Computational Economics
https://doi.org/10.1007/s10614-025-11024-w

Stock Market Forecasting: From Traditional Predictive Models to Large Language Models

Mahmoud Darwish[1] · Ehab E. Hassanien[1] · Amany H. B. Eissa[1,2]

Accepted: 7 June 2025

Abstract
Stock market forecasting is a complex research problem due to the complexity of the factors influencing stock market trends. This survey provides a comprehensive overview of recent advancements in stock market forecasting, focusing on the impact of large language models (LLMs) in financial analytics. The survey explores the strengths and challenges of feature engineering, ensemble methods, hybrid models, text-based prediction and reinforcement learning. It then presents the transformative impact of LLMs, highlighting their capabilities in utilizing transfer learning and few-shot learning to understand complex financial information, enhancing sentiment analysis, improving portfolio management, and stock forecasting accuracy. A key novelty of this survey lies in presenting comprehensive analysis of the strengths and weaknesses of LLMs for different financial tasks in addition to exploring how LLMs can be combined with machine learning and reinforcement learning approaches to overcome their limitations in handling unstructured data, improving model explainability, and enhancing generalizability. Finally, this survey identifies existing research gaps and limitations, proposing future research directions aimed at improving prediction accuracy and utilizing both LLMs and predictive models' capabilities in stock market forecasting.

Keywords Stock market forecasting · Large language models · Reinforcement learning · Feature engineering · Machine learning · LLM-based financial agents

✉ Mahmoud Darwish
mahmoud.fouad@pg.fci-cu.edu.eg

✉ Ehab E. Hassanien
e.ezat@fci-cu.edu.eg

✉ Amany H. B. Eissa
Amani.eissa@eui.edu.eg

1 Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt

2 Faculty of Computing and Information Sciences, Egypt University of Informatics, New Administrative Capital, Egypt

Published online: 26 June 2025

1 Introduction

Stock market and financial variables are hard to predict (Yang et al., 2023; Peng et al., 2021). Accurate prediction helps individuals, investors and institutions make better financial decisions, achieve profits and reduce risk. Stock market forecasting is considered a challenging research problem due to the complexity and diversity of factors affecting the prediction process. These factors include changing economic factors, unpredictable political factors, governmental regulations, investor's sentiment and other unexpected changes in local or global level. The forecasting process includes collecting different influencing factors, making proper analysis for huge and heterogeneous amounts of data to be able to predict future directions. In addition to that, the forecasting process should be updated in a continuous way to cope with rapidly changing economics and other influencing factors.

Over the years, stock market prediction methods have evolved from traditional approaches such as technical analysis, fundamental analysis, and statistical models to more modern techniques based on Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL). These advanced methods are better at identifying patterns and adapting to the complexity and rapid changes of the stock market. However, challenges remain in improving prediction accuracy while dealing with a dynamic stock market environment. traditional ML and DL models face key limitations:

1. Poor cross-market generalization: most are trained on a single index or exchange.
2. model explanatory: deep networks and ensemble methods lack transparent decision paths.
3. Unstructured-data challenges: require heavy pre-processing for text and sentiment inputs.
4. Scalability barriers: full retraining for new tickers or timeframes cause high compute cost.

By contrast, Large Language Models (LLMs) promise to directly address each of these via few-shot transfer, chain-of-thought reasoning, native text handling, and parameter-efficient adapters. LLMs introduced new possibilities, particularly in analyzing huge amounts of unstructured data, such as financial news, stock articles, social media sentiment, images, and videos as well in recent multimodal LLMs. The new capabilities of LLMs promise to provide a better understanding of various stock market data sources compared to existing models. Additionally, recent developments in LLMs offer improved contextual understanding and reasoning, which could lead to deeper insights into market dynamics and more accurate stock market predictions.

This survey provides a comprehensive overview of the latest stock market forecasting methods, with a particular emphasis on the transformative impact of LLMs on stock market analytics. To address the limitations of existing methods and highlight the potential of LLMs, this survey makes the following key contributions:

1. Presents the first unified taxonomy spanning traditional feature engineering, ML/DL, RL/MARL, and LLM-based forecasting methods (Sections 3 and 4).

2. Introduces a computational cost-aware categorization of LLM approaches into domain-specific versus generic/adapter pipelines, highlighting their respective trade-offs (Table 7).
3. Provides the first ML/RL/MARL/LLMs approaches comparative evaluation across four capability dimensions—explainability, cross-market transfer, unstructured-data handling, and training computational cost—using a standardized stacked bar chart (Fig. 6).
4. Identifies eleven concrete, actionable research challenges utilizing both LLMs and Traditional ML,DL, RL and MARL techniques —ranging from real-time scalability to prompt-engineering frameworks—that directly correspond to uncovered gaps (Section 6).

The remainder of this paper is organized as follows: Section 2 provides background information. Section 3 presents recent advancements in stock market analytics using feature engineering, ML, DL, and RL approaches. Section 4 presents the impact of LLMs on stock market forecasting. Section 5 presents an analysis and discussion of the key findings. Section 6 outlines research directions. Section 7 presents the limitations of the study. Finally, Section 8 concludes with a summary of the literature review.

2 Background

Quantitative trading utilizes mathematical functions and automated trading models to make trading decisions (Sun et al., 2023). Quantitative traders use a trading technique to create a mathematical model, then develop a computer program that applies the model to historical market data. The model is then tested again and optimized. The system is implemented in real-time markets with real capital if favorable results are achieved. Quantitative trading has some challenges, as financial markets are dynamic and change rapidly based on actions from many entities including traders, company owners, policy regulators, etc. Therefore, models should be adaptable to change and cope with surrounding changes. Prediction models depend on two major approaches: technical and fundamental analysis. Researchers depend on multiple combined factors from technical and fundamental analysis to improve prediction accuracy. Several researchers introduce other factors like sentiment analysis for economic news, macroeconomic indicators about the overall economy, and textual information about stocks and commodities (Sun et al., 2023; Rossi & Tinn, 2021).

Fundamental analysis is used for long-term investment strategies, focusing on a company's financial health and growth potential (Huang et al., 2019). It evaluates public financial statements, such as earnings, revenue, and profit margins, to estimate a stock's strength and value according to the company's financial conditions. If the market price is below this value, it suggests a buy; if higher, a sale. In contrast, technical analysis relies on historical price and volume data to predict future trends without considering underlying company performance (Picasso et al., 2019). Textual and sentiment analysis has become increasingly important in stock prediction, as analyzing news, social media, and other content reveals public sentiment's influence on stock movements. For example, social media posts by investors or influential figures, like Elon Musk's tweets about Tesla, can significantly impact stock prices (Picasso et al., 2019; Jing et al., 2021). By categorizing sentiment as positive, neutral, or negative, ML algorithms can enhance stock price predictions and market sentiment analysis.

ML and DL models are widely used in stock market analytics (Kumbure et al., 2022; Kehinde et al., 2023; Ashtiani & Raahemi, 2023). Supervised ML algorithms utilize historical stock data as a training dataset to learn relationships be-tween factors like stock prices, economic indicators, news events and other financial factors to provide better predictions about future prices and trend direction. In contrast, unsupervised learning identifies stock clusters and detects anomalies. DL models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory networks (LSTM), Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Network (GRU) have demonstrated significant use in forecasting stock trends due to their ability to process large and complex datasets. RL addresses sequential decision-making challenges by enabling agents to learn optimal actions through trial and error in uncertain environments. By maximizing cumulative rewards, RL offers promising applications in stock market trading. Unlike supervised methods, RL adapts to dynamic stock market conditions with limited previous knowledge, learning through rewards for correct trading decisions and penalties for wrong trading decisions. Multi-agent reinforcement learning (MARL) (Huang et al., 2024), an extension of RL, models scenarios involving multiple agents interacting together to achieve a common goal. Recent studies demonstrate MARL's potential for optimizing stock trading strategies, improving market predictions, and addressing complex multi-agent dynamics in trading environments.

Feature selection is an important preprocessing step in stock market prediction that aims to remove noisy and irrelevant factors while keeping the most effective and non-redundant features (Xu et al., 2023; Kumari & Swarnkar, 2023; Chaudhari & Thakkar, 2023; Chen & Zhou, 2020; Sethia & Raut, 2019; Chung & Shin, 2018; Htun et al., 2023). The complexity of financial factors impacting the stock market makes the stock prediction process more challenging. Feature selection methods are categorized into four types. The first type is filtering methods, which rely on statistical measures, such as correlation, ranking features and retain only the highest scoring ones. The second type is wrapper methods, where feature selection is integrated within the ML process. In these methods, the model is trained on different parts of features, and the subset leads to the best predictive performance is chosen. The third type is embedded methods, which perform feature selection during the model training process. With the increasing diversity and complexity of stock market data, advanced feature selection techniques are necessary for processing different complex stock features inputs, such as technical, fundamental, macroeconomic indicators and textual data, ensuring efficient and accurate stock forecasts. Figure 1 presents an overview taxonomy that organises the forecasting literature into traditional ML and DL approaches and LLM-centric approaches.

