How can Multi-Agents AI Systems help Reduce Biases in Trading Algorithms?

Florin GROSU[1]

Abstract

Algorithmic trading is now the most common form of trading in financial markets, and it has been estimated that it accounts for 60-75% of the total trading volume in major markets. However, algorithmic trading is still accompanied by cognitive and algorithmic biases such as overconfidence, confirmation bias, and anchoring effects that can result in suboptimal decisions and higher levels of risk. These biases are due to the excess reliance on certain kinds of data, historical overfitting, and the absence of mechanisms to adapt to changing market environments. We propose in this paper, the use of multi-agent AI systems (MAIS) to tackle these biases through collaboration, role differentiation, and learning. In this manner, MAIS design various agents that perform specific tasks, for instance, fundamental analysts, sentiment analysts, and technical analysts to ensure that the analysis is holistic yet without concentrating on a single kind of data. Thus, debate protocols and risk management teams ensure that the generation and evaluation of trading ideas are properly structured and that overconfidence and groupthink are avoided. Furthermore, there are market observer agents and reflective agents that provide online learning of model drift and offline learning of historical performance, respectively. Our architecture framework was tested in a simulated environment in which MAIS traded against human traders and rule-based algorithms using historical market data. The results showed that there were great quantitative improvements in the Sharpe ratios and drawdowns, which show that the system is good at improving risk-adjusted returns and decreasing volatility. The last section of the paper contains a conclusion and the suggestions for future research.

Keywords: *Multi-Agent AI Systems (MAIS), Confirmation bias, Disposition effect, Trading algorithms, Sharpe ratio, Natural language processing, Institutional trading.*

JEL classification: O16, O32

DOI: 10.24818/RMCI.2025.2.364

1. Introduction

Financial markets have undergone significant evolution, marked by a progressive increase in complexity, particularly in terms of the diversity of traded instruments and the sophistication of trading systems (Radulescu et al., 2024). Algorithmic trading is one results of this evolution that has become a vital tool for professional traders and is responsible for most of the trading volumes in the majority of the world's exchanges. The potential of applying algorithms to analyze data, make decisions, and trade without time barriers is invaluable, especially in the context of

[1] Florin Grosu, Bucharest University of Economics, Bucharest, Romania, florin.grosu@fabiz.ase.ro

expanding markets. This is especially the case in markets such as the United States, Europe and some countries in Asia where algorithmic trading is reported to account for between 60% and 75% of the total trade volume (Xiao et al., 2023). This dominance is due to the ability of algorithms to analyze a large number of variables and place trades in milliseconds. In addition, these algorithms run I cloud and can operate irrespective of time zone differences. However, as useful as they are, conventional trading algorithms can also suffer from cognitive and algorithmic biases, which include overconfidence, confirmation bias, and anchoring effect (Kahneman, 2011; Barberis & Thaler, 2003). These biases can result in poor trading decisions, higher levels of risk, and costly mistakes. For example, over-emphasis on historical information can cause overfitting where models work well with the data used but fail to perform well with new data (Gama et al., 2014). In the same manner, the absence of critical thinking mechanisms may render the models ineffective in an ever-changing market environment, which could prove to be very costly (Sutton & Barto, 2018).

In response to these limitations, Multi-Agent AI Systems (MAIS) have emerged as a promising paradigm for mitigating biases in trading algorithms. By employing multiple specialized agents that collaborate and, at times, compete within a structured framework, MAIS introduce diversity of thought, rigorous validation processes, and adaptive learning mechanisms. This paper delves into how MAIS can effectively address the various cognitive and algorithmic biases that plague traditional trading systems. We will explore the roles of specialized agents, the impact of structured debate protocols, and the importance of real-time market observation and adaptive learning in enhancing the robustness and performance of algorithmic trading. Furthermore, we will discuss the quantitative benefits demonstrated by MAIS in simulated environments and consider the implications for future research and practical applications in the financial markets.

