

Research paper

Are Large Language Models Rational or Behavioral? A Comparative Analysis of Investor Behavior Interpretation

 Özkan ŞAHİN^{a,*}

^aDüzce University, Faculty of Business, Department of International Trade and Finance, Düzce, Türkiye.

*Corresponding author: ozkansahin@duzce.edu.tr

Article information:

Received: 02/06/2025, Revision: 20/07/2025, Accepted: 24/07/2025

DOI: 10.29130/dubited.1711955

ABSTRACT

This study aims to evaluate the ability of Large Language Model (LLM)-based AI applications to understand and interpret the fundamental theories of behavioral finance. In this context, the responses of five current LLM applications (ChatGPT 4o, Deepseek, Gemini 2.0 Flash, QwenChat 2.5 Max, and Copilot) were comparatively analyzed based on ten distinct scenarios involving behavioral biases and investment decision-making. The findings reveal how each model responds to behavioral concepts such as conceptual depth, psychological insight, strategic recommendation level, and originality. The results indicate that while the applications demonstrate successful analyses in certain cases, they also differ significantly in terms of data source diversity, contextual sensitivity, and algorithmic approaches. In particular, notable discrepancies were observed in explainability, consistency, and theory-based interpretive capacity. Ultimately, the study concludes that LLM systems have the potential to assess investment decisions not only through a rational framework but also from a behavioral perspective. Accordingly, the research provides both theoretical and practical contributions to the development of AI-based financial decision support systems.

Keywords: Behavioral Finance, Large Language Models (LLMs), Financial Decision-Making, Generative Artificial Intelligence.

I. INTRODUCTION

For many years, financial decision-making processes were explained within the framework of classical finance theory, which is grounded in the assumption of rational individuals and investors. However, since the 1970s, this paradigm has undergone a significant transformation with the emergence and advancement of behavioral finance approaches (Kahneman & Tversky, 1979). Contrary to classical financial theory, behavioral finance highlights that psychological and cognitive biases—rather than pure rationality—play a key role in investor behavior, and that markets are not entirely efficient or rational, challenging the Efficient Market Hypothesis. Numerous studies have repeatedly shown that, unlike the predictions of classical finance, investors in financial markets frequently exhibit behavioral tendencies such as loss aversion, herd behavior, representativeness bias, and overconfidence (Barberis & Thaler, 2003).

Another important development has been the rapid evolution of generative artificial intelligence systems, particularly Large Language Models (LLMs), which have initiated a transformative shift in individual investment decisions and financial decision support systems. Bommasani et al. (2021) underscore the transformative impact of foundation models across various sectors, especially in enhancing

productivity, efficiency, and decision-making processes. Although their study does not extensively examine the finance sector, it is asserted that the growing prevalence of foundation models across domains—including finance—holds the potential for a major transformation.

Moreover, LLMs are not only effective in generating text, but they also demonstrate high performance in complex reasoning tasks such as performing arithmetic operations, understanding concepts in context, and generating tailored recommendations. This indicates that LLMs have evolved beyond mere text generators into multifunctional cognitive tools (Brown et al., 2020).

Large Language Models are artificial neural network-based systems trained on massive text corpora and typically contain billions of parameters. These models acquire syntactic and semantic structures of language, enabling them to generate novel text, perform linguistic tasks, and conduct contextual analysis (Vaswani et al., 2017). LLMs are commonly built on autoregressive architectures or encoder-decoder frameworks (Devlin et al., 2019). The development of LLMs can be broadly categorized into three major phases:

- **Rule-Based Systems (1950s–1980s):** These early systems processed linguistic structures through manually coded grammatical rules (Chomsky, 1959) but their flexibility was highly limited.
- **Statistical NLP (1990s):** With the rise of data-driven approaches, language modeling evolved through n-gram-based methods, paving the way for applications such as machine translation (Jurafsky & Martin, 2009).
- **Deep Learning and LLMs (Post-2010):** With the advent of neural networks, particularly the transformer architecture (Vaswani et al., 2017), LLMs emerged. The success of models such as BERT (Devlin et al., 2019), GPT (Radford et al., 2018), and subsequently GPT-3 (Brown et al., 2020) marked a turning point in this progression.

At the core of LLMs lies the Transformer architecture, introduced in 2017 (Vaswani et al., 2017). This architecture utilizes an attention mechanism to dynamically evaluate the contextual relevance of each word in relation to all others. The Transformer offers key advantages such as parallelizability and the ability to process long-range dependencies in text. LLMs are typically trained through a two-stage process:

- (1) **Pre-Training:** Where the model learns general language patterns from massive amounts of unstructured data, and
- (2) **Fine-Tuning:** Where the model is further trained on smaller, labeled datasets for more task-specific accuracy (Howard & Ruder, 2018).

One of the most critical factors affecting the performance of LLMs is model size, typically measured by the number of parameters. Kaplan et al. (2020) demonstrated a logarithmic relationship between model capacity, dataset size, and training time. For instance, GPT-3 has 175 billion parameters (Brown et al., 2020). LLMs are now being applied across a wide range of tasks, including text generation and summarization (Kaplan et al., 2020), machine translation (Liu et al., 2020), question-answering systems (Kwiatkowski et al., 2019), code generation (Chen et al., 2021), and even in legal and ethical analysis (Bommasani et al., 2021).

Despite their powerful performance, LLMs face several fundamental limitations:

- **Bias and Discrimination:** These models can replicate systemic biases present in their training data (Bender et al., 2021).
- **Lack of Transparency:** Their "black-box" nature makes explainability a challenge (Lipton, 2018).
- **Energy Consumption:** Training large models incurs significant environmental costs (Strubell et al., 2020).
- **Hallucination of Information:** A common issue where models generate content that is factually incorrect or fabricated (Maynez et al., 2020).

The literature highlights several emerging directions for enhancing the effectiveness and ethical use of LLMs:

- Development of multimodal models
- Adoption of compact and efficient modeling techniques
- Design of human-aligned and explainable AI systems

- Governance within the framework of universal ethical principles (Floridi & Cowls, 2022).

LLMs have triggered a paradigmatic transformation in the field of natural language processing. However, both their technical limitations and broader societal impacts necessitate a more responsible and interdisciplinary approach. Future research is expected to be shaped around principles such as transparency, explainability, and ethics in order to enhance the accountability and reliability of LLMs.

As in many other fields, LLMs have rapidly begun to find a place in finance as decision support systems. Recent studies have highlighted the potential of Large Language Models (LLMs) in identifying behavioral biases, modeling investor psychology, and predicting irrational investor behavior within financial decision-making systems. These models have the capacity to detect cognitive biases in investors' decision-making processes and can contribute to the development of more rational strategies. Research shows that LLMs are capable of recognizing cognitive biases in investment decisions, and that these biases are measurably reflected in the model outputs (Zhou et al., 2024).

According to the Efficient Market Hypothesis, which is one of the core tenets of classical finance theory, investors are assumed to behave rationally by correctly processing all available information. However, the tendency of investors to prematurely close profitable positions while holding onto losing ones for extended periods is regarded as a manifestation of irrational behavior. Behavioral finance addresses these phenomena using psychological concepts. Notably, (Shefrin & Statman, 1985) introduced the term "disposition effect" to describe investors' tendency to sell winning positions too early and to retain losing positions for too long. They explained this behavior through the frameworks of prospect theory and mental accounting.

Studies suggest that investors tend to realize gains early to secure profits, while losses are held in the hope of a price reversal. This behavior indicates that investor decisions may deviate from rational models and that such deviations are rooted in systematic, psychologically driven biases (Weidinger et al., 2021). Understanding the extent to which LLMs can comprehend these human-specific cognitive and emotional patterns is of critical importance for both academic research and practical financial applications.

In the literature, the potential impacts of Large Language Models (LLMs) have been examined from multiple perspectives, including ethical principles (Floridi & Cowls, 2022), reliability issues in information generation (Maynez et al., 2020), and broader societal effects (Weidinger et al., 2021). Today, LLMs are increasingly being used across a wide range of domains, including education, law, media analysis, and medicine (Bommasani et al., 2021). As in these fields, the active integration of LLM-based generative AI systems into the field of finance is also becoming increasingly apparent. In particular, models such as ChatGPT, DeepSeek, Gemini, Qwen 2.5-Max, and Copilot are frequently used by investors for tasks such as fundamental and technical analysis, portfolio management, macroeconomic data interpretation, and financial news evaluation.

LLMs have started to assume a consultative role in investment decisions by facilitating access to information—especially for individuals with low levels of financial literacy. However, the extent to which LLMs can respond effectively to cognitive biases that emerge in investor behavior, such as loss aversion, representativeness bias, and herd behavior, has not yet been thoroughly and systematically analyzed (Zhou et al., 2024). The ability of AI-assisted systems to properly guide investor behavior depends not only on technical capabilities, but also on their ability to understand behavioral finance theories, demonstrate contextual sensitivity, and provide decision support within an ethical framework.

In this context, the primary objective of this study is to conduct a comparative analysis of the conceptual consistency, explanatory depth, and bias awareness level of several state-of-the-art LLM-based generative AI applications—namely ChatGPT 4o, Deepseek, Gemini 2.0 Flash, QwenChat 2.5 Max, and Copilot—when confronted with ten original behavioral finance scenarios related to investor psychology. The study systematically evaluates the extent to which LLMs recognize the theoretical frameworks within behavioral finance literature, which concepts they rely on when offering decision suggestions, and how rational their responses are within the constructed scenarios.