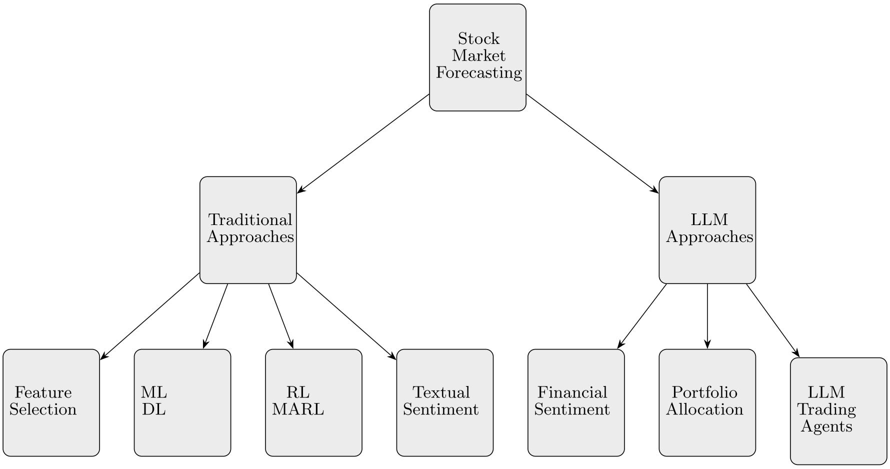


Fig. 1 Taxonomy of stock-forecasting research areas: traditional numeric approaches versus LLM-centric approaches

3 Recent Predictive Models for Stock Market Forecasting

3.1 Feature Engineering for Stock Market Prediction

Feature engineering is considered an important factor in stock market prediction. Figure 2 groups the feature-engineering studies by selection strategy. Selecting the right indicators would lead to accurate future predictions. Xu et al. (2023) introduced a stock market prediction model for China's major airlines based on Least Absolute Shrinkage and Selection Operator (LASSO) regression using multiple mixed features from technical indicators, fundamental indicators, competitor indicators, and economy indicators. the study depends on fundamental and technical indicators including previous daily stock prices for the top three Chinese airlines, sector index, stock market index, and trading volume. It also includes alternative related information and other economic indicators like interest rates. LASSO regression helps to give more value to important features and shrinks the value of remaining features. The study shows that combining LASSO with a group of technical and external features helps to improve prediction accuracy as it achieves fewer prediction errors than other regression algorithms like support vector regression, tree regression, and multilayer neural networks. Although the proposed approach is promising, it cannot achieve the same accuracy in predicting non-airline stocks with the same features so there is a need for a general approach that could tackle multiple types of stocks in different domains.

Kumari and Swarnkar (2023) proposed a stock market prediction approach that combines multiple feature selection techniques to improve prediction accuracy. The approach starts by calculating 83 technical indicators from historical stock prices and the stock index. These indicators are then normalized using hybrid normalization utilizing min-max and z-score. LASSO, Information Gain (IG), and Forward Feature Selection (FFS) are applied to generate three distinct lists of important features.

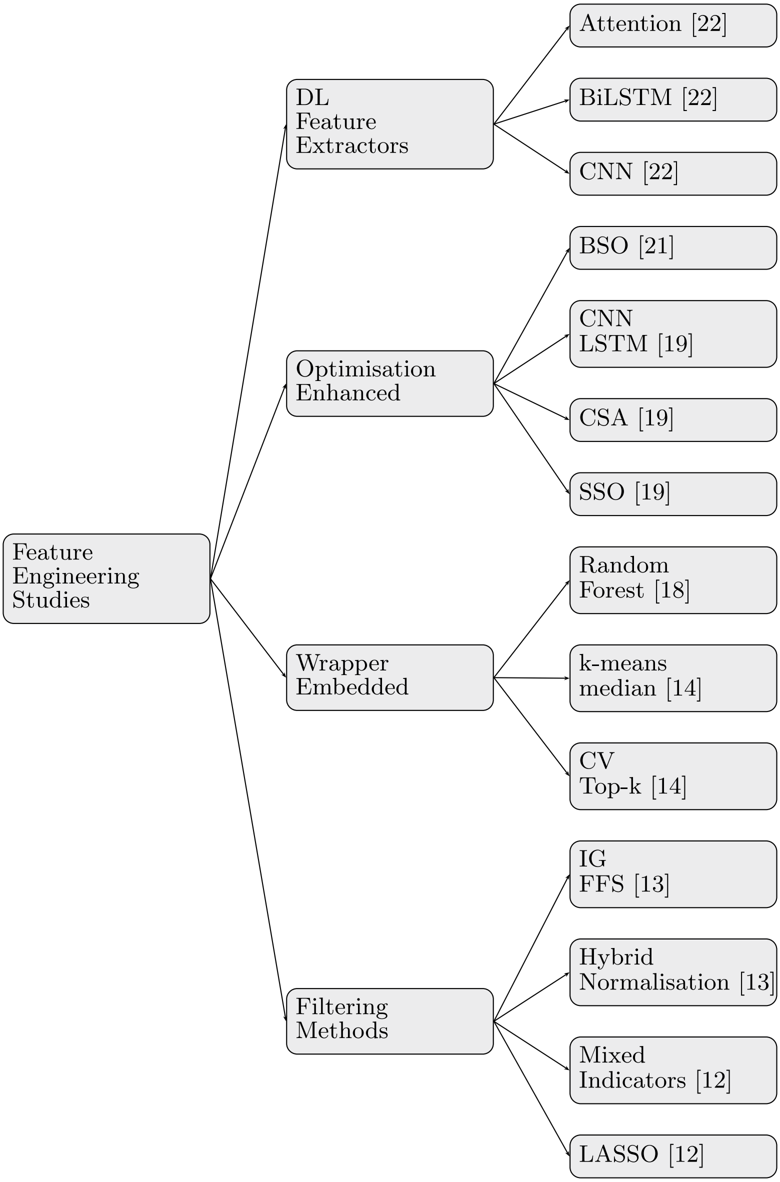


Fig. 2 Taxonomy of feature-engineering studies surveyed in Section 3.1

Intersection operation is performed to merge the outcome of the different feature selection mechanisms and generate only one common list of relative features. The selected features are fed into a prediction model to forecast stock market movements. The researchers compared several models, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). Experimental results showed that the proposed method improved prediction accuracy, with

ANN performing better than SVM and KNN. However, while the approach showed improvement, prediction accuracy remained below 90% in most datasets and under 80% in some, indicating room for further enhancement. Given the importance of financial decisions, higher accuracy is essential to avoid losses.

Chaudhari and Thakkar (2023) proposed a neural network-based approach utilizing the coefficient of variation (CV) feature selection mechanism to improve stock prediction accuracy. The CV selection mechanism identifies stock features based on their variability compared to the mean. The proposed approach integrates CV with k-means clustering, median, and top-k methods to generate a list of relevant features. Combining k-means with CV helps identify features belonging to the largest cluster, while integrating with median identifies features within a defined range from median. Finally, the top-k integration with CV identifies the top k features with the highest CV. The approach was compared with existing methods, such as variance threshold, Chi2, Principal Component Analysis (PCA), mutual information, and correlation coefficient. It also used established neural network-based models like CNN, LSTM, and backpropagation for prediction. Experimental results showed that the top-k CV-based approach outperformed other feature selection methods in terms of performance across most datasets. The other two CV methods based on k-means and median achieved similar accuracy to traditional methods. Although PCA sometimes provided equivalent or better prediction with the LSTM model, the proposed top-k CV method demonstrated lower execution times. However, determining the optimal number of features and the suitable feature selection algorithm for achieving the best prediction accuracy remains a challenge according to several other research studies (Chaudhari & Thakkar, 2023; Chen & Zhou, 2020; Sethia & Raut, 2019; Chung & Shin, 2018).

Htun et al. (2023) proposed a literature review for feature selection and extraction methods in stock market forecasting for the period from 2011 until 2022. Review shows that feature selection is crucial for improving stock model prediction. It also shows that correlation criteria, random forest, PCA, and autoencoder are widely used feature methodologies in stock market predictions. It also shows that several research publications in feature extractions became popular in 2019 and 2021 as multiple research papers have been published covering many feature selections including filter, wrapper, and embedded methods. Researchers showed that future research is needed in ensemble feature selection by combining more than one method or selection approach. Another research direction is to focus on information fusion and applying feature selection to different input features including technical indicators, fundamental indicators, and other types.

Recent research studies have explored the use of advancement in feature selection and optimization algorithms to optimize features input for predictive ML and DL models resulting in improvement in prediction accuracy. Patil et al. (2024) utilizes the Shuffled Shepherd Optimization (SSO) algorithm and the Crow Search Algorithm (CSA) as a feature selection methodology that feeds a hybrid DL model with the relevant subset of features that could improve prediction accuracy. Su et al. (2023) employed a feature selection algorithm to select a useful list of features that could improve model training performance and prediction accuracy. Li et al. (2024) utilizes Boruta feature selection algorithm and Brain Storm Optimization (BSO) to optimize feature selections integrated with support vector regression, improving stock price prediction and regression results. Zhao et al. (2024) proposed a hybrid DL model that utilizes CNN to perform feature selection for stock market prediction combined with BiLSTM to capture dependencies and attention to focus on important stock data. Zhao et al. approach shows improved performance in CAMBRIA and KRIRAN Datasets, however it has poor performance on some datasets.