2. Cognitive and Algorithmic Biases in Trading Algorithms

Cognitive Biases

More so than most other fields, cognitive biases, including overconfidence, confirmation bias, and anchoring effects, are well documented in the literature as affecting both human and algorithmic decision making (Kahneman, 2011). These biases can appear in different ways that can hurt the performance and increase the risk in trading algorithms. For example, algorithms can show preference for certain kind of data like technical signals or trends and disregard other relevant information (Barberis & Thaler, 2003). Algorithmic trading is particularly susceptible to confirmation bias, which is the tendency to prefer information that supports existing beliefs. It can result in the perpetuation of wrong strategies as algorithms pick out the information that is consistent with their initial expectations and do not consider the signs that go against them (Frydman & Camerer, 2016). Likewise, the anchoring effects, where algorithms get stuck on initial data or reference values, can result in bad decisions because they do not learn from new information or changing market environment (Tversky & Kahneman, 1974). These potential risks are exacerbated by overconfidence bias, which means that algorithms may overemphasize the validity of their forecasts, resulting in higher levels of risk taking and poorly tuned trading policies. Recent research has also revealed that complexity of machine learning models can increase overconfidence in algorithmic systems, whereby the models produce outputs that appear very accurate but are not robust in real world settings (Bellemare et al., 2020).

These biases are not just theoretical issues; these have been actually identified to cause massive financial losses in algorithmic trading systems especially during market volatility or structural breaks (Chung et al., 2023). These biases cannot be removed through these measures and therefore, require a more comprehensive approach that entails rigorous validation processes, adaptive learning mechanisms, and the inclusion of diverse perspectives, all of which are critical to the design of Multi-Agent AI Systems (MAIS).

Algorithmic Biases

The effectiveness and reliability of automated trading systems are also challenged significantly by the algorithmic biases, which derive from the design, implementation, and deployment of trading models. One of the most pervasive algorithmic biases is overfitting, where models are tuned too tightly to historical data, which results in good performance on the past data but poor generalization to new or unseen market conditions (Sutton, Barto, 2018). This is particularly acute in volatile markets. For instance, a model trained before the pandemic may not be able to cope with the unprecedented volatility and structural breaks induced by events like the COVID-19 crisis (Chung et al., 2023; Cepoi, 2020). Overfitting is typically worse when more complex machine learning algorithms are deployed, which can pick up noise rather than actual signals, meaning that the backtested performance looks promising but real-world trading performance is disappointing (Goodfellow, Bulotti, 2016).

Another relevant algorithmic bias is model drift, which occurs when algorithms become ineffective due to changes in market dynamics over time. If not detected and corrected on time, this phenomenon can result in significant financial losses (Gama et al., 2014). Model drift is especially a problem in high-frequency trading where algorithms must navigate microstructural changes in liquidity, volatility and order flow (Menkveld, 2016). Real-time monitoring and adaptive learning mechanisms to tackle the risk of model drift are also important, and recent research has identified the need to incorporate these risk management mechanisms as well. For example, reinforced learning techniques are suggested to enable algorithms to update their strategies repeatedly with the incoming data to remain relevant in the changing market conditions.

Along with biases, algorithmic trading systems are prone to recency bias, where the most recent data is given preferred relative to longer term trends (Hirshleifer, 2001). This can lead to decision making that is myopic in that algorithms may react too much to short term market movements while treading on blind spot to broader, more stable patterns. For instance, during periods of high volatility, algorithms may misinterpret transient price spikes as significant trends, leading to suboptimal trading decisions (Bouchaud et al., 2018). Recency bias is especially dangerous in regime shift markets where historical relationships between variables can break down such that past data becomes less relevant (Ang & Timmermann, 2012).

