in the case of traditional investments and likely does so to a different extent among various demographic segments. Finally, the digital financial literacy literature demonstrates that there is another type of competence that influences financial decision-making in a digital environment, including an AI-based one, and there has been at least one relevant study published recently.

None of the studies mentioned here considers all of these concepts within the same design, and this fact in itself says something significant about the state of the research at hand: research in this area has evolved along adjacent tracks, with researchers working in behavioral finance investigating bias and its demographics, scholars in technology adoption and human-computer interaction investigating robo-advisor trust and explainability, and scholars in financial literacy investigating how literacy can be a protective factor, all largely not speaking to each other. The story is one of a field where all the building blocks have been gathered to understand the relationship between bias, trust, transparency, and literacy and their influence on investment behavior through AI, except for one empirical puzzle.

9. Gaps and a Future Research Agenda

These gaps can be identified from the review of literature above. Firstly, there is very little research linking the existence of particular biases and the trust people exhibit towards AI-based investment instruments. Most studies either investigate the connection between the former concept and traditional decision-making or the latter concept and the use of robo-advisors, but hardly do researchers ever consider the two topics simultaneously using one and the same dataset. Secondly, although there has emerged some data suggesting the connection between digital financial literacy and trust, the former concept's moderating effect on the latter one has never been studied.

The third point of departure involves the algorithmic appreciation effect noted in the explainable AI literature, whereby confirmation bias influences the way in which investors interpret AI explanations, and whereby this effect has been discussed primarily in conceptual terms without any empirical testing of this phenomenon relative to the classic behavioral biases found in the finance literature. The fourth point involves the fact that much of the existing trust in AI literature employs cross-sectional designs, such that questions of trust dynamics – especially after algorithmic error – remain an open issue, especially considering the findings of the algorithm aversion literature, wherein a single error asymmetry is noted, which implies slow recovery compared to formation of trust.

Furthermore, the demographic literature surveyed here, which includes literature on gender, age, and generational bias and engagement with fintech applications, has largely evolved separately from both the AI-trust and the AI-literacy literature, although the results from each clearly indicate that they probably interact, with an investigation into how patterns of demographic bias influence differential trust in and usage of advice provided by AI applications needed. Lastly, the regulatory and ethical literature reviewed in Section 5 has largely been generated thus far by legal researchers and practitioners, rather than behavioral finance researchers, with no studies here measuring how awareness of fiduciary protections or regulatory oversight influences individual trust in, and susceptibility to bias in the usage of, AI applications. Seven, a great majority of the behavioral bias research literature and a significant number of the digital financial literacy research also use country-specific or region-specific samples, like those of India, Pakistan, Egypt, and Taiwan in some of the articles reviewed here. This brings into question the generalizability of the findings to other markets with varying degrees of fintech adoption and varying demographic makeups.

Such gaps point toward what could be seen as a relatively straightforward research agenda for future empirical research: investigations that assess behavioral biases, trust in AI, perceptions of algorithmic transparency, financial literacy, and digital financial literacy simultaneously in one group of investors; experimental designs that enable the examination of whether the quality of explanations moderates the influence of the investors' existing biases as proposed by the algorithmic appreciation hypothesis; longitudinal designs that enable an assessment of how trust in AI varies over time or as a result of manipulation of transparency and mistakes in particular of asymmetric recovery from mistakes in algorithms; and cross-national/cross-demographic comparisons in order to evaluate whether the findings of previous single-market/single-demographic studies apply more generally.

10. Conclusion

This narrative review has attempted to examine the evolution of the literature on behavioral biases in investment along with four newer fields. These fields are financial advice through AI technology, increased focus by scholars and regulators on algorithmic transparency and explainability, regulation and ethics on the issue of fiduciary duty and investor protection in relation to algorithmic advice, and increased focus on financial and digital financial literacy as contextualizing factors. The four fields are, on their own, individually developed fields, and they have already begun, in some cases, to interface with each other as seen in the recent research on the relation between digital financial literacy and trust in fintech, concerns of confirmation bias influencing the reception of AI explanations, and regulatory guidelines that make it clear that fiduciary duty applies to algorithmic advice in any form.

Nonetheless, they have evolved almost in parallel and not through a constructive dialogue with each other. The combination of the two in a narrative framework would imply the presence of a well-elaborated but empirically untested theory that links behavioral biases, themselves varying depending on the characteristics of gender, age, and experience, to not only investment decisions but also the investors’ trust in the AI-powered tools. This trust is, in turn, influenced by the extent of AI system’s transparency and the credibility of the regulatory environment, whereas the presence of financial and digital financial literacy acts as a buffer moderating the intensity of these influences. In light of the increasing embedding of the use of AI in investing, growing calls of regulators for algorithmic transparency and greater activity related to exposing overstated AI capabilities, it is clear that filling these gaps in behavioral finance would become an important area of research in the future.