3.2 Predictive Modeling Using Machine Learning & Deep Learning for Stock Market Forecasting

Figure 3 groups the ML/DL papers by algorithm family. Kehinde et al. (2023) presented a literature review of the last two decades of research in stock market prediction from 2001 until 2021 by investigating a total of 220 articles to map out trends and limitations. The review identified the need for models that can perform

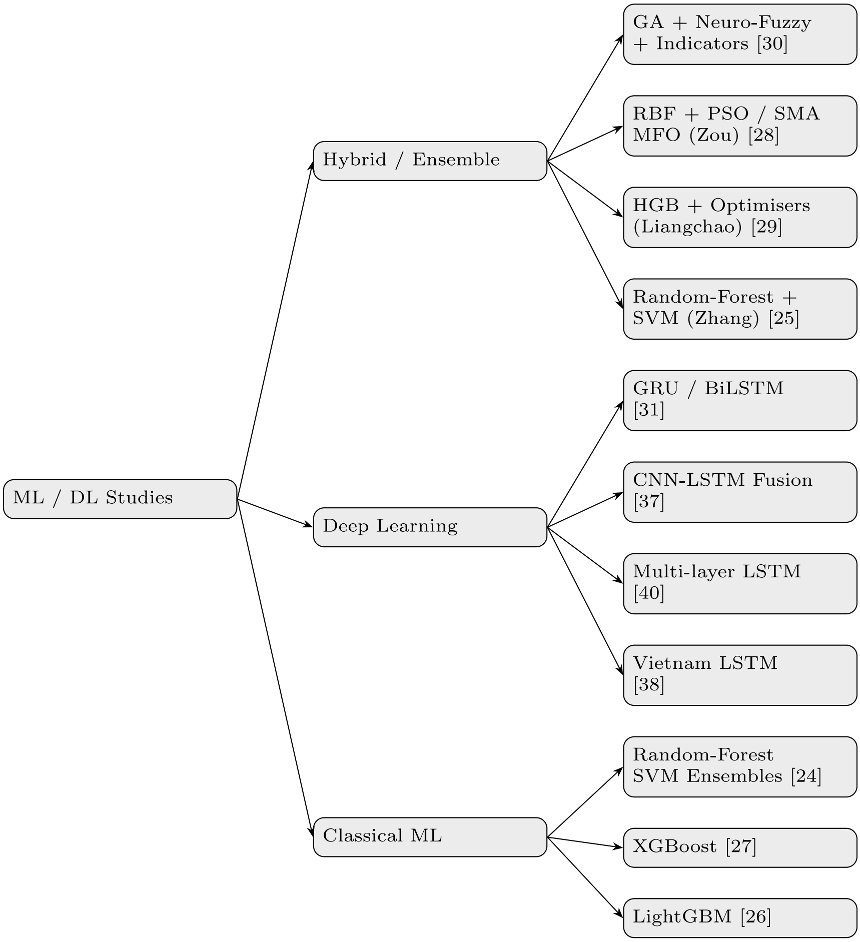


Fig. 3 Taxonomy of ML/DL predictive-modelling studies surveyed in Section 3.2

cross-market analysis prediction as previous studies consider only one market in their analysis. However, due to the differences in stock market policies, it has been challenging to test existing models in other markets. Although there have been significant breakthroughs in addressing stock market predictions using different neural networks, there is a need for collaborative models with less complexity that outperforms the existing models. Previous researchers have examined or forecasted stock market indices, but most of their analysis focuses on a specific market or stock. The use of earlier techniques might be inadequate for those investors who wish to make intelligent decisions about which stock market or stock to invest in because an analyst must assess the prospects of each market or stock individually.

Ashtiani and Raahemi (2023) presented a survey for research papers on stock market prediction from 2015 to 2022. Research shows economic news is a critical part of predicting stock markets. Literature reviews show there is a growing trend towards using AI networks like RNN, DL, and transformers as well. It also shows that stock market prediction still has great room for performance enhancements using BERT and other language models. Cagliero et al. (2023) proposed an approach to improve stock market forecasting by enhancing ML classifier performance based on integrating knowledge extracted from candlestick patterns to reduce the number of recommendations and false signals from the ML classifier. Although experimental results show the proposed approach is efficient, researchers proposed to study integrating technical indicators, news, and social media content as input to features, which might increase accuracy.

Significant research focuses on applying different ML for stock market predictions. Hybrid and Ensemble learning, which combines multiple models, has emerged as a promising research direction to improve prediction accuracy. Nti et al. (2020) and Zhang and Chen (2024) highlights the effectiveness of ensemble methods through combining techniques like random forests and SVM to improve stock prediction accuracy. Although random forests and SVM are widely used, other ML models have shown promising results. For example, Guennioui et al. (2024) utilized the LightGBM ML model to predict stock market pricing during market stress. Other ML models like XGBoost were utilized by Kianrad et al. (2024) to predict stock price movements as a response to company announcements. Zou and Xiao (2024) proposes hybrid learning approach to improve stock price prediction accuracy based on a combination of radical basis function models and three optimizers: particle swarm optimization, slime mold algorithm, and moth flame optimization. Liangchao (2024) proposes a hybrid learning approach for stock price prediction by integrating a histogram-based gradient boosting (HGB) model with optimization techniques which provide better accuracy compared to the base model. Further exploration for ensemble methods and hybrid models through integrating different ML and DL models is a potential research area to improve stock forecasting accuracy.

Building on the power of combining multiple models to improve prediction accuracy and the need to handle more than one stock at the same time in addition to tackling the problem of indicators lagging. Yeo et al. (2023) propose a novel approach that integrates a genetic algorithm, a SeroFAM neuro-fuzzy system, and two optimized technical indicators. The Proposed approach tackles the problem of indicators lagging by introducing new optimized technical indicators fMACDH and fMACDHfRSI to reduce the lagging effect. Some indicators give the right info after a certain delay which would cause loss so reducing the lagging effect enhances prediction and achieves better return. Proposed indicators provide buy/sell signals that are then provided to rule-based investment strategies (TBH and RBBC) for portfolio rebalancing. The main difference between TBH and RBBC is that TBH has multiple groups of stocks clustering high risk, medium risk, and low risk, while RBBC does not consider risk in clusters. Although experimental results are promising, the researchers shared the following limitations, the need to compare the proposed indicators against other forecasting techniques and to consider fundamental analysis and economic news. In addition to that, they suggest exploring alternative optimization algorithms and including transaction costs into model consideration.

DL prediction models are considered major research direction in stock market prediction. Beniwal et al.'s study (Beniwal et al., 2024) on long-term stock price forecasting using different DL models across multiple global indices demonstrated that LSTM outperformed RNN, CNN, BiLSTM and other DL models in terms of accuracy and generalization in 3 global indices while BiLSTM achieved better accuracy in Nikkei 225 index and GRU in Shanghai Stock Exchange composite index. Safari and Badamchizadeh (2024) employed sequence-oriented BiLSTM networks on amazon stock information and experimental results shows better performance in adjusted price and low accuracy in volume prediction. DL techniques are employed with time series analysis, ARIMA, for stock price prediction. Ma et al. (2024) utilized LSTM and Arima as hybrid models for Stock price prediction. ARIMA is employed as well as a method for performance evaluation against DL-based techniques for stock market price forecasting (Srivastava et al., 2022; Low & Sakk, 2023). Bagalkot et al. (2024) proposed optimized stock price prediction accuracy by optimizing the parameters of ARIMA using a grey wolf optimizer. Mulla et al. (2024) proposed a hybrid Approach for Stock Market Index Forecasting using the CNN-LSTM Fusion Model. Phuoc et al. (2024) utilized LSTM to predict the stock price trend in the Vietnam stock market and achieved 93% prediction accuracy. Tatane et al. (2024) proposed a CNN-based stock price prediction model by converting technical indicators from time series to images. Md et al. (2023) proposed multi-layer LSTM to understand long-term data dependencies to improve stock price prediction and achieved a prediction accuracy of more than 95% in the S&P index.

3.3 Reinforcement Learning for Stock Market Prediction & Portfolio Management

Supervised learning approaches face challenges in stock market forecasting as they require labeled data for training in addition to difficulty in adapting to changing market conditions. RL offers solutions for those challenges through learning from interacting with surrounding environment and learning from consequences of its action rather than depending on labeled data. Learning through consequences empowers RL with the ability to adapt to surrounding environment changes quickly. Wu et al. (2023) proposes a hybrid stock market prediction approach that combines the unsupervised learning algorithm GNG, and RL algorithm to resolve traditional supervised learning challenges and improve stock market prediction. A proposed approach utilizing the GNG model based on unsupervised learning to group historical stock trading data. This approach allows the model to represent the environment using a limited number of states for trading. The proposed approach introduces a novel trading agent algorithm, Triple Q-learning, which is designed to execute the corresponding trading behavior and make comprehensive predictions of the stock market. The solution architecture for the proposed approach clarifies GNG clustering's role in generating environment's state and triple Q-Learning RL that updates the Q values table and takes the right decision that maximizes rewards.

Building upon strength of RL in handling dynamic environment, MARL emerged as promising research direction for enhancing stock market prediction. Shavandi and Khedmati (2022) propose a multi-agent DRL framework to perform algorithmic trading in the stock market. Agents are organized in a hierarchical structure in which each agent is specialized in a specific period, and the agent who specializes in a longer time frame shares knowledge with a lower time frame to reduce noise for agents trading in lower time frames. The DQN algorithm is implemented to train the multi-agent framework and single independent agents. A sample of the proposed framework consists of three agents, trading in long-term, mid-term and short-term. Experimental results show multi-agents are better than a single agent and existing benchmark trading strategies. Despite the advantages of the proposed approach, there are multiple limitations and future improvements including employing other DRL algorithms, such as policy-based approaches for training, investigating utilizing other agent's hierarchies and roles and evaluating framework on different combinations of time frames and datasets.

Although RL and MARL have a clear advantage in dealing with dynamic environments, training RL is a challenging problem considering it needs to tackle different dynamic situations that agents might encounter in real situations. To solve the training problem, He et al. (2023) propose a multi-agent virtual market model (MVMM) comprised of multiple generative adversarial networks (GANs) that cooperate to reproduce market price changes. MVMM helps by replacing real historical data with simulated market data which enables MARL to be trained over training data generated by MVMM simulating real stock prices which results in a profit increase of RL trained by MVVM by 12%. This also solves the problem of over-fitting for RL agents. Future improvement for the approach is to expand it to tackle more than one single asset. Use other RL methods like the value-based approach and actor-critic approach.