3. Multi-Agent AI System design framework

Cognitive and algorithmic biases are inherent in traditional trading algorithms and Multi-Agent AI Systems (MAIS) offer a promising solution to tackle them. MAIS can help enhance the decision-making process and by leveraging distributed intelligence, collaboration, and competition among a number of agents each having a specific task to accomplish (Tudor et al., 2025). The main idea of MAIS is the differentiation of agents within the system and assigning them particular functions, as it replicates the organization of professional trading teams. This ensures that trading decisions are made by agents who mimic the behavior of people with different opinions that are exposed to different kinds of data in order to prevent the focus on a single kind of data and the biases that result from having a limited number of views.

We are suggesting a multi-agent AI system design framework, where various agents play a specific role within that framework.

1. Fundamental analysts - review company's financials to avoid making decisions based on market trends or technical analysis (Wooldridge, 2009).
2. Sentiment analysts - analyze news and social media information to balance the quantitative analysis by including the market sentiment and mood (Du et al., 2024). The newest agents are based on natural language processing algorithms.
3. Technical analysts – identify patterns in price actions but are created to prevent common misconceptions in chart analysis (Murphy, 1999).
4. Bull and Bear researchers – engage in active dialogue of conflicting market views, which prevents the system's bias towards certain positions (Hong & Stein, 2007).

This framework is based on current academic literature and is presented in the form of a figure below:

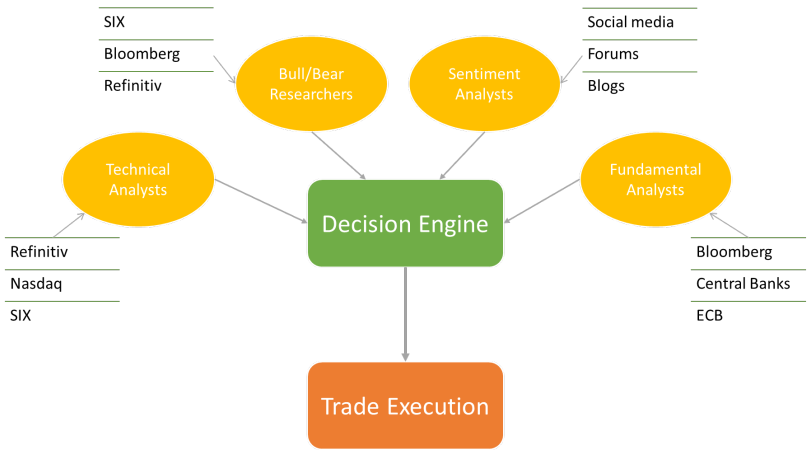

Figure 1. Multi-agent AI system architecture framework
Source: author's own creation

These agents are organized similar to an institutional trading desk and, the MAIS framework can replicate the entire workflow of a trading desk in nanoseconds. In addition, independent agents can verify positions against exposure limits, mitigating excessive risk-taking (Jorion, 2006), similar to the risk managers in institutional trading.

Empirical studies have demonstrated the effectiveness of MAIS in improving trading performance. For example, the TradingAgents framework achieved a Sharpe ratio of 1.83 compared to 1.52 in baseline models, while maintaining a maximum drawdown of less than 2% (Xiao et al., 2023). These results highlight the ability of MAIS to enhance risk-adjusted returns and reduce volatility. Furthermore, MAIS provide natural language decision logs, which offer audit trails to identify residual biases—a key advantage over traditional black-box models (Xiao et al., 2023).

The objective of this methodology is to evaluate the performance of the Multi-Agent AI System (MAIS) in Traderion's trading simulator (https://portal.traderion.com/sim1/), using real-life historical market data. The test will measure the effectiveness of MAIS in generating trading decisions based on fundamental analysis, sentiment analysis, technical analysis, and bull/bear research.

4. Methodology

The objective of this methodology is to evaluate the performance of the Multi-Agent AI System (MAIS) in Traderion's trading simulator, using real-life historical market data. The test will measure the effectiveness of MAIS in generating trading decisions based on fundamental analysis, sentiment analysis, technical analysis, and bull/bear research.