In this regard, the study offers a pioneering framework that examines behavioral finance theories through the lens of AI-generated outputs. It not only focuses on technical accuracy but also aims to assess LLMs in terms of conceptual integrity and psychological sensitivity. Thus, the research contributes a novel

dimension to the behavioral finance literature and stands out as one of the first studies to explore, from a holistic perspective, the potential influence of LLM-based AI systems on investor behavior and their applicability in financial advisory processes.

Furthermore, the current literature contains very few comprehensive analyses evaluating LLM-based generative AI applications within the context of behavioral finance. Therefore, this study offers an original contribution by: uncovering how well behavioral finance theories are understood by AI systems, and comparing the decision-making approaches of various LLM applications in terms of their conceptual depth and sensitivity to cognitive biases.

In particular, the comparative analysis of AI responses to ten scenario-based investor behavior models is, to the best of our knowledge, the first to be conducted using such a structured methodological approach. The main contribution of the study comes from presenting a pioneering evaluation that comparatively analyzes the attitudes of large language model (LLM) based generative artificial intelligence systems towards behavioral finance scenarios within the framework of multidimensional criteria such as conceptual depth, psychological insight, strategic recommendation level and conceptual originality. Secondly, the extent to which LLMs can recognize behavioral biases and their capacity to develop solution-oriented suggestions through original scenarios related to investor psychology is systematically coded. Thirdly, the study produces guiding suggestions for ethical, contextual and expertise-based artificial intelligence systems to be developed in the future by revealing the axes on which different LLMs differ in financial decision support processes. In this respect, the research provides original theoretical and methodological contributions to the field of behavioral finance and artificial intelligence interaction.

II. MATERIAL AND METHOD

This study analyzes how generative artificial intelligence applications based on Large Language Models (LLMs) respond to investor decision-making scenarios derived from behavioral finance theory.

Within the scope of the research, ten scenarios were developed, each structured around cognitive and emotional investor biases commonly discussed in the behavioral finance literature. Each scenario presents a narrative that reflects a specific behavioral bias within an investment decision-making context. These scenarios were presented to five widely used and up-to-date LLM applications—ChatGPT 4o, DeepSeek, Gemini 2.0 Flash, QwenChat 2.5 Max, and Copilot—and independent responses were collected from each model.

The responses were evaluated across four analytical dimensions, as described in Table 1:

- **Conceptual Depth:** This dimension assesses the extent to which the LLM refers to the behavioral finance literature and the technical terminology it employs in its responses. Particular attention was paid to the diversity of theoretical frameworks used, the degree of alignment with established literature, and the contextual appropriateness of terminology.
- **Psychological Insight:** This dimension evaluates how well the LLM can explain investor behaviors presented in the scenarios through psychological foundations. It reflects the model's awareness of individual investors' emotional, cognitive, and social biases and the degree of sensitivity to behavioral psychology exhibited in its answers.
- **Strategic Solution Orientation:** This dimension examines the clarity and explainability of the model's recommendations, its ability to propose strategies to mitigate behavioral biases, and the overall coherence of its guidance. The model's potential not only to diagnose issues but also to guide investors toward more rational decisions is assessed within this axis.
- **Conceptual Originality:** The final dimension evaluates the originality of the LLM's thinking in its responses—specifically whether the model forms novel and meaningful conceptual links that go beyond common patterns and whether it demonstrates the ability to think outside conventional frameworks.

Each of these four dimensions was independently scored for each of the five LLM applications based on their responses to the ten constructed scenarios. The results were then comparatively tabulated and analyzed. Through this approach, the study offers a holistic evaluation of each model's capacity to generate knowledge about investor behavior, its theoretical alignment with behavioral psychology, and the nature of the strategic guidance it provides.

The evaluation framework assesses participants' abilities across four key dimensions: Conceptual Depth, Psychological Insight, Strategic Orientation, and Conceptual Originality. Each dimension is rated on a four-level scale: Low (1), Medium (2), High (3), and Very High (4).

Conceptual Depth examines how effectively participants use behavioral finance concepts. Higher scores indicate the use of multiple concepts, establishing relationships among them, and referencing the literature with logical consistency.

Psychological Insight evaluates attention to emotional and cognitive aspects of investor behavior. Scores improve as explanations incorporate deeper psychological dynamics, such as cognitive dissonance and emotional strategies.

Strategic Orientation measures solution-focused thinking. Higher levels reflect structured, actionable strategies grounded in financial reasoning and conceptual understanding.

Conceptual Originality assesses the novelty of interpretations. Scores increase when participants move beyond conventional ideas, offering creative and innovative conceptualizations.

The framework emphasizes a progression from basic to sophisticated analysis, encouraging a blend of theoretical knowledge, psychological awareness, strategic thinking, and originality.

Table 1. Evaluation scale.

Evaluation Dimension	Low (1)	Med. (2)	High (3)	Çok Yüksek (4)
Conceptual Depth <i>This criterion evaluates how many distinct and accurate concepts from the behavioral finance literature the participant employs, and how effectively these concepts are applied to the given scenarios.</i>	Uses 1–2 basic behavioral finance concepts (e.g., only “Loss Aversion”)	Uses 2–3 concepts with moderate-level explanation	Uses at least 3 concepts with relationships established between them	Uses 4+ concepts, relationships among them articulated, with literature references and logical consistency
Psychological Insight (Emotional and Cognitive Depth) This criterion evaluates the extent to which the participant draws attention to the investor’s internal processes, including emotions, biases, and psychological conflicts.	Little to no reference to emotions or internal processes	Superficial acknowledgment of psychological effects on decisions	Explanations reflecting cognitive dissonance, emotional reactions, and awareness	In-depth analysis using concepts like defense mechanisms, emotional strategies, and cognitive dissonance
Strategic Orientation (Solution-Focused Thinking) This criterion assesses how well the participant not only identifies the problem but also models potential solutions in a structured and actionable manner.	Vague or generic suggestions (e.g., “be more careful”)	1–2 specific suggestions (e.g., “use a stop-loss order”)	Coherent plan supported by financial strategies (e.g., “50/30/20 budgeting”)	Multiple strategic suggestions grounded in conceptual reasoning

Table 1 (cont). Evaluation scale.

Conceptual Originality	No new perspectives introduced	Accurate but conventional use of standard concepts	Innovative interpretation of concepts within context	Novel conceptualizations (e.g., “hope strategy,” “emotional relief cycle”), creative approaches
This criterion evaluates whether the participant produces original interpretations that go beyond conventional frameworks and whether novel or uncommon conceptual connections are made, as opposed to merely repeating well-established ideas from the literature.				

III. RESULTS AND DISCUSSIONS

The responses of LLM-based generative AI applications to each of the investor behavior scenarios are summarized below. The theories, concepts, and models derived from the full analysis of all responses are presented in the tables in the subsequent sections.

A. Scenario 1 – Case Pattern

Zeynep is a novice investor who has recently started trading in the stock market. The share price of company ABC in her portfolio has declined by 15% over the past two months. In response to this loss, she refrains from selling the shares, thinking, “I can recover my losses if I just wait a bit longer.” Meanwhile, she realizes a 10% profit on another investment, which makes her feel satisfied, and she immediately sells the position, saying, “At least I made money here.” Zeynep’s behavior reflects an unwillingness to accept losses and a tendency to prematurely realize gains.

A.1. Scenario 1 – LLM Responses

A.1.1. ChatGPT 4o

Zeynep’s behavior may provide short-term emotional satisfaction, but it could negatively affect her long-term investment performance. Closing profitable positions too early may result in missed potential returns. Meanwhile, holding onto losing positions due to emotional attachment may create a “drag effect,” pulling down the overall performance of the portfolio. This behavior reflects a deviation from rationality and emotionally driven decision-making. Therefore, it is advisable for investors to rely more on objective analyses rather than emotions when making decisions.

A.1.2. DeepSeek

Zeynep’s actions are the opposite of what is expected from a rational investor. A rational investor would implement a stop-loss strategy for risk management and avoid impatience in profit-taking, as selling too early when a trend continues can result in opportunity costs. The thought “I can recover my losses” is an example of the sunk cost fallacy. Past losses do not determine future performance, but Zeynep seems emotionally attached to her investment.

A.1.3. Gemini 2.0 Flash

Zeynep’s decisions illustrate the significant impact of emotional factors in investment behavior. A rational investor should evaluate each investment individually and focus on its future potential. Past gains or losses should not solely dictate future decisions. Holding onto the losing stock may be viewed as an averaging down strategy, which carries risk if the company’s fundamentals have deteriorated or the industry outlook is negative. Accepting the loss and reallocating capital to better opportunities could be a wiser approach. Selling a profitable stock too early may result in missed opportunities for larger gains. Investors should have a clear strategy aligned with their investment goals and risk tolerance. Selling immediately after a small profit might hinder long-term objectives.