To optimize decisions for a group of stocks, Chakole et al. (2021) integrated clustering along with RL to predict stock behavior. This would support setting optimal decisions for groups of stocks that have similar behaviors. The proposed solution architecture contains clustering and RL parts on the cluster level. K-Means algorithm is used as a clustering approach. The trading agent is implemented using RL's Q-learning method. At the start, all Q-values were zero. The trading agent performs trading actions based on the Q-table and states in response to getting profit or loss as rewards, then updates the Q-value for each state-action pair of Q-tables based on rewards received for trading action. The trading strategy is decided based on just one day, which might not be optimal as things change from one day to another and a single day might not be enough to decide stock trends. Trading strategy identification only considers price-related information while it neglects news on the company along with external factors like prices increase or inflation or external laws or international prices for the same stocks. In addition to that, a comparison is needed to compare other RL methodologies such as deep Q-Network, policy gradient, deep deterministic policy gradients, and others to identify optimal methodology.

Although using RL and other ML techniques to predict optimal decisions regarding a single stock is an important research problem, the ability to work on multiple stocks and select a list of best stocks in terms of high revenue and low risk is an important research problem. Several researchers proposed solutions for portfolio optimizations (Jang & Seong, 2023; Gu et al., 2024). A Deep Reinforcement Learning (DRL) approach for stock portfolio optimization is proposed by Jang and Seong (2023), which integrates DRL with modern portfolio theory. The proposed approach by Jang et.al depends on deep deterministic policy gradient RL which combines the Q-learning RL technique and the Policy Gradients RL technique to train a RL agent to take the best action in a continuous action space. In addition to using deep deterministic policy gradients, it also uses tucker decomposition of a model with the input of technical analysis and stock return. experimental results applied on twenty-nine assets through merging technical indicators along with correlations between stocks shows significant improvement over existing legacy techniques. However, the proposed approach does not adjust the reward function to different outcomes and doesn't consider risk. Addressing risk gaps, the Pro Trader RL (Gu et al., 2024) framework provides a RL-based solution with separate buy and sell RL agents trained to mimic professional trader decisions and integrated with stop loss module to provide risk-adjusted trading strategy. Future research could enhance these models by integrating additional inputs like economic events and fundamental indicators for improved decision-making.

Lin et al. (2022) proposed an innovative approach for stock portfolio management considering risk shifting based on multi-agent DRL. The proposed approach is composed of a two-level nest RL agent structure. Input for the first layer is a pair representation of stocks, and output is a view based on the portfolio view vector of each local agent. Local RL agents are trained using a multi-agent deep deterministic policy gradient. The second level is composed of one global agent layer that processes all local agent decisions in the form of a view matrix as an input in addition to the current state. Global agents produce the final global portfolio weights considering risk, maximizing rewards, and most importantly transaction costs. Although experimental results show the proposed approach by Lin et al. is efficient in comparison to other existing approaches in terms of improved return ratio along with decreased risk, however there are a lot of limitations and future improvements. The first limitation is that input data considers only price-related information, therefore future research is recommended to consider other factors including technical, fundamental, and economic events that might affect stock profitability and trends. Experimental results and training are based on historical datasets, which might not accurately predict the future as future conditions might be different from historical stock market circumstances. Actual stock market data should be used for training and validation for future improvement.

Credit assignment is a challenging problem in MARL which employs centralized training and decentralized execution (Chen & Tan, 2023; Gronauer & Diepold, 2022;

Kapoor, 2018). It is important to define each agent's contribution in MARL, as a poor performing agent would decrease the overall performance of all agents leading to the lazy agent phenomenon. Chen and Tan (2023) propose Predictive contribution measurement, an explicit credit assignment method that compares prediction errors between agents and allocates rewards depending on their relevance to global state transitions for assigning credit for each RL agent. Researchers introduced a new approach to credit assignment in MARL called Predictive Contribution Measurement (PCM) (Zou & Xiao, 2024). PCM uses a predictive model to estimate the contribution of each agent to the team's reward. The predictive model is trained using a dataset of historical state-action pairs and rewards. The authors evaluated PCM on several MARL benchmarks, including StarCraft II and the GridWorld environment. The results showed that PCM outperforms other credit assignment methods, such as Q-learning and Actor-Critic. However, evaluating PCM on more real benchmarks is still important to better understand its performance, especially real stock market datasets. In addition, exploring the impact of different hyper-parameters on the performance of PCM will be interesting in future work.

Yang et al. (2023) propose a novel DRL framework for stock trading. The framework combines a transformer and U-Net networks to learn a single stock trading strategy. The transformer network is used to capture complex dynamic long-term dependencies in stock prices, while the U-Net network is used to capture short-term price fluctuations. The input of the model is a windowed stock price sequence, and the output consists of a trading action and action weight. The benefit of having two outputs is that the agent can control the share of buying and selling to reduce investment risk. As the policy and target network of the trading agent, the UTrans Network learns and generates trading signals (trading actions and action weights). The framework is evaluated on a dataset of historical stock prices, and the results show that it outperforms other DRL-based trading strategies. Yang et al.'s framework has several limitations. First, it can only handle a single stock, not multiple stocks. Second, while trained and evaluated on historical data, its performance in real-world trading conditions remains unclear. Third, training is computationally expensive. Finally, future research should explore incorporating more advanced attributes beyond price-related information to assess the impact of other factors on prediction accuracy.

3.4 Textual Analysis for Stock Market Prediction

Regardless of the specific ML or DL model employed, the effective analysis of textual data, such as stock market news articles and social media sentiment is increasingly critical for enhancing stock market prediction accuracy (Gangopadhyay & Majumder, 2023; Fazlija & Harder, 2022; Li & Pan, 2022; Bouadjenek et al., 2023; Li et al., 2020; Du et al., 2024b). Economic news and investors' sentiments have a big impact on stock market directions and stock prices. Khan et al. (2022) proposed ML classifier based on social media and news sentiment to predict stock market movements. Proposed approach results show that social media news sentiment enhances accuracy of stock market prediction based on ML. Mu et al. (2023) proposed a stock price prediction model that explicitly integrates investor sentiment with optimized

DL techniques. The proposed approach shows that investor sentiment can improve DL models ability in stock market forecasting.

Building on the understanding that sentiment analysis can significantly enhance stock market prediction, researchers have explored various hybrid techniques and approaches to improve its effectiveness. Berradi et al. (2022) explored integrating sentiment analysis for news with LSTM and attention mechanisms for forecasting stock prices. Proposed hybrid approach shows that considering sentiment analysis from sources like news and social media enhances the accuracy of predictive models however the limitation is ability to handle fake news. Li and Zhang (2023) introduced a RL model that integrates social media sentiment with historical stock prices to improve stock forecasting along with addressing unbalanced data classification. Abdelfattah et al. (2024) proposed an approach to enhance stock market movement prediction through improving sentiment analysis using neutrosophic logic. The Proposed model improves stock prediction accuracy through integrating sentiment scores derived from tweets with historical stock market data, using LSTM algorithm for forecasting. Das et al. (2024) surveys integrating sentiment analysis with Graph Neural Networks (GNNs) for stock prediction, and it shows that capturing market sentiment and investor behavior improve forecasting accuracy. However future improvements are needed in terms of considering risk management, improving scalability of handling large scale graphs, integrating with RL and including additional financial data in GNN in addition to sentiment. Finally, it's clear that textual news analysis and sentiment analysis have a major impact in stock market forecasting, and we do recommend combining it with other financial information types like technical indicators. In addition to find a proper method to give a weight for news according to impact and credibility.

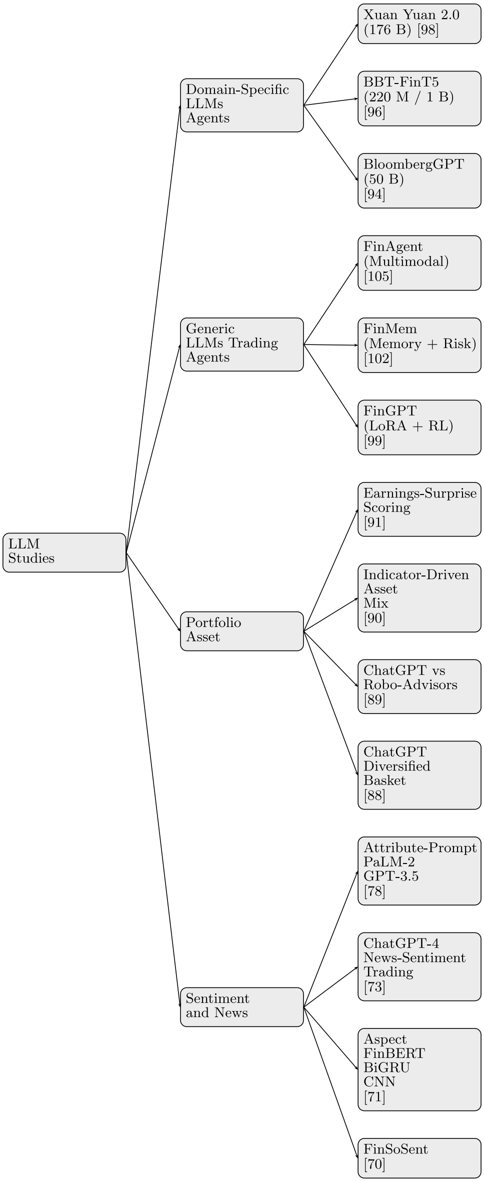
4 Large Language Models in Stock Market Prediction

Building on the advancements in Natural Language Processing (NLP) and textual analysis in stock market forecasting discussed in the previous section. Considering LLMs power in understanding complex relations and contextual information in textual information, LLMs offer a potential to improve stock prediction through better understanding of financial data. This section presents advancements and potential applications of LLMs in stock market prediction and research directions in financial LLMs. Figure 4 organises the LLM-based studies surveyed in this paper.