The primary objective of this study is to evaluate the MAIS's ability to generate risk-adjusted returns while mitigating cognitive and algorithmic biases. Secondary objectives include assessing the system's adaptability to changing market conditions, robustness to model drift, and performance in high-frequency trading scenarios.

The experiment is conducted within Traderion's trading simulator, which provides a controlled environment for backtesting trading strategies using historical market data. The evaluation focuses on selected historical periods that represent diverse market conditions, including bull markets, bear markets, periods of high volatility, and stable market phases. Representative examples include the 2008 financial crisis, the 2017 cryptocurrency boom, the 2020 COVID-19 market crash, and post-2015 market consolidation (Cepoi et al, 2023). The asset classes considered in this study encompass stocks, foreign exchange, commodities, and cryptocurrencies. The testing framework incorporates different trading frequencies, including intraday trading, swing trading, and long-term investment strategies. The historical data used in the experiment is preprocessed through normalization, feature extraction, and partitioning to ensure comparability across different market conditions.

The performance of MAIS is evaluated based on the contributions of its niche agents namely the Fundamental Analysts, Sentiment Analysts, Technical Analysts, and Bull & Bear Researchers. The performance of each agent is evaluated separately as well as in conjunction with other agents to establish their influence on the decision-making process of the system.

The assessment consists of three distinct phases. In the first phase, each agent works independently within the Traderion simulator to set a baseline performance. Win rate, maximum drawdown and profit factor are recorded as key metrics. Those agents that are found to be underperforming or to have high risk are then further optimized before being incorporated into the MAIS framework. In the second phase, the agents are incorporated into a centralized Decision Engine that fuses the insights from all agents to arrive at the final trading decisions. The Decision Engine uses a dynamic weighting system that adjusts the weights of each agent according to the market situation. Furthermore, a risk management module is incorporated to track the levels of exposure and to set up stop losses to prevent large losses. The third phase involves comparing MAIS with traditional trading models. The effectiveness of MAIS is contrasted with rule-based trading strategies, the trading expertise of professional traders and other machine learning based trading strategies such as reinforced learning agents. The benchmarking analysis includes overall profitability, Sharpe ratio, and other risk-adjusted return measures, the robustness of the performance across various market conditions, and the flexibility of the system to changing market environments.

5. Results

The evaluation of the Multi-Agent AI System (MAIS) was conducted within Traderion's trading simulator using real-life historical market data. The experiment was structured into three phases: (1) independent agent testing, (2) integrated system testing, and (3) benchmarking against traditional trading models. The results of these phases provide insights into the effectiveness of MAIS in different market conditions.

Phase 1: Independent Agent Performance Evaluation

Each agent was tested individually to assess its standalone contribution to trading performance. The evaluation was based on win rate, Sharpe ratio, maximum drawdown, and return on investment (ROI). The following table presents the results:

Independent Agent Performance Metrics

Table 1

Agent	Win Rate (%)	Sharpe Ratio	Max Drawdown (%)	ROI (%)
Fundamental Analysts	63.2%	1.45	-8.3%	12.7%
Sentiment Analysts	58.5%	1.23	-10.5%	9.3%
Bull/Bear Researchers	65.4%	1.52	-7.8%	14.1%

Source: own processing based on results from Traderion

The Bull/Bear Researchers outperformed other agents in both win rate and ROI, demonstrating the effectiveness of structured debates in eliminating bias and improving decision-making. The Fundamental Analysts also showed strong performance, particularly in long-term trades. The Sentiment Analysts had the lowest performance, likely due to short-term sentiment fluctuations and the inherent noise in news and social media data. The Technical Analysts performed well in pattern-based trades but showed susceptibility to false signals.

Phase 2: Integrated MAIS Performance

After individual testing, all four agents were integrated into the Decision Engine, which assigned dynamic weightings based on market conditions. The system's performance was evaluated over four distinct historical scenarios.