A.1.4. Qwen 2.5-Max

Zeynep's behavior clearly reflects the influence of emotional and cognitive factors on investor decisions. Financial decisions, especially in volatile environments like the stock market, are shaped not only by mathematical calculations but also by psychological dynamics. Her reluctance to accept losses and eagerness to realize gains are indicators of how emotional intelligence impacts her financial choices. Her hope to recover the 15% loss in ABC shares may ignore the company's actual performance or broader market conditions. If ABC is fundamentally weak or the sector trends are negative, her decision could lead to greater losses. The 10% profit on another investment brought her short-term satisfaction, but if the asset had higher future potential, the early exit may have imposed an opportunity cost. This illustrates how the pursuit of short-term gratification can undermine long-term returns. Emotional reactions such as these may adversely affect her portfolio performance. Hence, managing emotional decision-making processes and adopting a more rational approach is crucial.

A.1.5. Copilot

Zeynep's behavior may negatively affect her long-term portfolio performance. Selling profitable investments too early deprives her of potential gains, while holding on to losing investments can result in larger losses. Such patterns of behavior may prevent investors from making rational decisions.

B. Scenario 2 – Case Pattern

Ahmet works at a company and receives a regular monthly salary. At the end of the year, due to his performance, he receives a bonus of 10,000 TL. He considers this bonus as “extra” money and decides to allocate it for a summer vacation. However, during the same period, Ahmet has accumulated high-interest credit card debt. While his family advises him to use the bonus to pay off the debt, Ahmet insists, “My salary is for paying bills, but this is my fun money,” and spends it on the vacation. In Ahmet's mind, the bonus constitutes a separate mental budget.

B.1. Scenario 2 – LLM Responses;

B.1.1. ChatGPT 4o

Ahmet's mental categorization of his bonus may provide short-term satisfaction, but it is detrimental to his financial health in the long run. In cases involving high-interest debt, the rational choice is to prioritize debt repayment. The psychological dynamics behind Ahmet's decision may include the emotional reward associated with unexpected income, such as a bonus, which encourages spending on leisure rather than addressing painful financial obligations. This is a common example of short-term gratification taking precedence over long-term goals—even when individuals are aware of the more logical course of action.

B.1.2. DeepSeek

Ahmet's decision is financially irrational: Delaying the repayment of high-interest debt significantly increases total borrowing costs. For example, a 10,000 TL debt with 30% interest results in 3,000 TL in additional cost annually. The perception of the bonus as “extra income” disrupts budgetary discipline. All income—including bonuses—should be directed first toward reducing debt. The idea of “rewarding oneself” stems from a desire for instant gratification.

B.1.3. Gemini 2.0 Flash

Ahmet's behavior illustrates how emotional and psychological factors can override rational economic principles. Although it is understandable that the bonus is seen as “extra” and earmarked for a different purpose, this is not an optimal decision given his financial situation. Spending liquid resources on leisure rather than paying off high-interest debt is not financially sound. Interest accumulation translates into greater future payments. Ahmet is choosing short-term emotional satisfaction (a vacation) over long-term financial stability (debt elimination). His rejection of his family's advice reveals a strong commitment to mental accounting—he views the bonus as personal and reserved for a specific purpose, making him less receptive to rational external suggestions.

B.1.4. Qwen 2.5-Max

Ahmet's decision illustrates that financial choices are shaped not only by calculations but also by psychological tendencies. By treating the bonus as "extra," he perceives it as less valuable. This behavior can be analyzed in the following ways: People tend to spend unexpected or bonus income (e.g., performance rewards, gifts) more freely, often ignoring its real economic value. Yet this income is still part of overall earnings and should be used in the most rational way possible. Ahmet's disregard for high-interest debt demonstrates how the pursuit of short-term pleasure can undermine long-term gains. The compounding interest on his unpaid debt will worsen his financial situation over time. While using the bonus for a vacation may provide temporary happiness, such emotionally driven decisions are ultimately detrimental to long-term financial planning. Prioritizing debt repayment would reduce interest burdens and promote greater savings in the future.

B.1.5. Copilot

Ahmet's behavior may lead to a financially unsustainable situation. The cost of maintaining high-interest credit card debt may far outweigh the temporary joy gained from a vacation. However, the desire to allocate funds for leisure can also be understood as an attempt to enhance personal well-being and quality of life. Ahmet's approach reflects both mental accounting and a tendency toward short-term gratification.

C. Scenario 3 – Case Pattern

Bora enters the cryptocurrency market and makes several successful trades within his first six months. In particular, the rapid surge of altcoins during a bull market generates significant profits for him. As a result, Bora begins to believe he can predict every market movement in advance. He starts using leverage and invests in high-risk tokens. Convinced of his own competence—"I understand the market; it can't shake me"—he eventually experiences a sharp market correction and loses 60% of his portfolio. Nevertheless, he attributes the loss to external factors rather than his own decisions.

C.1. Scenario 3 – LLM Responses;

C.1.1. ChatGPT 4o

Bora's psychological fallacies during this period hindered his ability to make rational investment decisions. Early successes often reinforce investor confidence while creating a dangerous illusion: "I know what I'm doing." This belief can lead to risky behaviors such as using leverage and neglecting risk management entirely. Such behavior ultimately impairs overall portfolio performance.

C.1.2. DeepSeek

Bora's behavior reflects common irrational patterns seen during speculative bubbles. Attributing profits to personal skill and losses to external conditions prevents learning. His use of leverage and investment in risky tokens aligns with gambling-like behavior. There is also a false causality bias—profits made during a bull market do not prove the soundness of an investment strategy.

C.1.3. Gemini 2.0 Flash

Bora's experience offers critical lessons, particularly for newcomers in highly volatile markets. Fast gains can lead to unrealistic confidence and an increased appetite for risk. However, markets are not always predictable, and excessive risk-taking can lead to substantial losses. While leveraged trades and risky altcoins offer high return potential, they also carry significant downside risk. Successful investors manage risks prudently and diversify their portfolios. Bull markets can enable profit even without deep market knowledge or skill, but do not test real competence. Bear markets and sharp corrections can be painful for overconfident investors. Bora's loss should be seen as a learning opportunity. Acknowledging mistakes and embracing the complexity of markets can help improve future decision-making.

C.1.4. Qwen 2.5-Max

Bora's behavior clearly demonstrates that financial decisions are influenced not only by logic but also by emotional and psychological factors. Early wins gave him a false sense of confidence. However, crypto markets are highly volatile, and short-term success is not a guarantee of long-term performance. His initial gains distorted his perception of risk. People often interpret gains as universal competence, which leads to

riskier strategies. Despite suffering heavy losses, Bora attributes the outcome to external conditions rather than acknowledging his own mistakes—reflecting an avoidance of personal responsibility. This mindset may cause repeated errors in the future. His actions—especially the use of leverage and risky tokens—have put his portfolio at serious risk. Managing emotional decision-making and adopting a disciplined investment strategy are essential.

C.1.5. Copilot

Bora's overconfidence bias posed serious risks in a highly volatile environment like the crypto market. His use of leverage and risky tokens significantly weakened his portfolio. Furthermore, his refusal to take responsibility and blame external factors hinders his ability to improve his investment strategies. This may trap him in a cycle of repeating the same mistakes.

D. Scenario 4 – Case Pattern

Cemre is a long-time stock investor with two major positions in her portfolio. The first stock has appreciated by 5%, and she immediately realizes this small gain. The second stock has declined by 30%, but Cemre refuses to sell it. She continues to hold the losing stock, thinking, "It has already dropped a lot—if I wait, it will bounce back." Over time, despite the prolonged presence of the losing position in her portfolio, Cemre begins to experience a sense of psychological relief.

D.1. Scenario 4 – LLM Responses:

D.1.1. ChatGPT 4o

Cemre's behavior is common but financially harmful. A rational investor should base decisions on future expectations rather than past costs. However, she holds the losing position due to the sunk cost fallacy, and even derives a form of psychological comfort from keeping the stock. This reflects a dangerous scenario where emotions take precedence over information in investment decisions.

D.1.2. DeepSeek

Cemre's actions may lead to various financial problems. Holding a losing stock ties up capital and creates opportunity costs, as these funds could be reallocated to better-performing investments. Additionally, risk concentration increases, distorting the portfolio's overall risk profile. The "psychological relief" she experiences is an illusion and does not reflect the actual financial reality.

D.1.3. Gemini 2.0 Flash

Cemre's strategy is risky and potentially harmful from a rational investing perspective. Every investment decision should be based on future outlooks and opportunities—not on past gains or losses. Taking small profits too early can result in missed opportunities for greater gains. A successful strategy should allow winning positions to grow and benefit from long-term trends. Holding onto a losing stock solely in the hope that it will recover to the original cost is known as the "hope strategy," which rarely yields good outcomes. If there are fundamental or sector-specific issues, further losses are possible. Although Cemre may experience psychological relief by holding the losing stock, this does not mean the problem is resolved. It merely indicates cognitive adjustment, not financial recovery.

D.1.4. Qwen 2.5-Max

Cemre's behavior clearly demonstrates that financial decisions are not solely based on logic but are also shaped by emotional and psychological dynamics. Realizing a small 5% gain is a sign of the Disposition Effect, reflecting a preference for short-term satisfaction. Her reluctance to sell a stock with a 30% loss is a classic manifestation of Loss Aversion, where investors avoid the discomfort of accepting losses. The belief that "if I wait, it will recover" is often unrealistic and can lead to greater financial damage. Using the losing position as a source of psychological comfort illustrates her emotional decision-making tendencies. However, such behavior can harm both the performance of her portfolio and her psychological well-being. Holding onto unprofitable positions may expose the portfolio to greater risks. Therefore, controlling emotional impulses and adopting a more disciplined investment strategy is crucial.