4.1 Introduction to LLMs and Pre-Trained Language Models

LLMs are a class of DL models that have revolutionized the field of NLP. These models are pre-trained on massive text corpora, enabling them to learn complex language patterns and representations. For example, GPT-3 (Brown et al., 2020) was trained on a massive dataset of 570 GB of text, encompassing a wide range of sources like books, articles, code, and websites. This pre-training process allows LLMs to develop a broad understanding of language and its nuances, which can then be adapted to specific domains like finance. LLMs are typically based on transformer architecture,

Fig. 4 Taxonomy of LLM-based studies surveyed in Section 4



a neural network architecture that relies heavily on attention mechanisms. The transformer, as introduced by Vaswani et al. (2017), allows for distributed processing of the input sequence which makes it faster and more efficient than RNN models. LLMs architecture has potential to be highly efficient for different NLP tasks, including language modeling, sentiment analysis, and question answering. The self-attention mechanism enables LLMs to capture complex relationships, long-range dependencies, and focus on the most relevant parts of the input text, which is important for efficiently understanding complex financial information.

LLMs like GPT-3 (Brown et al., 2020), GPT-4 (Achiam et al., 2023), FinBERT (Araci, 2019), and BloombergGPT (Yang et al., 2024) are pre-trained over a huge number of parameters, making them efficient in capturing language patterns, context, and performing complex tasks like sentiment analysis, news summary, company reports analysis, and trend prediction in financial markets. For example, GPT-3 is a general-purpose LLM trained with 175 billion parameters. FinBERT is a specialized pre-trained LLM designed for sentiment analysis trained over financial textual data and is based on 110 million parameters. BloombergGPT is a specialized financial LLM tailored for financial applications and trained with 50 billion parameters from general and financial information. Bloomberg's innovative approach of combining the power of LLMs with domain-specific financial knowledge is opening research towards more tailored LLMs specifically designed for finance and the stock market domain. In addition, LLMs capability for few-shot learning allows them to adapt to new information, which is critical for stock markets due to rapid changes that can dramatically and quickly alter stock trends and prices. GPT-4 extended GPT-3 capabilities with the ability to handle non-textual data in addition to enhanced reasoning skills, which has a potential for a better understanding of complex financial queries (Xue, 2024).

4.2 Applied Research Domains of Pre-Trained LLMs in Stock Market Domain

4.2.1 Financial Sentiment Analysis and Summarization

Recent research has explored the effectiveness of pre-trained LLMs in financial sentiment analysis. Delgadillo et al. (2024) proposed FinSoSent, a novel sentiment analysis model based on pre-trained LLMs. The proposed approach utilizes BERT architecture through pre-training on a huge corpus of financial news articles from Thomas Reuters, followed by fine-tuning on different social media datasets. Experimental results show that FinSoSent outperforms existing sentiment analysis models, including Amazon Comprehend, FinBERT, GPT-3.5-Turbo, and IBM Watson, albeit with modest improvements in accuracy. Researchers utilized ensemble learning through integration of FinSoSent with other sentiment models using soft voting and majority voting techniques. However, they achieved a modest improvement. Despite proposed model improvements, achieving high accuracy in financial sentiment analysis remains a challenge as accuracy is still below 60% for FinSoSent and other models. This indicates the need for further research to improve accuracy and tackle FinSoSent limitations such as the use of small fine-tuning datasets and the complexities of processing informal social media language, including slang, abbreviations, sarcasm, and incomplete information. Future research directions include utilizing larger LLMs with greater parameter capacity, developing multi-sentiment analyzers capable of identifying complex motivations behind social media posts and exploring the use of ChatGPT and generative AI to improve preprocessing complex social media language and give deeper understanding of social media text.

Traditional sentiment analysis in finance focus on overall sentiment of text. However, in certain cases financial text contains multiple sentiments emotions about different aspects therefore aspect-based sentiment analysis is considered an important research field especially in finance domain for a better understanding for textual news. Yan and Qin (Yan & Qin, 2024) utilize the power of pre-trained LLMs to handle aspect-based senti-ment analysis through proposing FinBERT-BiGRU-CNN model. This model combines FinBERT, Bidirectional Gated Recurrent Unit and CNN which enables the model to capture long-range dependencies in addition to local features in financial news. Pro-posed models utilize attention-pooling strategy to empower the model with ability to focus on important relative aspects. Experimental results show improved performance compared to existing base models with a F1 score of 0.87 for aspect-based sentiment classifications in a dataset of financial news texts form "THS iFund" and "East Money Choice". Future research is needed on real financial datasets along with techniques for handling social media language in addition to utilizing GPT-3 and GPT-4 LLMs for as-pect-based sentiment analysis.

Fatouros et al. investigated using zero-shot learning capability of ChatGPT 3.5 for financial sentiment analysis in the forex market (Fatouros et al., 2023). Their proposed approach depends on multiple ChatGPT prompts applied to news headlines sourced from Forex Live and FXstreet between January 2023 and May 2023. The researchers utilized a dataset outside of ChatGPT 3.5's training cutoff to ensure an unbiased evaluation of the LLM's ability to identify sentiment without fine-tuning. In addition to sentiment prediction, the proposed model calculated the correlation between predicted sentiment and market returns as an evaluation criterion. Experimental studies showed that the proposed ChatGPT approach always outperformed Fin-BERT, which is a well-known financial sentiment analysis pre-trained model, achieving a 35% enhancement in performance and a 36% higher correlation with market returns. These results present the potential of LLMs in improving financial sentiment analysis and market return prediction. Future research is recommended to explore the full potential of LLMs in the financial domain by applying the proposed approach to other datasets with longer timeframes and different financial markets, such as the stock market. Further investigation is needed to identify optimal prompts for various financial tasks, as ChatGPT prediction accuracy depends on the prompt design. In addition to prompt optimization, it is recommended to investigate other LLMs models based on GPT-4.

Lopez-Lira and Tang (2023) explored Chatgpt-4 model ability to predict expected stock price movements based on stock price information and news sentiments then design trading strategies according to predicted sentiment and stock movement. Proposed framework provides an explanatory framework that helps to explain factors impacted LLMs prediction decision. experimental results for applying Chatgpt-4 on news headline for U.S. common stocks from October 2021 to December 2023 shows that Chatgpt-4 LLMs manage to achieve cumulative return of 650% when applying investment strategy that depends on buying stocks with a positive ChatGPT-4 score and selling stocks with a negative ChatGPT-4 score. The second-best trading strategy is the short strategy that depends on selling stocks associated with negative sentiment with cumulative return of 300% during same period while the third best strategy is the long strategy that depends on buying stocks associated with a positive Chatgpt-4 score. This result shows that ChatGPT-4 prediction power is stronger in negative news than positive news and it also shows utilizing LLMs has big potential in financial sentiment and investment decisions with high return. Topic modeling and regression has been utilized to explain LLMs decision from textual data utilizing Cong et al. textual factors approach (Cong et al., 2024). Future research is needed to consider real-time prediction for stock market considering real-time conditions including transaction cost. Exploring integrating LLMs with ML techniques along with utilizing non-textual factors like technical indicators is another research direction to have a realistic complete prediction model with higher accuracy.

Xing (2024) proposes a heterogeneous LLM agent framework for financial sentiment analysis. This framework is designed according to Minsky's theory of mind and emotions (Minsky, 2007), where each agent specializes in a specific linguistic or financial aspect. Specialized agents focus on mood, rhetoric, or investor perspective. Specialized agents include those focusing on mood, rhetoric, dependency, aspect, and reference, in addition to agents simulating the perspectives of institutional and individual investors. A single summative agent aggregates the outputs of all specialized agents to generate the final financial sentiment. This design enables a more detailed and accurate analysis of financial sentiment compared to traditional methods. Each agent is assigned a specific prompt reflecting its purpose. The proposed framework is evaluated using three instruction-fine-tuned LLMs: GPT-3.5, BLOOMZ, and LlaMa3. It is also compared to three different multi-agent architectures: a voting-based homogeneous multi-agent architecture, a debate-based homogeneous multi-agent architecture (Du et al., 2023), and a voting-based heterogeneous multi-agent architecture. Experimental results show proposed framework improve accuracy compared to other frameworks and major improvements achieved with GPT-3 and LlaMa3 compared to BLOOMZ. Future research is needed to apply framework on new fresh datasets to mitigate the risk that evaluations datasets were part of LLM training in addition to checking impact of using different prompts along with exploring other techniques to achieve consensus between agents.

Du et al. (2024a) explores the reasoning capabilities of LLMs in Financial Sentiment Analysis. They identify six important financial attributes relevant to financial sentiment analysis namely semantic, numerical, temporal, comparative, causal and risk. Authors propose a financial attribute prompting framework that includes these attributes to guide LLMs to achieve high accuracy in financial sentiment analysis. Study shows that without explicitly following prompting framework, LLMs like PaLM-2 and GPT-3.5 have weakness in reasoning numerical. Other research studies investigated ChatGPT reasoning skills, and it shows it struggle in some tasks that require deep analysis (Gao et al., 2023). Bang et al. (2023) research indicating that LLMs capacity for reasoning in specialized do-mains like finance, often requires explicit prompting or fine-tuning. Zhang et al. improved LLM ability in financial sentiment analysis through combination of instruction-tuned prompts tailored for financial sentiment analysis and retrieval-augmented generation (RAG) to provide LLM with more context information from external sources (Lewis et al., 2020; Zhang et al., 2023). Therefore, Tailoring prompting strategies and utilizing techniques like RAG could enhance LLMs ability in financial sentiments and future research is needed to enhance LLMs reasoning mechanism and find enhanced techniques to improve LLMs accuracy including prompts optimization and the development of more sophisticated prompting frameworks (Mao et al., 2024).