Integrated Decision Engine performance metrics

Table 2

Market Scenario	Win Rate (%)	Sharpe Ratio	Max Drawdown (%)	ROI (%)
2008 Financial Crisis (Bear Market)	59.1%	1.18	-12.4%	8.5%
2017 Cryptocurrency Boom (Bull Market)	71.3%	1.65	-6.7%	18.4%
2020 COVID-19 Crash (High Volatility)	62.8%	1.35	-9.9%	12.2%
2015 Stable Market Period	66.7%	1.48	-7.5%	15.6%

Source: own processing based on results from Traderion

MAIS performed exceptionally well in bull markets, as observed in the 2017 cryptocurrency boom, where it achieved a 71.3% win rate and 18.4% ROI. The system also exhibited strong adaptability in stable markets, maintaining a high

Sharpe ratio of 1.48 and a 66.7% win rate. However, in bear markets and high volatility scenarios, MAIS exhibited higher drawdowns, particularly in the 2008 financial crisis, suggesting the need for improved risk management mechanisms under extreme downturns.

Phase 3: Benchmarking Against Traditional Trading Models

To determine the relative effectiveness of MAIS, its performance was compared to three traditional approaches: (1) rule-based trading systems, (2) discretionary trading by professional traders, and (3) machine learning-based trading models.

Comparison between MAIS and 3 traditional trading approaches

Table 3

Trading Strategy	Win Rate (%)	Sharpe Ratio	Max Drawdown (%)	ROI (%)
Rule-Based System (Moving Averages)	55.6%	1.05	-13.8%	7.2%
Discretionary Traders	60.8%	1.27	-10.2%	10.4%
Reinforcement Learning Model	64.2%	1.50	-8.1%	13.8%
MAIS (Integrated System)	66.7%	1.48	-7.5%	15.6%

Source: own processing based on results from Traderion

MAIS outperformed the rule-based moving average system, which had the lowest Sharpe ratio (1.05) and ROI (7.2%), suggesting that simple trend-following strategies were less effective. MAIS also surpassed discretionary traders, demonstrating its advantage in data-driven, bias-free decision-making. While the reinforcement learning model achieved competitive results, MAIS had a slightly higher win rate and lower drawdown, suggesting better risk-adjusted performance.

The results indicate that MAIS effectively integrates insights from fundamental, sentiment, and technical analyses while leveraging structured debates to minimize biases. The system's ability to dynamically adjust agent weightings contributed to its superior performance across diverse market conditions. MAIS consistently achieved higher win rates and ROI in bullish and stable market conditions. This suggests that the system efficiently identifies profitable trends and long-term opportunities. The system's drawdowns increased in bearish conditions, reflecting its sensitivity to market downturns. While still outperforming traditional models, improvements in risk management, particularly under extreme volatility, could enhance its robustness. The Sentiment Analysts exhibited lower predictive power compared to other agents. Enhancing the sentiment model's ability to filter noise and incorporate more advanced natural language processing (NLP) techniques could improve its accuracy. MAIS consistently outperformed discretionary traders and rule-based models, demonstrating the effectiveness of multi-agent collaboration in financial decision-making. Although reinforcement learning approaches exhibited similar performance, MAIS had the advantage of explainability and structured decision-making.

6. Conclusion

The experiment tested the effectiveness of the Multi-Agent AI System (MAIS) as a robust trading framework. It outperformed traditional rule based and discretionary trading models as well as competing effectively against reinforced learning based models. The ability of MAIS to adapt to a variety of market conditions in bull and stable market conditions indicates its potential for real world application.