D.1.5. Copilot

Cemre's behavior may negatively impact her portfolio performance in the long run. Selling profitable positions too early may cause her to miss future gains, while holding onto losing positions may result in larger losses. Her sense of psychological relief while holding the losing stock can be explained through cognitive dissonance theory, which suggests that individuals tend to rationalize their decisions to reduce internal conflict. In this case, she justifies the loss by convincing herself that holding the stock is the right choice.

E. Scenario 5 – Case Pattern

Mert is an investor who frequently follows social media while making investment decisions. On platforms like Twitter and Telegram, he repeatedly sees messages such as “XYZ stock is guaranteed to rise.” Without conducting his own research, he joins the trend and buys shares of XYZ. Within a few days, the stock rises briefly, but soon after, it drops sharply. Mert defends his behavior by saying, “Everyone was buying it, so did I—I didn’t make a mistake.”

E.1. Scenario 5 – LLM Responses;

E.1.1. ChatGPT 4o

Mert's behavior illustrates how fragile financial decisions become when based on perception rather than information. Social media claims like “guaranteed to rise” are often exaggerated or manipulative. Making informed, research-based investment decisions is essential for long-term success. Mert's vulnerability to speculation reflects a lack of critical analysis. Furthermore, his external attribution—claiming “I didn’t make a mistake”—blocks opportunities for learning from errors.

E.1.2. DeepSeek

Mert's approach is a classic example of participation in a speculative bubble. Twitter and Telegram groups are prone to “pump and dump” schemes. While the stock initially rose, the movement lacked fundamental support and was unsustainable. His justification—“everyone was buying”—shows an avoidance of individual responsibility. This irrational behavior exposes investors to high risks. The absence of original research bypasses fundamental analysis and is often driven by emotionally fueled reactions.

E.1.3. Gemini 2.0 Flash

Mert's investment approach is particularly dangerous, given the complexity and risk of financial markets. While social media can provide valuable community interactions, it should not be treated as a primary or reliable source for investment decisions. Sound investing requires thorough research, including financial statements, industry trends, and management evaluations. Recommendations spread via social media are often manipulative or self-serving. Fraudulent schemes, such as pump-and-dump, can spread rapidly online. Investment decisions should be based on rational analysis and aligned with one's personal financial goals—not driven by emotional reactions or crowd behavior.

E.1.4. Qwen 2.5-Max

Mert's behavior clearly demonstrates that financial decisions are influenced not only by rational calculations but also by social and psychological factors. He followed the crowd after seeing messages claiming a guaranteed rise in XYZ stock. People often imitate others in uncertain situations, which can lead to speculative bubbles and subsequent crashes. By investing without independent research, Mert ignored the company's fundamentals and broader market dynamics—one of the most common investor mistakes. His refusal to acknowledge the mistake and his defense—“everyone else was doing it”—represent a psychological defense mechanism. This avoidance of responsibility may result in repeated errors, exposing his portfolio to significant risk. Emotional awareness and a disciplined investment strategy are crucial to avoiding such pitfalls.

E.1.5. Copilot

Mert's behavior undermines the rationality of his financial decisions. Herd behavior contributes to market bubbles and can lead to substantial losses for individual investors. Additionally, external attribution

bias impairs Mert's ability to learn from mistakes, increasing the likelihood of repeating them. This scenario clearly highlights the influence of investor psychology on decision-making processes.

F. Scenario 6 – Case Pattern

Ayşe purchases shares of ABC company at a price of 150 TL. The stock price later declines to 110 TL. Believing that “this stock was already worth 150 TL, so the drop is temporary,” Ayşe decides not to sell. For her, the reference point remains the initial price of 150 TL, and she continues to evaluate the stock based on that figure. Even when the price falls to 90 TL, Ayşe maintains her original valuation as the benchmark.

F.1. Scenario 6 – LLM Responses

F.1.1. ChatGPT 4o

Ayşe's behavior reflects a common but financially harmful mistake made by many retail investors—anchoring to past prices. In financial markets, value should be based not on historical prices but on future expectations and actual company performance. Ayşe's perception of 150 TL as a fixed and “correct” value distances her from rational risk assessment. Moreover, holding a losing position based on “psychological accounting” results in opportunity costs and capital inefficiency.

F.1.2. DeepSeek

Ayşe's approach entails several financial risks. A decline from 150 TL could signal a deterioration in the company's fundamentals—signals she appears to ignore. Capital tied up in a losing position could have been allocated to more promising investments. Her belief that the stock will “return to its old price” reflects unrealistic optimism. This behavior impairs adaptability to changing market conditions, undermines stop-loss discipline, and erodes the effectiveness of rational portfolio management.

F.1.3. Gemini 2.0 Flash

Ayşe's behavior demonstrates how past purchase prices and emotional attachment can unduly influence investment decisions. A rational investor should focus on future potential rather than allow previous purchase prices to dictate choices. The price paid in the past does not affect future returns. Investment decisions should be based on current market conditions, company fundamentals, and future outlook. Her emotional anchoring to 150 TL prevents objective reassessment. The market environment may have changed, and the stock may never return to that level. By holding a loss-making asset, Ayşe forfeits better investment opportunities. Tied-up capital could be redirected to more profitable alternatives.

F.1.4. Qwen 2.5-Max

Ayşe's behavior illustrates that financial decisions are not solely based on logical calculations but are also heavily influenced by emotional and psychological factors. She continuously uses the 150 TL purchase price as her reference point. However, markets are dynamic, and past prices are not indicative of future performance. Her reluctance to sell suggests loss aversion and emotional attachment. Many investors avoid selling losing investments to escape the discomfort of acknowledging losses. Her belief that “the decline is temporary” reflects hope-based thinking, often rooted in unrealistic expectations. Such behavior can lead to both financial and emotional harm. Ignoring the actual value of the stock may expose her portfolio to greater risk. Managing emotional impulses and adopting a more disciplined strategy is essential.

F.1.5. Copilot

Ayşe's actions may negatively impact her portfolio's long-term performance. The anchoring effect can lead investors to make irrational decisions and misinterpret market conditions. Loss aversion, by encouraging her to hold onto losing positions, increases the likelihood of deeper losses. Using the purchase price as a reference point undermines rational investment strategy.

G. Scenario 7 – Case Pattern

Deniz strongly believes that the wind energy sector is poised for a major breakthrough. Before investing, she only reads news articles and reports that support the sector. She dismisses any critical analyses as “incorrect or exaggerated.” After making her investment, she continues to follow only positive developments and remains closed off to risks or negative information.

G.1. Scenario 7 – LLM Responses;

G.1.1. ChatGPT 4o

Deniz's behavior exemplifies a dangerous form of selective information bias for investors. Belief in a sector or stock is not sufficient for sound investment decisions. In fact, analyzing opposing viewpoints and objectively evaluating risks is a cornerstone of successful investing. Relying solely on "good news" can lead to ignoring potential threats and weakens risk management. In investment, emotional attachment should be replaced by critical thinking and balance.

G.1.2. DeepSeek

Deniz's approach introduces significant risks to her investment decisions. A strategy based solely on positive information fosters unrealistic confidence. She neglects critical aspects such as regulatory challenges, technological limitations, and market competition in the wind energy sector. Relying on like-minded sources reduces her objectivity. Financially irrational, this behavior leads to: (1) concentrated portfolios (lack of diversification), (2) vulnerability to unexpected shocks, and (3) missed learning opportunities due to the absence of a critical perspective.

G.1.3. Gemini 2.0 Flash

Deniz's investment strategy prevents objective and balanced evaluation, leading her to ignore potential risks and downside scenarios. A successful investment decision requires weighing both the sector's potential and its challenges. Focusing exclusively on supportive information results in a skewed and potentially misleading understanding of the market. Critical analyses provide insights into possible issues, enabling more informed decisions. Investment decisions should be based on rational analysis—not emotional conviction. Even sectors with great potential may be subject to competition, regulatory changes, and economic risks. Ignoring risks leaves portfolios vulnerable to potential losses. Effective risk management involves anticipating negative outcomes and taking proactive measures.

G.1.4. Qwen 2.5-Max

Deniz's behavior clearly illustrates that financial decisions are shaped not only by rational calculations but also by emotional and psychological influences. She actively seeks information that confirms her belief and avoids dissenting views. This lack of objectivity compromises the realism of her investment decisions. Her emotional connection to the wind energy sector leads her to ignore risks—exposing her portfolio to potential harm. Investors often avoid acknowledging risks when emotionally attached to an asset. This kind of bias can result in both financial and psychological losses. Ignoring risks especially makes her vulnerable to unexpected developments. Therefore, controlling emotional decision-making and adopting a more disciplined strategy is essential.

G.1.5. Copilot

Deniz's behavior undermines the rationality of her investment decisions. Confirmation bias leads investors to under-assess risks and overlook potential threats, which can be particularly dangerous in volatile sectors. By focusing only on positive developments, Deniz loses touch with market realities and creates imbalanced investment strategies.

H. Scenario 8 – Case Pattern

Ali is frequently exposed to statements within his social circle such as "NFTs are the new form of art" and "blockchain is revolutionizing the art world." He also regularly encounters similar narratives on social media. Without conducting any analysis of NFT projects, he purchases several NFTs simply because he believes in the narrative. When the market crashes, he suffers major losses, yet continues to believe for a time that "the narrative was right."