Extracting useful information from textual news along with converting unstructured content into effective structured information is very important to increase financial sentiment accuracy. Several researchers explore LLMs capabilities in news summarizing, Kawamura et al. (2024) investigated prompt design impact on LLMs accuracy in generating financial news summaries and highlights that using leading role sentence in prompt improve accuracy of financial summary outcome in capturing important insights. Chhikara et al. (2024) introduced, LaMSUM, a multi-level approach that enhances summary extraction of user-generated content through integrating LLMs with voting algorithms which improves LLMs ability to extract and capture public sentiment from large dataset. Wan et al. (2024) utilized LLMs to provide text mining framework that automatically create and categorize large taxonomies for unstructured textual data which is critical for efficient financial sentiment analysis. Jin et al. (2024) presented a survey in automatic text summarization using LLMs which shows LLMs methods have advantages in quality and flexibility compared to traditional methods. Although LLMs have advantages in news summarization, there is a need to investigate effective prompt design since the quality of LLMs outcome depends heavily on prompt design.

4.2.2 Portfolio Management and Asset Selection

Lopez-Lira and Tang (2023) proposed LLMs based methodology to create groups of stocks and perform long or short investment decisions for selected stocks according to predicted returns based on financial sentiment analysis. However, it doesn't consider portfolio diversification and risk management. Ko and Lee (2024) investigated ChatGPT 3.5 ability to create dynamic diversified portfolio from different types of assets including stocks, bonds, cryptocurrencies, and commodities. The authors utilized multiple evaluation metrics to measure ChatGPT portfolio diversity including Simpson and Shannon diversity indices and average correlation. Experimental results show that ChatGPT portfolio has higher diversity, low correlation and low risk compared to randomly chosen assets. This implies ChatGPT and LLMs can understand relations between assets and associated risks. Although research is promising, there is a need to compare LLMs portfolio management with expert techniques for portfolio management. There is also a need for future research to explain LLMs decision making process as portfolio diversification depends on risk acceptance of investor which was not specified in prompt. Future research could focus on developing explainable AI techniques to make LLMs recommendations clear.

Oehler and Horn (2024) investigated ChatGPT ability to recommend personalized investment portfolios according to different investors risk profiles. Researchers compared ChatGPT recommendation outcome for different risk profiles with 17 robo-advisor globally available platform. Study considered three investor profiles, the first profile is a high-risk aversion with a loss tolerance of five percent, the second profile is a moderate risk aversion with loss tolerance ten percent, the last is low risk aversion with loss tolerance twenty-five percent. ChatGPT is also provided with other essential financial information needed for building investment portfolios that is also shared with robo-advisors including initial investment amount, investment horizon time, monthly net income and other. Experimental studies show that

ChatGPT provided better portfolios for one-time investment than robo-advisors. It shows ChatGPT recommendations align with investors risk profiles and compiled with a benchmark delivered from an academic study. It also shows that most robo-advisors failed to comply with the benchmark and only 3 out of the 17 robo-advisors meet with benchmark for the three risk profiles. Future research is needed to study the impact of input information to prompt on the outcome. Monitoring investment portfolio long-term actual performance against predicted profile is needed for reliable financial advice.

Kim (2023) explores the use of LLMs in recommendation of asset portfolio classes based on economic indicators. The study utilized ChatGPT to identify the most relevant economic indicators impacting stock returns and identify most relevant asset classes according to different economic conditions. Therefore, two unique prompts are used for the two different use cases. Portfolios constructed based on ChatGPT recommendation were back tested and compared to benchmark portfolios. Results show that ChatGPT portfolios outperformed benchmark portfolios especially for out-sample portfolios recommendations on data outside ChatGPT training period from October 2022 to July 2023. This indicates LLMs ability to enhance asset allocation, generalize to unseen market conditions and understand relation between economic factors and asset classes. Study depends on ChatGPT recommendations for 3 indicators and 3 asset classes. However, future research is needed to consider additional technical indicators and a dynamic number of asset classes, rather than specifying a fixed number. The current study focuses on portfolio optimization and rebalancing outside the scope of LLMs, so future research is recommended to explore these topics using LLMs.

Pelster and Val (2024) studied the potential of LLMs to assist in stock selection by analyzing stock financial information to predict earnings surprises and stock returns. The study employed ChatGPT-4 to predict upcoming stock earning compared to analyst expectation and to rate stock attractiveness according to recent stock news and earning expectation. The study employed a live experiment during the 2023 Q2 earnings season which is outside of ChatGPT training period and recent stock information collected from different internet sources and augmented to ChatGPT prompt to make sure ChatGPT model has latest stock news. Experimental results show a positive correlation between ChatGPT rating and subsequent stock performance. It also shows ChatGPT can adjust stock ratings as a response to new negative or positive news. The authors created and back-tested equally weighted portfolios over the next month according to ChatGPT attractiveness ratings, the portfolios with higher ratings achieve higher returns before end of the sample period. Experimental results indicate ChatGPT has the capability to analyze stock news and convert it into a score representing expected stock performance for next month. Although ChatGPT provides promising results in stock selection, it has limitations regarding black box nature, lack of transparency and factual accuracy that might lead to different response over time so future research is needed to address these limitations (Chen et al., 2023).

4.2.3 LLMs Trading and Financial Agents

Instead of employing general LLMs to address the complexity of finance, several researchers investigated developing specific LLMs tailored for the finance sector. According to a study by Li et al. (2023), these models perform better when handling tasks in the finance area. Wu et al. (2023) proposed BloombergGPT which represents a 50-billion parameter specialized financial LLM. BloombergGPT is a decoder-only causal language model based on BLOOM (Le Scao et al., 2022). It is trained on 363 billion tokens from Bloomberg's proprietary financial datasets and 345 billion tokens from general purpose datasets. Experimental studies show that BloombergGPT outperforms existing general purpose LLMs on financial tasks including sentiment analysis and question answering. This success is a result of mixed training strategy that allows models to learn financial aspects along with general purpose aspects as well. Although BloombergGPT shows promising results in financial tasks, there are limitations as it is not publicly available due to proprietary nature and in addition to that it may not generalize beyond Bloomberg's financial dataset. These limitations align with discussion in Li et al. (2023) survey about the limitations of accessing and fine-tuning certain LLMs due to data privacy and security concerns.

Like BloombergGPT, Lu et al. (2023) introduced BBT-FinT5, a Chinese financial LLM trained on a 300 GB corpus with 220 million parameters for the base version and 1 billion parameters for the large one. BBT-FinT5 outperforms general-purpose models on Chinese financial tasks. While BloombergGPT utilizes a mixed dataset and is based on the BLOOM architecture, BBT-FinT5 is based on the T5 architecture (Raffel et al., 2020) and trained solely on financial data from multiple Chinese financial datasets. This difference in training data and model architecture may lead to variations in performance and capabilities between the two models however comparison is not doable since BloombergGPT is not open source and since BBT-FinT5 is trained only on Chinese financial datasets. In another approach, Zhang and Yang (2023) proposed Xuan Yuan 2.0, a 176 billion parameter decoder-only model based on BLOOM architecture for the Chinese financial domain. Unlike BloombergGPT and BBT-FinT5, which focus on various financial NLP tasks, Xuan Yuan 2.0 is specifically designed for conversational AI applications in Chinese finance. Like BloombergGPT, Xuan Yuan 2.0 uses mixed datasets composed of general datasets and financial specific datasets. Xuan Yuan 2.0 specialization in Chinese financial chat and mixed training enables it to have better performance in financial question answering and dialogue generation and interactive personal financial advice in Chinese domain.

Although specialized financial LLMs like BloombergGPT shows improved performance in financial domain compared to general LLMs however training process requires high computational resources and cost. For Example, BloombergGPT requires 1.3 million GPU hours for training which indicates a huge cost for retraining models from scratch (Wu et al., 2023). Furthermore, specialized financial LLMs don't fully reuse the capabilities of existing pre-trained LLMs that are trained over billions of parameters. To address these challenges, FinGPT provides open source, real-time, data-centric framework for financial LLMs that utilize the power of pre-trained general-purpose LLMs (Yang et al., 2023). FinGPT employs Low-Rank Adaptation (LoRA) (Hu et al., 2021; Dettmers et al., 2023) to fine-tune LLMs for financial tasks using labeled stock market dataset which reduces the number of trainable parameters 6.17 billion to a mere 3.67 million and making the model update process faster and less resource-intensive. FinGPT utilizes RL from Stock Prices to adapt model behavior according to market response to financial news and previous model predictions. FinGPT also tackles the problem of high temporal sensitivity and low signal-to-noise ratio in financial data through real-time data gathering from multiple financial data sources in addition to real-time data processing to filter noise. FinGPT can handle different financial tasks including quantitative trading, portfolio management, financial sentiment analysis, personalized investment advice, and others.

Although LLMs like FinGPT (Yang et al., 2023) and BloombergGPT (Wu et al., 2023) show promising results in the financial domain, they have limitations in understanding the time sensitivity of events and the impact of risk profiles on investor decisions. Yu et al. (2024) proposed FinMem, an LLM-based autonomous financial stock trading agent designed to address these limitations. FinMem builds upon the memory concept introduced by Park et al. (2023), who proposed a hierarchical memory for improving language model performance in tasks requiring long-term context. FinMem enhances this concept by introducing a working memory for immediate tasks and a layered long-term memory to hold important events with long-term impacts. It is also equipped with three risk profiles, risk-seeking, risk-averse, and self-adaptive. FinMem allows dynamic adjustment for risk according to market circumstances, a feature not fully addressed in previous research on risk-aware LLMs (Oehler & Horn, 2024; Kim, 2023). FinMem achieves higher cumulative returns and accuracy compared to other algorithmic agents, including DRL-based agents and other LLM-based agents, on a real-world financial dataset. It also has superior performance with limited training data and short timeframes unlike DRL models that require extensive training data (Dang, 2019). Furthermore, FinMem offers better transparency and interpretability because it allows access to memory layer content, which provides a better understanding of the important reasons behind LLM decisions, unlike RL and DRL, which lack interpretability.