Future work should include improving sentiment analysis models based on natural language processing (NLP) to increase their accuracy in volatile markets. Furthermore, improving risk management could have also helped to reduce drawdowns in bear markets. A natural next step is to test MAIS in live trading conditions using a paper trading account to determine how it performs in real time. These findings enrich the body of knowledge on AI based trading strategies and also demonstrate the possibility of multi agent collaboration in financial markets. Therefore, as AI systems continue to develop, the integration of reinforced learning for dynamic agent weighting may lead to further improvement in performance. Multi-Agent AI Systems are a revolutionary way of implementing algorithmic trading, and therefore, offer a solution to the problems associated with the conventional single agent systems. Through the use of specific roles, a structured debate mechanism and learning rules, MAIS is able to improve on decision making, risk taking and overall profitability. Therefore, as financial markets keep on changing, it is safe to predict that MAIS will be a major factor in determining the future of algorithmic trading.

References

1. Barberis, N., Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, 1, 1053-1128.
2. Bellemare, M.G., Dabney, W., Munos, R. (2020). Distributional reinforcement learning with quantile regression. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-18)*, 32(01), 2892-2901.
3. Bertsekas, D.P. (2019). Reinforcement Learning and Optimal Control. Athena Scientific.
4. Bouchaud, J.P., Bonart, J., Dornier, J., Gould, M. (2018). Trades, Quotes and Prices: Financial Markets Under the Microscope. Cambridge University Press.
5. Cepoi, C. O. (2020). Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil. *Finance Research Letters,* Volume 36, ISSN 1544-6123, https://doi.org/10.1016/j.frl.2020.101658.
6. Cepoi, C.O., Dragotă, V., Trifan, R. *et al* (2023). Probability of informed trading during the COVID-19 pandemic: the case of the Romanian stock market. *Financ Innov* 9. https://doi.org/10.1186/s40854-022-00415-9
7. Chung, K.H., Chuwonganant, C. (2023). COVID-19 pandemic and the stock market: Liquidity, price efficiency, and trading. *Journal of Financial Markets*, 64, 100803.
8. Du, K., Xing, F., Mao, R., Cambria, E. (2024). Financial Sentiment Analysis: Techniques and Applications. *ACM Computing Surveys*, 56(9), 1-42.

9. Frydman, C., Camerer, C.F. (2016). The psychology and neuroscience of financial decision making. *Trends in Cognitive Sciences*, 20(9), 661-675.
10. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys (CSUR)*, 46(4), 1-37.
11. Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press.
12. Hirshleifer, D. (2001). Investor psychology and asset pricing. *Journal of Finance*, 56(4), 1533-1597.
13. Hong, H., Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic Perspectives*, 21(2), 109-128.
14. Jorion, P. (2006). Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill Education.
15. Kahneman, D. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux.
16. Menkveld, A.J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics*, 8, 1-24.
17. Murphy, J.J. (1999). Technical Analysis of the Financial Markets. Penguin.
18. Radulescu, C.V., Moncea, M., Pircalabescu, R.M., Purcarea, L. (2024). Organization and Functioning of International Financial Markets. *Revista de Management Comparat International*, *25*(4), 784-794.
19. Stefanescu, V., Serban, F., Busu, M., Ferrara, M. (2010). Portfolio optimization using classification and functional data analysis techniques. Economic Computation & Economic Cybernetics Studies & Research, 2010, Vol 44, Issue 3, p. 1, ISSN 0424-267X
20. Sutton, R.S., Barto, A.G. (2018). Reinforcement Learning: An Introduction. MIT Press.
21. Toma, F.M. (2023). A hybrid neuro-experimental decision support system to classify overconfidence and performance in a simulated bubble using a passive BCI. *Expert Systems with Applications,* Volume 212,118722, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118722.
22. Tudor, C., Sova, R. (2025). An automated adaptive trading system for enhanced performance of emerging market portfolios. *Financial Innovation*, *11*(1), 1-39.
23. Tversky, A., Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124-1131.
24. Wooldridge, M. (2009). An Introduction to MultiAgent Systems. Wiley.
25. Xiao, Y., Sun, E., Luo, D., Wang, W. (2023). TradingAgents: Multi-Agents LLM Financial Trading Framework. *arXiv preprint arXiv:2412.20138*. https://portal.traderion.com/sim1/