H.1. Scenario 8 – LLM Responses

H.1.1. ChatGPT 4o

Ali's fundamental mistake was basing his investment decisions solely on narratives and popular perception. In innovative but highly volatile markets like NFTs, individual investors must avoid making decisions without proper fundamental analysis. Emotional or social pressures should not drive investment behavior. Regardless of how compelling a narrative may seem, if it is disconnected from economic reality, it can mislead investors and result in significant losses.

H.1.2. DeepSeek

Ali's approach reflects key characteristics of speculative bubbles. Influenced by his social circle, he experienced FOMO (fear of missing out) and acted hastily. Instead of evaluating the artistic or technological value of NFTs, he focused exclusively on the narrative. His continued belief in the narrative even after the market crash demonstrates resistance to accepting losses. Financially risky, this behavior: (1) makes the strategy susceptible to manipulation, (2) results in large losses post-bubble, and (3) inhibits learning and critical reflection.

H.1.3. Gemini 2.0 Flash

Ali's investment behavior highlights the difficulty of making rational decisions in hype-driven markets. Powerful narratives and peer influence can discourage fundamental analysis and expose investors to substantial risks. Every investment decision should be based on a thorough assessment of value, potential, and risk. In the case of NFTs, this includes evaluating the project team, technology, use cases, community support, and long-term sustainability. While narratives can make an investment opportunity seem urgent or irresistible, they may distort reality and contribute to speculative bubbles. Social media and peer groups can significantly influence decisions, but such information must be critically evaluated, and independent research is essential.

H.1.4. Qwen 2.5-Max

Ali's behavior clearly shows that financial decisions are driven not only by rational calculation but also by social and psychological dynamics. He invested in NFTs based on emotionally appealing narratives without conducting any fundamental analysis. Such behavior contributes to the formation of speculative bubbles and is typically unsustainable in the long run. Relying on information from social media, which often contains speculative or emotionally charged content, led to misguided confidence. His insistence that "the narrative was right" even after suffering losses is a form of emotional self-justification. People tend to reject personal failure by attributing it to external circumstances. This pattern of behavior can have both financial and psychological costs. Investing without analysis exposes one's portfolio to higher risk. Emotional discipline and a more structured strategy are essential.

H.1.5. Copilot

Ali's decision-making is negatively impacted by narrative bias, which causes investors to ignore risks and make irrational decisions based on compelling stories. Additionally, social proof bias leads individuals to follow the majority without conducting their own analysis. His tendency to justify losses by continuing to believe in the narrative illustrates external attribution bias—the inclination to blame outside factors rather than accept personal responsibility.

I. Scenario 9 – Case Pattern:

Can learns that a brand he has loved since childhood—one he associates with positive emotions through advertisements—is going public. Without reviewing the company's financials, he decides to buy its shares. His investment decision is heavily influenced by his personal affinity and emotional attachment to the brand.

I.1. Scenario 9 – LLM Responses

I.1.1. ChatGPT 4o

Can's critical mistake is treating stock purchasing not as an investment, but as a personal preference. Liking a brand is not a valid financial reason to buy its stock. Fundamental indicators such as profitability, debt levels, market position, and competitive strength should be assessed before investing in an IPO. Emotionally driven investors often romanticize the stock's value, ignoring its actual fundamentals.

I.1.2. DeepSeek

Can's behavior reflects the classic risks of emotional investing. Ignoring key financial metrics such as P/E ratio and debt levels exposes him to potential losses. Investing based on emotional attachment to a single brand violates the principle of diversification. The belief that "a brand I like will perform well" hinders objective judgment. Economically, this is problematic because brand affection does not necessarily correlate with company performance. It also increases the risk of post-IPO decline and may cause cognitive dissonance, where the investor holds onto losing positions out of emotional loyalty.

I.1.3. Gemini 2.0 Flash

Can's approach highlights how emotional factors can lead to irrational financial decisions. Successful investing should be based on objective data and analysis, not personal sentiment. Evaluating an IPO requires a careful review of the company's income statement, balance sheet, cash flow, profitability ratios, growth outlook, sector conditions, and competition. Emotional investments may impair risk assessment and result in substantial losses. Liking a brand does not guarantee that it is a good investment. Avoiding personal biases and maintaining a rational outlook is critical in investment decisions.

I.1.4. Qwen 2.5-Max

Can's behavior illustrates that financial decisions are significantly influenced by emotional and psychological factors. His decision to invest is rooted in emotional attachment to the brand, not financial analysis. Emotional investments often lead to unrealistic expectations and increase exposure to unforeseen risks. Investing without examining financial performance or profitability ratios is speculative and may lead to losses. People tend to overlook the real market performance of brands they feel connected to. Can's behavior may have both financial and psychological consequences. Managing emotional biases and adopting a disciplined investment approach is essential.

I.1.5. Copilot

Can's investment decision reflects impaired rationality due to emotional attachment and familiarity bias. These biases prevent investors from properly evaluating risks and lead to irrational choices. Investing without reviewing financial statements or profitability metrics can result in long-term losses. This case clearly demonstrates how investor psychology influences decision-making in financial markets.

J. Scenario 10 – Case Pattern

Serkan has recently been watching numerous news reports about ongoing wars on television. Influenced by these reports, he becomes convinced that defense industry stocks are guaranteed to perform well in the long run. Without analyzing the financial health of individual companies or sector trends, he makes investment decisions solely based on the information frequently presented in the media.

J.1. Scenario 10 – LLM Responses

J.1.1. ChatGPT 4o

Serkan's investment approach is extremely risky and superficial. News media often present emotionally charged, speculative content with short-term impact. Sectors like defense operate with long-term dynamics that require investors to examine company-level performance, innovation capacity, supply chains, and government contracts. By ignoring structural analysis and relying only on media frequency and emotional cues, Serkan demonstrates irrational behavior, making himself vulnerable to speculative market bubbles.

J.1.2. DeepSeek

Serkan's behavior represents a textbook case of media-driven investment bias. He neglected critical data such as debt-to-equity ratios or R&D investment figures. The volatile nature of defense contracts and political risk were also overlooked. Treating the entire defense sector as uniformly profitable ignores firm-level variability. Financially problematic, this behavior: (1) leaves the investor open to media manipulation, (2) fosters an overemphasis on temporary trends, and (3) increases the risk of panic selling during crises.

J.1.3. Gemini 2.0 Flash

Serkan's investment decision highlights the influence of media narratives and emotional responses in financial decision-making. Media attention can draw investors toward certain themes, but that interest doesn't automatically translate into sound investments. Short-term news cycles can mislead long-term strategy. Investing without evaluating a company's financial strength, competitive positioning, or growth potential exposes the portfolio to considerable risk. Investment decisions should be grounded in comprehensive sector and company analysis, not in emotional reactions such as fear or FOMO (fear of missing out). Rational, deliberate thinking is essential for making sound financial choices.

J.1.4. Qwen 2.5-Max

Serkan's actions demonstrate that investment decisions are shaped not only by rational calculations but also by psychological and emotional influences. He invested based on frequently encountered news content without analyzing corporate financials or market trends. Media often amplifies short-term events, which may not support sustainable investment strategies. By neglecting risk assessment and reacting emotionally to war coverage, Serkan increases the likelihood of long-term losses. His behavior may result in both financial damage and psychological stress. Investment strategies require emotional control and analytical discipline to avoid these outcomes.

J.1.5. Copilot

Serkan's behavior impairs the rationality of his investment decisions. Availability bias causes individuals to overestimate the importance of information that is recent or frequently seen. Confirmation bias may further compound the issue by leading Serkan to seek only information that supports his beliefs, ignoring potential risks. These biases can be particularly dangerous in volatile sectors, potentially resulting in serious financial losses.

Following the presentation of all ten scenarios, the explanatory elements, reasoning depth, and interpretative capabilities demonstrated by LLM-based generative AI applications in relation to behavioral finance and investor psychology are synthesized in Table 2, titled:

Table 2. Explanatory criteria of llm-based generative ai applications in the evaluation of behavioral finance and investor psychology scenarios.

Scenario	LLM App	Type of Justification	Concepts/Models Used
1	ChatGpt 4o	Theoretical/Multi-Conceptual	Disposition Effect, Mental Accounting, Endowment Effect
	DeepSeek	Strategic & Psychological	Loss Aversion, Opportunity Cost
	Gemini 2.0 Flash	Psychological	Averaging Down, Emotional Attachment
	Qwen 2.5-Max	Emotional Intelligence Focused	Disposition Effect
	Copilot	Practical & Concise	Loss Aversion
2	ChatGpt 4o	Theoretical/Multi-Conceptual	Mental Accounting, Hedonic Framing, Opportunity Cost

	DeepSeek	Strategic & Psychological	Interest-Debt Comparison, 50/30/20 Budgeting
	Gemini 2.0 Flash	Psychological	Persistence Bias, Peer Influence
	Qwen 2.5-Max	Emotional Intelligence Focused	Instant Gratification, Debt Avoidance
	Copilot	Practical & Concise	Spending-Routine Balancing
	ChatGpt 4o	Theoretical/Multi-Conceptual	Overconfidence, Illusion of Control, Confirmation Bias
3	DeepSeek	Strategic & Psychological	Revenge Trading, Risk Neglect
	Gemini 2.0 Flash	Psychological	Survivorship Bias, Emotional Control
	Qwen 2.5-Max	Emotional Intelligence Focused	Bias Typologies, Lack of Learning
	Copilot	Practical & Concise	Risk Perception, Exclusion Bias
	ChatGpt 4o	Theoretical/Multi-Conceptual	Inertia Bias, Sunk Cost Fallacy
4	DeepSeek	Strategic & Psychological	Cognitive Dissonance
	Gemini 2.0 Flash	Psychological	Hope Strategy
	Qwen 2.5-Max	Emotional Intelligence Focused	Emotional Attachment + Lack of Analysis

Table 2 (cont). Explanatory criteria of llm-based generative ai applications in the evaluation of behavioral finance and investor psychology scenarios.