Disclosure Statement: No potential conflict of interest was reported by the authors**.**
Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

[1]. Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T
[2]. Aren, S., & Hamamci, H. N. (2020). Relationship between risk aversion, risky investment intention, investment choices: Impact of personality traits and emotion. Kybernetes, 49(11), 2651–2682. https://doi.org/10.1108/K-07-2019-0455
[3]. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
[4]. Baker, H. K., & Ricciardi, V. (2014). How biases affect investor behaviour. The European Financial Review, 10(7), 7–10.
[5]. Baker, T., & Dellaert, B. G. C. (2018). Regulating robo-advice across the financial services industry. Iowa Law Review, 103(2), 713–750.
[6]. Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. The Journal of

Finance, 55(2), 773–806. https://doi.org/10.1111/0022-1082.00226
[7]. Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics, 116(1), 261–292. https://doi.org/10.1162/003355301556400
[8]. Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In G. Constantinides, M. Harris, & R. Stulz (Eds.), Handbook of the Economics of Finance (Vol. 1, pp. 1053–1128). Elsevier.
[9]. Bashir, T., Azam, N., Butt, A. A., Javed, A., & Tanvir, A. (2013). Are behavioral biases influenced by demographic characteristics and personality traits? Evidence from Pakistan. European Scientific Journal, 9(29), 277–293.
[10]. Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. Review of General Psychology, 1(3), 311–320. https://doi.org/10.1037/1089-2680.1.3.311
[11]. Belanche, D., Casaló, L. V., & Flavián, C. (2019). Artificial intelligence in fintech: Understanding robo-advisor adoption among customers. Industrial Management & Data Systems, 119(7), 1411–1430.
[12]. Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2019). Service robot implementation: A theoretical framework and research agenda. Service Industries Journal, 40(3–4), 203–225. https://doi.org/10.1080/02642069.2019.1672666
[13]. Biais, B., Hilton, D., Mazurier, K., & Pouget, S. (2005). Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. Review of Economic Studies, 72(2), 287–312. https://doi.org/10.1111/0034-6527.00333
[14]. Bikhchandani, S., & Sharma, S. (2001). Herd behavior in financial markets. IMF Staff Papers, 47(3), 279–310.
[15]. Devenow, A., & Welch, I. (1996). Rational herding in financial economics. European Economic Review, 40(3–5), 603–615. https://doi.org/10.1016/0014-2921(95)00073-9
[16]. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033
[17]. European Parliament & Council of the European Union. (2024). Artificial Intelligence Act. Official Journal of the European Union.
[18]. European Securities and Markets Authority. (2023). Artificial intelligence in EU securities markets. ESMA Report.
[19]. Financial Conduct Authority. (2022). Artificial intelligence and machine learning in UK financial services. FCA Discussion Paper.
[20]. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
[21]. Glaser, M., Langer, T., & Weber, M. (2007). On the trend recognition and forecasting ability of professional traders. Decision Analysis, 4(4), 176–193.
[22]. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057
[23]. Goyal, K., Kumar, S., & Xiao, J. J. (2022). Behavioural biases in investment decision making: A systematic literature review. Qualitative Research in Financial Markets, 14(2), 275–305. https://doi.org/10.1108/QRFM-07-2021-0109
[24]. Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. Journal of Chiropractic Medicine, 5(3), 101–117. https://doi.org/10.1016/S0899-3467(07)60142-6
[25]. Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. MIS Quarterly, 23(4), 497–530. https://doi.org/10.2307/249487
[26]. Jung, D., Dorner, V., Weinhardt, C., & Pusmaz, H. (2018). Designing a robo-advisor for risk-averse, low-budget consumers. Electronic Markets, 28(3), 367–380. https://doi.org/10.1007/s12525-017-0279-9
[27]. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291. https://doi.org/10.2307/1914185
[28]. Kumar, S., & Goyal, N. (2015). Behavioural biases in investment decision making – A systematic literature review. Qualitative Research in Financial Markets, 7(1), 88–108. https://doi.org/10.1108/QRFM-07-2014-0022
[29]. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005
[30]. Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. Journal of Consumer Research, 46(4), 629–650. https://doi.org/10.1093/jcr/ucz013
[31]. Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. Journal of Economic Literature, 52(1), 5–44. https://doi.org/10.1257/jel.52.1.5
[32]. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21. https://doi.org/10.1177/2053951716679679
[33]. Morgan, P. J., Huang, B., & Trinh, L. Q. (2019). The need to promote digital financial literacy for the digital age. Asian Development Bank Institute Working Paper No. 1008.
[34]. Mousa, G. A., Elamir, E. A. H., & Abdelaziz, H. M. (2023). Behavioral biases and investment decisions in the Egyptian stock market: A review of recent evidence. Review of Behavioral Finance, 15(4), 512–529.
[35]. OECD. (2023). OECD/INFE toolkit for measuring financial literacy and financial inclusion 2023. OECD Publishing.
[36]. OECD. (2024). OECD consumer finance risk monitor 2024. OECD Publishing.
[37]. Pompian, M. M. (2012). Behavioral finance and investor types: Managing behavior to make better investment decisions (2nd ed.). Wiley.
[38]. Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5
[39]. Rastogi, S., Chaudhary, R., & Sharma, A. (2022). Digital financial literacy and financial well-being in emerging economies: Evidence from India. International Journal of Social Economics, 49(8), 1147–1163.
[40]. Ricciardi, V., & Simon, H. K. (2000). What is behavioral finance? Business, Education and Technology Journal, 2(2), 1–9.
[41]. Securities and Exchange Commission. (2017). Guidance update: Robo-advisers. U.S. Securities and Exchange Commission.
[42]. Securities and Exchange Commission. (2024). 2024 examination priorities. U.S. Securities and Exchange Commission.
[43]. Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. The Journal of Finance, 40(3), 777–790. https://doi.org/10.1111/j.1540-6261.1985.tb05002.x
[44]. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

[45]. Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356–367. https://doi.org/10.1177/1534484305278283

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

[47]. Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5(4), 297–323. https://doi.org/10.1007/BF00122574

[48]. Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly, 36(1), 157–178. https://doi.org/10.2307/41410412

[49]. Xiao, J. J., & Porto, N. (2022). Financial literacy and investment behaviors among young adults. Journal of Consumer Affairs, 56(2), 604–628.