Zhang et al. introduced FinAgent (Zhang et al., 2024), the first multimodal LLM agent for financial trading that is capable of processing different data sources including numerical, textual and visual information utilizing recent LLM capabilities of processing different data types including visual information (OpenAI, 2023). Unlike FinGPT (Yang et al., 2023) and FinMem (Yu et al., 2024) which primarily focus on numerical and textual data, FinAgent processes various types of data including trading charts and visual input. FinAgent utilizes memory concept previously utilized by FinMem and research of Yu et al. (2024); Park et al. (2023) to learn from historical information and improve decision making process. FinAgent memory includes three main components, market intelligence memory which is used to store market information for trend analysis, Low-Level Reflection memory to analyze relationships between market observations and resulting price movements, and high-level reflection memory to store and retrieve reflections of past trading decisions to adapt trading strategies. FinAgent includes a tool-augmented decision-making module that analyzes information in different memory components in addition to external information from expert knowledge and traditional trading strategies to improve prediction accuracy. Experimental results on six financial datasets show that FinAgent outperforms other trading agents including LLM-based agents like FinGPT and FinMem, Other ML models, DL models including LSTM (Yang et al., 2020), LGBM (Yang et al., 2020), Transformer (Yang et al., 2020), RL agents (Haarnoja et al., 2018; Schulman et al., 2017a) and traditional strategies with 36% average improvement on profit.

Explaining prediction decisions and rationale behind LLM financial agent decisions is necessary for building trust in prediction outcome. Yu et al. improve LLM interpretability in their FinMem agent by utilizing a layered memory module that allows for examination of the financial information stored in it along with tailoring prompt to provide underlying rationale behind financial decision (Yu et al., 2024). This provides insight into the agent's reasoning process, but it doesn't offer detailed explanations. FinAgent (Zhang et al., 2024) utilizes a Chain-of-Thought (COT) (Wei et al., 2022) prompting strategy to explicitly provide reasoning steps from the LLM, making its decision-making process more transparent. Both FinAgent and FinMem rely on inherit LLM model capabilities for reasoning using tailored prompting techniques. Lopez-Lira and Tang (2023) utilized topic modeling and regression to explain LLM agent financial decision through presenting relation between LLM decision and main factors behind LLM decision. Koa et al. (2024) proposed the Summarize-Explain-Predict (SEP) framework, which utilizes a self-reflective LLM agent capability and Proximal Policy Optimization (PPO) (Schulman et al., 2017b) to improve LLM interpretability. SEP summarizes information then use LLM to generate explanations for financial decision predictions then LLM is fine-tuned using PPO to produce confidence-based predictions. This approach shows RL can be integrated with LLM to improve interpretability and stock prediction accuracy.

Several of the financial LLM agents we surveyed already embed interpretability into their core architectures. FinMem uses a two-stage "working + long-term" memory to surface key past events when justifying a trade, while FinAgent layers reflection memories to record and recall rationale steps. Beyond memory, Chain-of-Thought prompting enables LLMs to generate human-readable reasoning chains for predictions, and the SEP framework fine-tunes models to output both a forecast and a confidence-scored explanation. Together, these approaches move LLMs in finance from black-box predictors toward systems whose decisions can be traced and audited.

5 Analysis & Discussion

This literature review has explored different recent feature engineering techniques for stock market prediction, ranging from traditional methods like LASSO regression to more advanced optimization algorithms, ensemble methods and hybrid models. Table 1 presents a summary of recent feature selections research in stock market forecasting along with key strengths and limitations. Although feature selection techniques offer improved prediction accuracy along with reduced complexity by selecting relevant and important factors, challenges remain in generalizability, achieving high accuracy across diverse stock market datasets and considering features in unstructured data. Several studies focus on technical indicators and price related information neglecting other valuable stock information including fundamental or macroeconomic and other features exist in unstructured data source like social media news and news articles.

Table 1 Summary of recent feature selections research in stock market forecasting

Research	Data Source	Selection Method	Key Strengths	Challenges & Research Gaps
Xu et al. (2023)	Technical and Fundamental Indicators for Chinese Airline	LASSO	Better Prediction Accuracy on Chinese airlines	Tested only on Chinese Airlines; limiting model generalization
Kumari and Swarnkar (2023)	Eighty-Three Technical Indicators	Ensemble Methods [LASSO, IG, FFS]	Better Stock Prediction accuracy on 6 different stock indices	Ignores non-technical features; Limited indices tested
Chaudhari and Thakkar (2023)	Stock Price; Technical Indicators; Google Trends	CV+K-Means, CV+Median, CV+top-k	Better Performance on most datasets and faster execution time	Ignores non-technical features; Finding an optimal number of features is a challenge
Htun et al. (2023)	Structured Input features	Different Methods	Highlight important research directions of feature selection including ensemble methods	Focus only on structured input.
Patil et al. (2024)	Historical Stock data for two companies	SSO and CSA	Enhances accuracy with hybrid CNN-LSTM and BSO optimization.	Limited to specific datasets; lacks broader testing and comparison.
Su et al. (2023)	Taiwan stock exchange (10 years)	New proposed Method	Enhances accuracy with feature selection and ensemble methods.	Limited to specific datasets; lacks broader testing and comparison.
Li et al. (2024)	Stock Price; Technical Indicators	Boruta and BSO	Better accuracy with ensemble feature selection and denoising.	Limited to specific datasets; lacks broader testing and comparison.
Zhao et al. (2024)	Stock Price; Trading Volume from two datasets	CNN	Uses CNN, BLSTM, and attention mechanism for better accuracy	Poor performance on some datasets due to limited feature extraction

Designing an effective feature selection mechanism that deals with different structure and unstructured data types along with achieving a consistent high prediction accuracy in cross stock markets and different stock datasets is still a challenging research problem. Several research has been conducted that shows LLMs ability to transform unstructured data into meaningful structured data and summarize text which reflects how LLMs can enhance feature selection limitations in dealing with unstructured data and social media news (Kawamura et al., 2024; Chhikara et al., 2024; Achiam et al., 2023; Jin et al., 2024).

Table 2 presents a summary of the key trends and challenges in stock market prediction using ML, DL and RL. Recent advancements in predictive modeling using ML and DL show better prediction accuracy compared to traditional technical and statistical stock market prediction methods. Figure 5 presents accuracy and profit improvement percentages reported by six representative studies, allowing a direct comparison.

LSTM shows high accuracy in stock price prediction, while hybrid models and ensemble methods that integrate different ML, DL and optimization methodologies show potential improvements. Text-based prediction also shows potential improvement utilizing sentiment analysis from social media and news to enhance prediction accuracy. However, challenges such as handling fake or inaccurate information along with sentimental ambiguity. The development of cross-market prediction models

Table 2 Summary of key trends and challenges of traditional predictive models in stock market forecasting

Category	Studies	Methodologies	Key Trends and Findings	Challenges and Research Gaps
Predictive Modeling using ML and DL	Kumbure et al. (2022); Kehinde et al. (2023); Ashtiani and Raahemi (2023); Guennioui et al. (2024); Beniwal et al. (2024); Safari and Badamchizadeh (2024); Ma et al. (2024); Phuoc et al. (2024); Tatane et al. (2024); Md et al. (2023)	ML Classifiers, DL Classifiers like SVM, ANN, CNN, RNN, LSTM, etc.	● Different ML and DL techniques show potential better stock prediction compared to statistical and technical methods. ● DL, particularly LSTM, achieves high accuracy in stock price prediction. ● There is a growing trend towards hybrid and ensemble learning models.	● Lack of model explainability. ● Models should consider different data sources (technical indicators, news, social media, and investor sentiment). ● Limited cross-market evaluation; challenge in creating models that perform well across markets. ● Hybrid and ensemble learning for performance improvements.
Text-Based Prediction	Picasso et al. (2019); Jing et al. (2021); Ashtiani and Raahemi (2023); Gangopadhyay and Majumder (2023); Fazlija and Harder (2022); Li and Pan (2022); Bouadjenek et al. (2023); Li et al. (2020); Du et al. (2024b); Khan et al. (2022); Mu et al. (2023); Berradi et al. (2022); Abdelfattah et al. (2024); Das et al. (2024)	ML and DL classifiers for textual analysis, including RNN, LSTM, attention, GNN, and Transformers, along with Neutrosophic logic	● Integrating sentimental data from social media and news with ML/DL models enhances stock prediction accuracy. ● There is a growing focus on utilizing language models for improvement.	● Lack of model explainability. ● Address challenges related to handling fake and inaccurate news, sentiment ambiguity, risk management, and scalability. ● Explore utilizing advanced language models.
Hybrid and Ensemble Methods	Jing et al. (2021); Patil et al. (2024); Nti et al. (2020); Zhang and Chen (2024); Kianrad et al. (2024); Zou and Xiao (2024); Liangchao (2024); Yeo et al. (2023); Ma et al. (2024); Bagalkot et al. (2024); Mulla et al. (2024); Wu et al. (2023); Khan et al. (2022)	Combine statistical, technical, ML, and DL techniques, like ARIMA, Random Forest, SVMs, XGBoost, LightGBM, CNN, LSTM, and others, along with optimization algorithms (Genetic Algorithm, Particle Swarm, etc.)	● Both Ensemble Learning with optimization techniques and hybrid models combining multiple approaches improve stock prediction accuracy and investment decisions compared to single model/method.	● Lack of model explainability. ● Cross Market Prediction Model. ● Compare different ensemble/hybrid model combinations across different market conditions. ● Improve accuracy across different datasets, parameters, and market indices.