5	Copilot	Practical & Concise	Need for Diversification
	ChatGpt 4o	Theoretical/Multi-Conceptual	FOMO, Herding, Social Contagion
	DeepSeek	Strategic & Psychological	Lack of Research, Behavioral Inertia
	Gemini 2.0 Flash	Psychological	External Attribution, Responsibility Avoidance
	Qwen 2.5-Max	Emotional Intelligence Focused	Misinformation, Behavioral Shaping
	Copilot	Practical & Concise	FOMO

6	ChatGpt 4o	Theoretical/Multi-Conceptual	Anchoring, Regret Aversion
	DeepSeek	Strategic & Psychological	Past Price Fixation
	Gemini 2.0 Flash	Psychological	Hope of Recovery
	Qwen 2.5-Max	Emotional Intelligence Focused	Gain-Loss Imbalance
	Copilot	Practical & Concise	Violation of Risk Threshold
7	ChatGpt 4o	Theoretical/Multi-Conceptual	Confirmation Bias, Selective Exposure
	DeepSeek	Strategic & Psychological	Media Manipulation
	Gemini 2.0 Flash	Psychological	Filter Bubble Effect
	Qwen 2.5-Max	Emotional Intelligence Focused	Need for Information Diversity
	Copilot	Practical & Concise	Critical Thinking Emphasis
8	ChatGpt 4o	Theoretical/Multi-Conceptual	Narrative Bias, Media Influence
	DeepSeek	Strategic & Psychological	Narrative vs. Analytical Evidence
	Gemini 2.0 Flash	Psychological	Regret Aversion, Story Framing
	Qwen 2.5-Max	Emotional Intelligence Focused	Strategic Analysis Deficiency
	Copilot	Practical & Concise	NFT Analysis Inadequacy

Table 2 (cont). Explanatory criteria of llm-based generative ai applications in the evaluation of behavioral finance and investor psychology scenarios.

9	ChatGpt 4o	Theoretical/Multi-Conceptual	Familiarity Bias, Emotional Attachment
	DeepSeek	Strategic & Psychological	Perception–Decision Mismatch
	Gemini 2.0 Flash	Psychological	Emotional Satisfaction Bias
	Qwen 2.5-Max	Emotional Intelligence Focused	Lack of Analysis
	Copilot	Practical & Concise	Risk Awareness and Personal Connection

10	ChatGpt 4o	Theoretical/Multi-Conceptual	Availability Bias
	DeepSeek	Strategic & Psychological	Emotional Triggers in News
	Gemini 2.0 Flash	Psychological	Representativeness Heuristic
	Qwen 2.5-Max	Emotional Intelligence Focused	Need for Source Verification
	Copilot	Practical & Concise	Diversification of Information Sources

The diversity of answers provided by LLM capable AI applications for scenarios and the structure of depth analyses are given in Table 3.

Table 3. Differences in scenario assessments and depth analyses of llm proactive AI applications

Scenario	LLM App	Difference/Depth
1	ChatGpt 4o	One of the most comprehensive responses; incorporated Mental Accounting, Endowment Effect, and Sunk Cost Fallacy with layered analysis.
	DeepSeek	Emphasized opportunity cost, emotional awareness, and provided practical investment strategies.
	Gemini 2.0 Flash	Focused heavily on the psychological justification of holding losses; introduced Averaging Down as a unique angle.
	Qwen 2.5-Max	Framed Zeynep's behavior within emotional intelligence; presented concise but conceptually strong explanations.
	Copilot	Offered a brief yet clear evaluation with direct suggestions; theoretical scope was more limited.
2	ChatGpt 4o	Delivered a multi-layered analysis involving hedonic framing, sunk cost, and opportunity cost violations.

Table 3 (cont). Differences in scenario assessments and depth analyses of llm proactive AI applications

	DeepSeek	Provided impactful numerical examples (interest burden) and practical budgeting rules (e.g., 50/30/20).
	Gemini 2.0 Flash	Interpreted behavior through mental framing; assessed opportunity cost in depth.
	Qwen 2.5-Max	Strongly emphasized emotional impulses such as FOMO, instant gratification, and debt normalization.
	Copilot	Short but balanced response; focused on life quality vs. financial sustainability.

3	ChatGpt 4o	Offered a robust theoretical framework; cited Confirmation Bias, Neglect of Risk, and literature references.
	DeepSeek	Skillfully examined “gambling behavior” and revenge trading using behavioral psychology.
	Gemini 2.0 Flash	Introduced deeper concepts such as Survivorship Bias and the emotional learning cycle.
	Qwen 2.5-Max	Provided one of the most structured analyses; listed biases and grounded them in theory.
	Copilot	More concise but focused; emphasized objective reflection.
4	ChatGpt 4o	Included rare biases like Inertia Bias; conducted a multilayered analysis of the investor’s inertia.
	DeepSeek	Strong emphasis on Cognitive Dissonance and critique of emotional anchoring.
	Gemini 2.0 Flash	Explained with behavioral terms such as “hope strategy” and emotional relief.
	Qwen 2.5-Max	Focused on psychological adaptation and decision evolution.
	Copilot	Presented a more general view; addressed emotional bias and diversification.
5	ChatGpt 4o	Combined FOMO, diffusion of responsibility, and self-justification within social psychology.
	DeepSeek	Proposed preventive strategies to avoid repeating the same behavior.
	Gemini 2.0 Flash	Highlighted external attribution bias and artificiality in social media narratives.
	Qwen 2.5-Max	Focused on risk control and portfolio diversification with actionable suggestions.
	Copilot	Concise yet effective; balanced in conceptual and prescriptive content.

Table 3 (cont). Differences in scenario assessments and depth analyses of llm proactive AI applications

6	ChatGpt 4o	Included both behavioral and financial analysis; integrated Regret Aversion.
	DeepSeek	Focused on the psychological root of “being stuck in the past.”
	Gemini 2.0 Flash	Explained sunk cost effect with strong examples.
	Qwen 2.5-Max	Evaluated behavior within the context of portfolio strategy.
	Copilot	Offered behavior training suggestions and systematic strategic correction.

7	ChatGpt 4o	Analyzed behavior in both pre- and post-decision phases.
	DeepSeek	Emphasized informational imbalance and detachment from market realities.
	Gemini 2.0 Flash	Highlighted selective exposure and media manipulation risks.
	Qwen 2.5-Max	Offered strategic recommendations: data diversity, media literacy, balanced decision-making.
	Copilot	Suggested escaping “filter bubbles” to support objective decision-making.
8	ChatGpt 4o	Strong framing of media influence and the social bubble phenomenon.
	DeepSeek	Assertively stressed the need for data-driven decisions over narratives.
	Gemini 2.0 Flash	Enriched the discussion with regret aversion, external attribution bias, and storytelling impact.
	Qwen 2.5-Max	Proposed critical social media literacy and long-term investment planning.
	Copilot	Critiqued lack of technical NFT analysis; offered detailed corrective strategies.
9	ChatGpt 4o	Skillfully revealed the perceptual blindness of emotional investors.
	DeepSeek	Clearly articulated the investment vs. consumption preference dichotomy.
	Gemini 2.0 Flash	Analyzed regret aversion as a mechanism of self-defense.
	Qwen 2.5-Max	Explained the risks of avoiding fundamental analysis.
	Copilot	Highlighted emotional awareness and need for portfolio diversification.
10	ChatGpt 4o	Clearly explained how media frequency induces cognitive prioritization.

Table 3 (cont). Differences in scenario assessments and depth analyses of llm proactive AI applications

DeepSeek	Offered compelling analysis of the emotional-media investment link.
Gemini 2.0 Flash	Addressed Representativeness Heuristic and causality errors.
Qwen 2.5-Max	Provided strong strategic suggestions on verification, analysis, and media literacy.
Copilot	Emphasized bias awareness and encouraged diverse information sourcing.

IV. CONCLUSION

This study aimed to systematically examine the extent to which large language model (LLM)-based generative artificial intelligence applications can understand, interpret, and evaluate investor psychology within the framework of behavioral finance theories. As part of the analysis, each LLM's responses to 10 fictional investor behavior scenarios were assessed according to four key criteria: Conceptual Depth (Table 4), Psychological Insight (Table 5), Strategic Solution Orientation (Table 6), Conceptual Originality (Table 7).

The findings emphasize that AI systems should not only be evaluated in terms of technical or computational accuracy, but also in their ability to recognize and reason through psychological biases and cognitive distortions inherent in real-world investment decisions. This multidimensional approach underscores the importance of developing emotionally-aware and behaviorally-informed AI in financial decision support systems.

Table 4. Evaluation of conceptual depth and theoretical familiarity.

LLM Application	Evaluation Summary
ChatGpt 4o	Demonstrates the strongest conceptual depth and closest alignment with academic literature. Frequently references core contributors to behavioral finance such as Thaler, Kahneman, Tversky, and Shefrin. Biases are not merely identified but also effectively linked to investor behavior in a well-integrated analytical framework.
DeepSeek	Balances theoretical knowledge with practical investment insights. Offers a hybrid approach, incorporating both behavioral concepts and strategic suggestions (e.g., 50/30/20 budgeting rule). Shows solid conceptual depth while maintaining applied relevance.
Gemini 2.0 Flash	Displays strong psychological depth, with a conceptual structure rooted more in emotional and cognitive mechanisms than in academic references. Analyzes emotional triggers and defense mechanisms with nuance, though literature citations are less explicit.
Qwen 2.5-Max	Adopts a structured and systematic approach, using behavioral finance terminology appropriately. Explanations tend to be clear, accessible, and instructive, with a strong emphasis on media literacy, emotional awareness, and applied behavioral insight.
Copilot	Provides concise and actionable explanations. While the use of academic concepts is occasionally limited, it compensates with pragmatic clarity and implementable recommendations. Generally adopts a "practical inference" style with low cognitive load for users.

Tablo 5. Evaluation of approaches to psychological and emotional dimensions.

LLM Application	Evaluation Summary
ChatGpt 4o	Interprets psychological impacts primarily through conceptual frameworks. Emotional dynamics are approached from a technical and theoretical perspective, with clear reference to established behavioral biases.
DeepSeek	Focuses on explaining emotional effects using concrete, scenario-based examples, while maintaining a solution-oriented tone. Prioritizes investor actionability over emotional elaboration.
Gemini 2.0 Flash	Exhibits the most psychologically nuanced approach among the models. Provides in-depth analysis of the investor's inner emotional conflicts, defense mechanisms, and

	cognitive-emotional inconsistencies. Demonstrates a deep understanding of the human psyche within financial contexts.
Qwen 2.5-Max	Strives to establish a balance between emotional responses and rational strategies. Places strong emphasis on emotional awareness, media influence, and decision-making under uncertainty. Offers a practical-emotional blend.
Copilot	Adopts a simplified and intuitive style. While it does not explore emotional processes in depth, it often delivers accurate emotional cues through brief but insightful observations.

Tablo 6. Evaluation of critical thinking and strategic solution orientation.

LLM Application	Evaluation Summary
ChatGpt 4o	Provides solutions grounded primarily in theoretical frameworks. Recommendations such as avoid sunk cost fallacy.
DeepSeek	Offers the most comprehensive and strategic recommendations. Incorporates concrete rule-based systems (e.g., stop-loss), portfolio structuring, financial literacy, and planning tools. Exhibits a multidimensional decision-support approach.
Gemini 2.0 Flash	Emphasizes emotional and psychological strategies such as acceptance, self-awareness, and cognitive readiness. Solutions focus more on the emotional regulation side of investor behavior than strict financial planning.
Qwen 2.5-Max	Demonstrates systematic and instructive solutions, often emphasizing financial education, risk literacy, and the importance of professional consultation. Balances cognitive and behavioral intervention techniques.

Tablo 7. Evaluation of conceptual originality and introduction of novel constructs.

LLM Application	Evaluation Summary
ChatGpt 4o	Demonstrates a rich variety of behavioral finance concepts, including mental accounting, endowment effect, sunk cost fallacy, and regret aversion. The application shows a broad command of established terminology and applies it meaningfully to investor behavior.
DeepSeek	Exhibits originality through its systematic reasoning and real-world examples. While not always introducing new terminology, its grounded and applicable framing makes it distinctly practical.
Gemini 2.0 Flash	Uses psychologically evocative but non-standard terminology such as "hope strategy", "emotional relief", and "cognitive defense". These original phrases reflect deep insight, even though they are not formal behavioral finance constructs.
Qwen 2.5-Max	Occasionally ventures into less common conceptual territory, referencing terms like inertia bias, media manipulation, and filter bubbles. Its originality lies in the integration of cognitive psychology and media studies into financial behavior analysis.
Copilot	While not highly original in conceptual terms, it offers clear, simplified examples and reasoning suited to non-expert or low-financial-literacy users. Its strength lies in making core ideas accessible rather than introducing new ones.

The evaluation results reveal that the examined LLM-based generative AI applications demonstrate a high degree of familiarity with the behavioral finance literature, effectively applying theoretical constructs in the context of scenario-based investor behavior analysis. The findings show that LLMs can correctly recognize the types of biases frequently discussed in the behavioral finance literature (e.g. loss aversion, disposition effect, overconfidence, mental accounting). In this respect, it has been observed that theoretical infrastructures such as Kahneman and Tversky's (Kahneman & Tversky, 1979) prospect theory and Barberis and Thaler's (Barberis & Thaler, 2003) mental accounting approach can be internalized by LLMs at the conceptual level. However, significant differences were observed across models in terms of psychological depth, strategic solution generation, and conceptual originality.

Notably, behavioral biases such as the Disposition Effect, Loss Aversion, Sunk Cost Fallacy, Mental Accounting, and Overconfidence Bias emerged as recurring themes across nearly all scenarios. This suggests that the models possess a strong capability to recognize cognitive patterns, which appears to be closely tied to the architectural design, training data quality, and ethical alignment of each LLM.

When compared with findings from the behavioral finance literature—particularly Zhou et al. (2024) and Shefrin and Statman (1985)—the results indicate that while classical cognitive biases (e.g., anchoring, confirmation bias, regret aversion) are frequently recognized by the models, such recognition does not always translate into actionable or theoretically consistent investment strategies. This highlights a gap between bias identification and solution-oriented financial reasoning, pointing to a key area for future development in AI-driven decision support systems.

When the results of the study, specifically for artificial intelligence models, are examined;

ChatGPT 4o's literature-based, systematic approach overlaps with the behavioral portfolio modeling processes described by Barberis and Thaler (Barberis & Thaler, 2003). Its deep theoretical explanations and the level of consistency between concepts reveal the potential of LLMs to be an academic reference.

Gemini 2.0 Flash's focus on deep psychological layers such as cognitive dissonance, emotional release and irrational commitment, especially in investment decisions, can be directly associated with the "risk-as-feelings" theory developed by Loewenstein, Weber, Hsee and Welch (Loewenstein et al., 2001). The "psychological insight" capacity offered by this model is parallel to numerous studies indicating that investment decisions are based not only on rational information processing but also on emotional processes (Ariely & Berns, 2010). On the other hand, some models, such as Copilot, have been found to offer more superficial, intuitive, and pragmatic answers. This suggests that the differences in the scope, timeliness, and contextual orientations of LLMs' training datasets are directly reflected in the results. Therefore, the capacity of LLMs to recognize and interpret behavioral finance biases is not fixed; it varies depending on many factors such as the nature of the algorithms used, ethical filtering systems, data diversity, and model architecture (Bubeck et. al., 2023).

The overall findings of the study suggest that LLM-based generative AI tools have achieved a high level of success in understanding and interpreting the core principles of behavioral finance. However, each model tends to emphasize different interpretive dimensions—some focusing on internal psychological processes, others on strategic financial instruments, and yet others on the influence of external information sources. This interpretive variation reflects the multifaceted nature of behavioral finance, underscoring that investor behavior cannot be adequately assessed through a singular analytical lens.

Finally, the summarized performance scores based on the evaluation criteria are presented in Table 8. In addition to the analytical findings, the results of this study raise critical concerns about the use of LLM-based systems as standalone decision-making tools in financial advisory services. Particularly for individuals with low financial literacy, the potential for overreliance on such systems underscores the need to improve LLMs' ability to recognize behavioral biases and to ensure that their recommendations are developed in alignment with principles of transparency, ethical sensitivity, and explainability.

Tablo 8. Overall comparative evaluation of llm-based generative ai applications

LLM App	Conceptual Depth	Psychological Insight	Strategic Recommendation	Originality	Overall Evaluation
ChatGpt 4o	<i>Demonstrates the strongest connection to academic literature, incorporating concepts from Thaler, Kahneman, Tversky, and Shefrin. Biases are not only defined but also effectively linked to behavioral patterns.</i>	<i>Addresses emotional influences through technical and theoretical constructs, offering an analytical view of investor psychology.</i>	<i>Provides theory-based solutions, such as "avoid sunk cost fallacy," but is less practical in application.</i>	<i>Offers a wide conceptual repertoire, including mental accounting, endowment effect, sunk cost, and regret aversion.</i>	<i>It stands out for its high-level command of the behavioral finance literature. In each scenario, it explains multiple biases within a coherent theoretical framework, structuring its interpretations in an academically rigorous manner. Furthermore, it successfully establishes conceptual interconnections and maintains consistency across different scenarios, reflecting a strong grasp of theoretical integration.</i>
DeepSeek	<i>Balances theory and practical application, integrating strategic financial tools like the 50/30/20 rule.</i>	<i>Explains emotional dynamics using realistic examples, but remains solution-focused.</i>	<i>Delivers the most pragmatic and strategic suggestions, including stop-loss rules, portfolio construction, and financial planning.</i>	<i>Displays originality through structured reasoning and real-life relevance.</i>	<i>It stands out with its solution-oriented approach and strategic recommendations. Particularly notable is its contribution to practical investor behavior improvement, offering applicable strategies such as stop-loss rules, portfolio diversification, and the 50/30/20 budgeting rule. While it adopts a simpler conceptual language, its responses are well-structured and directive, reflecting a clear orientation toward real-world applicability.</i>
Gemini 2.0 Flash	<i>Conceptually sound, but primarily emphasizes psychological mechanisms such as cognitive dissonance and emotional defense strategies.</i>	<i>Provides the deepest emotional and psychological analysis, exploring internal contradictions and rationalization strategies.</i>	<i>Prioritizes emotion-based strategies like "acceptance," "alertness," and "self-reflection."</i>	<i>Introduces non-standard but impactful constructs, such as "hope strategy" and "emotional relief."</i>	<i>It is the application that offers the highest level of psychological depth. It specifically focuses on dimensions such as cognitive dissonance, emotional defense mechanisms, and internal rationalization strategies within the investor decision-making process. Through the use of original conceptualizations such as the "hope strategy," "emotional relief," and "affective attachment," it introduces innovative content into behavioral analysis.</i>
Qwen 2.5-Max	<i>Uses a structured and systematic approach, applying terminology correctly with clear, often pedagogical explanations.</i>	<i>Balances emotional and rational strategies; emphasizes awareness, media literacy, and emotional regulation.</i>	<i>Suggests educational and structured interventions, including financial literacy and professional support.</i>	<i>Occasionally employs less common concepts like inertia bias and media manipulation.</i>	<i>It stands out with its systematic analytical framework. By emphasizing external factors such as media influence, information accessibility, and narrative bias, it broadens the scope of behavioral finance beyond individual cognitive processes to a more holistic perspective. The inclusion of solution-oriented recommendations such as financial education, awareness-building, and critical media literacy reflects the model's comprehensive and multidimensional analytical orientation.</i>
Copilot	<i>Offers concise, intuitive insights. Conceptual coverage is limited but easy to grasp.</i>	<i>Takes a simplified approach, offering intuitive but less elaborated psychological interpretations.</i>	<i>Provides simple, actionable suggestions ("diversify, stay disciplined, know your emotions").</i>	<i>Conceptually limited but contributes via clear examples suited to users with low financial literacy.</i>	<i>It contributes through short and concise explanations. Although it adopts a simpler approach compared to other models in terms of conceptual richness, its responses exhibit high intuitive accuracy and contain practical recommendations. In this respect, it approaches behavioral finance concepts not from a purely academic standpoint, but rather from a practice-oriented perspective.</i>

In this regard, it is crucial that policy-makers and technology developers clarify the ethical and regulatory frameworks governing the deployment of AI systems that may resemble or function as financial advisory tools. To ensure that LLM-based applications can provide more accurate and consistent recommendations in the future, the following guiding principles are proposed:

- **Data Quality and Representativeness:** The diversity, neutrality, and reliability of the data used to train LLMs directly influence the quality of their outputs. Training models on balanced and representative datasets, free from socio-cultural biases, is essential to preventing misleading suggestions and erroneous inferences.
- **Contextual Sensitivity and Domain Knowledge Integration:** Although LLMs are capable of generating content across diverse topics, they may produce incomplete or superficial interpretations when domain-specific expertise is required. Therefore, enhancing context awareness through integration with domain knowledge or hybrid expert systems is recommended to improve judgment in specialized financial contexts.
- **Explainability and Justification:** It is not sufficient for AI systems to merely generate correct outputs; their reasoning processes must be transparent and intelligible. Structuring LLM outputs in an explainable format strengthens user trust and facilitates the identification of erroneous reasoning or faulty advice.
- **Feedback Loops and Human Oversight:** Incorporating human supervision and interactive feedback mechanisms into the decision-making processes of LLMs is critical for refining model outputs over time. This approach aligns with the broader goal of human-centered artificial intelligence, ensuring that AI augments rather than replaces human judgment.
- **Alignment with Ethical and Normative Principles:** LLM-based decision systems must comply not only with technical standards but also with ethical principles and societal norms. Upholding values such as fairness, privacy, and transparency enhances public acceptance and mitigates the risks of algorithmic harm in financial decision-making.
- **Continuous Updating and Adaptation:** Given the dynamic nature of financial information, training data can become outdated over time. Therefore, regular updates, continuous learning, and adaptive mechanisms are necessary to ensure that models remain responsive to evolving market conditions and knowledge developments.

In conclusion, the most significant contribution of this study to the existing literature lies in its systematic and comparative analysis of how generative AI-based large language models (LLMs) respond to behavioral finance scenarios, particularly in terms of their theoretical grounding, explainability, conceptual consistency, and terminological coherence. By evaluating the extent to which each model relies on behavioral finance constructs, the study provides a nuanced understanding of the conceptual adequacy and interpretive capacities of LLMs within investor psychology contexts.

Furthermore, the observed differences in terminology usage, bias recognition ability, and recommendation styles among the LLMs offer important insights into the future development of AI-based financial decision support systems. In this regard, the study represents a pioneering effort that bridges the behavioral finance and artificial intelligence literatures through an interdisciplinary perspective.

However, the study has several limitations. The scenarios were exclusively constructed within the framework of behavioral finance theory; only five LLM applications were analyzed; and the evaluations were primarily based on qualitative comparisons. Moreover, the study does not assess how these models affect actual investment decisions or user interactions. Therefore, while the findings are informative, their generalizability is limited, and contextual interpretation is recommended.

Future research should expand on this work by using larger samples, conducting quantitative assessments based on user experience, and incorporating diverse investment scenarios. Longitudinal studies could also investigate how LLMs evolve over time in their alignment with behavioral theories and decision-making performance. Beyond recognizing behavioral biases, further inquiry should examine how LLMs actively guide users away from such biases and whether they can foster more rational decision-making in real-world financial contexts.

DECLARATIONS

Acknowledgements: The author do not wish to acknowledge any individual or institution.

Author Contributions: All parts of the study were conducted by the Author.

Conflict of Interest Disclosure: The author declare no conflict of interest.

Copyright Statement: The author retain the copyright of their work published in the journal, which is licensed under the CC BY-NC 4.0 license, allowing others to share and adapt the work for non-commercial purposes with appropriate attribution.

Funding/Supporting Organizations: This research received no external funding.

Ethical Approval and Participant Consent: This study does not involve human or animal participants. All procedures followed scientific and ethical principles, and all referenced studies are appropriately cited.

Plagiarism Statement: This article has been evaluated for plagiarism and no instances of plagiarism were detected.

Use of AI Tools: In accordance with the aim of this study, various Large Language Model (LLM)-based AI tools — including ChatGPT 4o, Deepseek, Gemini 2.0 Flash, QwenChat 2.5 Max, and GitHub Copilot — were systematically used as subjects of analysis. These tools generated responses to ten scenario-based prompts related to behavioral finance, which were then evaluated by the researchers based on conceptual depth, psychological insight, strategic orientation, and originality.

No AI tools were used in the writing, editing, or structuring of the manuscript itself. All content analysis, interpretation, and academic writing were conducted solely by the authors.

REFERENCES

- Ariely, D., & Berns, G. S. (2010). Neuromarketing: the hope and hype of neuroimaging in business. *Nature Reviews Neuroscience*, 11(4), 284-292. <https://doi.org/10.1038/nrn2795>
- Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance, Volume 1A Corporate Finance*, (1st ed., pp. 1053–1128.) Elsevier. [https://doi.org/10.1016/S1574-0102\(03\)01027-6](https://doi.org/10.1016/S1574-0102(03)01027-6)
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). <https://doi.org/10.1145/3442188.3445922>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2108.07258>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901. <https://doi.org/10.48550/arXiv.2005.14165>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T. & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2303.12712>
- Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., Ryder, ... Zaremba, W. (2021). Evaluating large language models trained on code. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2107.03374>
- Chomsky, N. (1959). *Syntactic structures*. Mouton & Co.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North*

- American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (Vol. 1, Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.1810.04805>
- Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In R. Stouffs, S. Roudavski, & B. Davis (Eds.), *Machine Learning and the City: Applications in Architecture and Urban Design* (pp. 149–164). Springer. <https://doi.org/10.1002/9781119815075.ch45>
- Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 328–339). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P18-1031>
- Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing* (2nd ed.). Prentice Hall.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291. <https://doi.org/10.2307/1914185>
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J. & Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2001.08361>
- Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., Epstein, D., Polosukhin, I., Devlin, J., Lee, K., Toutanova, K., Jones, L., Kelcey, M., Chang, M. W., Dai, A. M., Uszkoreit, J., Le, Q. & Petrov, S. (2019). Natural questions: A benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7, 453–466. https://doi.org/10.1162/tacl_a_00276
- Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31–57. <https://doi.org/10.1145/3233231>
- Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M. & Zettlemoyer, L. (2020). Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8, 726–742. <https://doi.org/10.48550/arXiv.2001.08210>
- Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267–286. <https://doi.org/10.1037/0033-2909.127.2.267>
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2005.00661>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training [Technical report]. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- 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/10.2307/2327802>
- Strubell, E., Ganesh, A., & McCallum, A. (2020, April). Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(9), 13693–13696. <https://doi.org/10.48550/arXiv.1906.02243>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>
- Zhou, Y., Ni, Y., Gan, Y., Yin, Z., Liu, X., Zhang, J., Liu, S., Qiu, X., Ye, G. & Chai, H. (2024). Are LLMs rational investors? A study on detecting and reducing the financial bias in LLMs. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2402.12713>