Table 2 (continued)

Category	Studies	Methodologies	Key Trends and Findings	Challenges and Research Gaps
Reinforcement-Based Prediction	Sun et al. (2023); Huang et al. (2024); Yeo et al. (2023); Wu et al. (2023); Shavandi and Khedmati (2022); He et al. (2023); Chakole et al. (2021); Jang and Seong (2023); Gu et al. (2024); Lin et al. (2022); Chen and Tan (2023); Gronauer and Diepold (2022); Kapoor (2018); Yang et al. (2023); Li and Zhang (2023); Dang (2019)	Q-learning, Deep Q-learning, Deep Deterministic Policy Gradient, Multi-agent Deep Deterministic Policy Gradient, Predictive Contribution Measurement	• RL and MARL can deal with surrounding changes in stock markets and can consider different factors through utilizing multi-agent concepts leading to improvement in stock trading profitability and portfolio management optimization.	• Lack of model explainability. • Challenges in training RL agents due to extensive training data needed. • Evaluate and compare different RL methods and MARL credit assignments solutions on real conditions across multiple datasets and different stock markets.

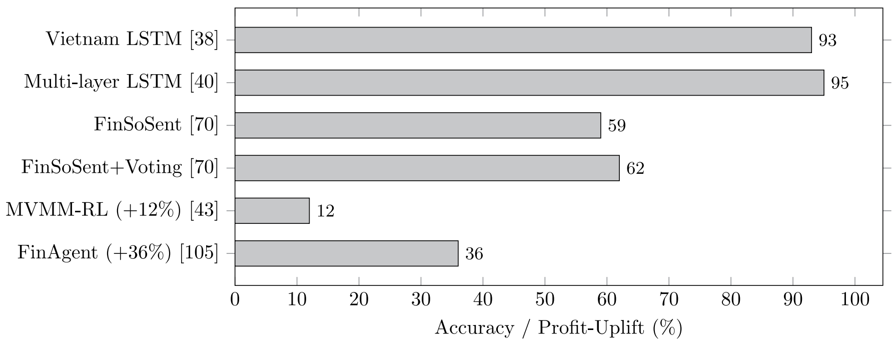


Fig. 5 Headline accuracy or profit Improvement percentages reported by six research studies

remains a challenge for all research categories, as existing models often lack generalizability across different datasets and markets. RL and MARL show potential improvement in dealing with stock market complexity leading to better stock trading accuracy and enhanced portfolio management. Despite these advancements, the extensive training data requirements for RL and credit assignment for MARL models are a challenge. In addition to previous challenges, ML, DL, and RL model decisions lack explainability. LLMs promise to handle these challenges by enhancing model explainability through chain-of-thought reasoning, improving textual and context understanding through understanding complex structures, which improve ability to handle cross-market analysis and different stock datasets.

Table 3 presents a summary of recent advances in LLM-based sentiment analysis that cover key strengths and research gaps. Recent LLMs research shows promising results for LLMs in handling financial sentiment analysis. Delgadillo et al.'s BERT-based FinSoSent model outperforms traditional models but faces challenges with informal language. Yan and Qin's FinBERT-BiGRU-CNN integrates FinBERT

Table 3 Summary of key strengths and challenges of LLM-based financial sentiment analysis approaches

LLM Approach	Data Source	Key Strengths	Challenges and Research Gaps
Delgadillo et al. (2024)	Financial news articles from Thomas Reuters and social media	• Proposes FinSoSent, a novel sentiment analysis using LLMs and ensemble learning. • Outperforms existing sentiment analysis models.	• Modest accuracy improvement. • Limited by small fine-tuning datasets. • Limited by informal language challenges.
Yan and Qin (2024)	Financial news texts from ”THS iFund” and ”East Money Choice”	• Uses LLMs for aspect-based sentiment analysis. • Outperforms base models, capturing long-range dependencies.	• Limited evaluation on real financial datasets. • Needs to address social media language. • Needs to utilize LLMs.
Fatouros et al. (2023)	Forex news headlines from Forex Live and FXstreet	• Explores ChatGPT's zero-shot learning for financial sentiment analysis. • Outperforms FinBERT in accuracy and return correlation.	• Limited to forex news. • Limited to a specific timeframe. • Requires evaluation on non-forex datasets.
Lopez-Lira and Tang (2023)	News headlines for U.S. common stocks	• Explores ChatGPT-4 for stock prediction and trading strategies using sentiment analysis. • Achieves high cumulative returns with long/short strategies.	• Black box nature and lacks transparency. • Consider real-time prediction and transaction costs.
Xing (2024)	Financial sentiment from news, social media, and crypto forums	• Provides financial sentiment analysis based on Minsky's theory. • Improved accuracy compared to other frameworks.	• Evaluation dataset may be part of LLM training data. • Needs to explore different prompts and consensus techniques.
Du et al. (2024a)	Financial sentiment from news and twitter	• Proposes a financial attribute prompting framework to guide LLMs in sentiment analysis. • Improved reasoning capabilities and addresses limitations in numerical reasoning.	• Requires further fine-tuning. • Requires prompt engineering for specific tasks. • Requires cross-market validation on different datasets.

with CNN to effectively capture long-range dependencies, however, limited dataset diversity restricts its evaluation and the need to use larger LLMs. Fatouros et al. utilize ChatGPT-3.5's zero-shot capability, achieving high forex sentiment accuracy, but the approach is context-specific. Lopez-Lira et al. utilize ChatGPT-4 for utilizing sentiment analysis for trading strategies, showing high returns however, lack of model transparency and real-time market considerations remain challenges. Xing et al. introduce a heterogeneous LLM agent framework that improves sentiment accuracy by using specialized agents; however, future enhancements in prompt techniques are needed. Finally, Du et al.'s attribute prompting framework advances LLMs in reasoning but needs further fine-tuning and prompt engineering. Overall, each approach highlights LLM advancements in financial sentiment analysis, however future research is needed to enhance model transparency, prompt engineering methodologies and cross-market evaluation to overcome biases and secure model's generalizability.

Table 4 summarizes recent LLM approaches in portfolio management, covering strengths, challenges, and factors like risk management, asset classes, and evaluation

Table 4 Summary of key strengths and challenges of LLM-based financial portfolio management approaches

LLM Approach	Asset Class	Risk Management	Evaluation	Key Strengths	Challenges and Research Gaps
Lopez-Lira and Tang (2023)	Stocks	Not considered	• Sharpe ratio • Mean daily returns • Standard deviation daily returns • Maximum drawdown	• Early exploration of LLM-based stock selection using sentiment analysis.	• No portfolio diversification or risk management. • No cross-market evaluation or real-world comparison. • Lacks transparency.
Ko and Lee (2024)	• Stocks • Bonds • Cryptocurrencies • Commodities	Considers portfolio diversity and correlation	• Simpson and Shannon diversity indices • Average correlation	• Explores LLMs for diversified, low-risk portfolios across asset classes.	• Lacks explanation of decision-making. • Needs comparison with expert techniques.
Oehler and Horn (2024)	• Stocks • Bonds • Others	Considers investor risk profiles (high, moderate, low)	• Comparison with robo-advisors • Compliance with academic benchmark	• Compares LLM portfolios with robo-advisors and suggests personalized, risk-aligned portfolios.	• Needs long-term performance monitoring. • The impact of prompt input needs further study.
Kim (2023)	Asset classes based on economic indicators	Not explicitly addressed	• Backtesting of portfolios • Comparison with benchmark portfolios	• Shows LLM's ability to recommend asset classes based on economic indicators and market conditions. • Identifies relevant indicators and asset classes, outperforms benchmark portfolios.	• No cross-market evaluation or real-world comparison. • Limited number of indicators and asset classes as they considered only 3. • No portfolio optimization within LLM.
Pelster and Val (2024)	Stocks	Not explicitly addressed	• Correlation between ratings and stock performance • Backtesting of portfolios	• Shows LLMs ability to predict stock returns, earnings surprises and adjusting recommendation based on recent negative or positive financial news.	• Black box nature and lack of model transparency. • Potential factual inaccuracies.

metrics. LLMs research LLM research shows promising results in stock selection, asset allocation, and diversification. Lopez-Lira et al. used LLMs for stock selection through sentiment analysis but lacked diversification across asset classes and risk management. Ko et al. expanded the scope to include stocks and other assets for improved portfolio diversity but did not clarify their decision-making process or compare results to expert methods. Ko et al. also used LLMs to create diversified portfolios based on correlation and risk tolerance. Oehler et al. compared LLM-generated portfolios to robo-advisors, showing LLMs' potential for personalized investment strategies. Kim demonstrated LLMs' ability to recommend asset classes based on economic indicators, reflecting their reasoning capabilities. Pelster et al. used LLMs for predicting stock returns and earnings surprises, highlighting their advanced reasoning. These approaches show significant progress in LLM-based portfolio management, but challenges remain, such as model transparency, limited evaluation, generalizability across market conditions and datasets, and prompt engineering bias.

Tables 5 and 6 present a detailed analysis of LLM-based financial agents, covering key strengths, limitations, and other factors. Several studies have explored using LLMs for stock market prediction and other financial tasks. There is a trend toward specialized financial LLMs, such as BloombergGPT and BBT-FinT5, which are trained on extensive financial datasets. These models show improvements over general LLMs, but face challenges such as high training costs, bias in proprietary data (e.g., BloombergGPT), and language limitations (e.g., BBT-FinT5 only supports Chinese). Xuan Yuan 2.0 focuses on conversational finance within the Chinese market but lacks adaptability to other financial tasks. Another trend is customizing existing LLMs for stock market tasks, like FinGPT, which uses LoRA and RL. Incorporating memory and explainability to improve decision-making and transparency, as seen in FinMem, FinAgent, and Koa et al.'s work, is also a key trend. The use of RL to autonomously refine decision-making without human intervention, seen in FinGPT and Koa et al., is gaining traction. Overall, these trends show progress in financial LLMs, but challenges remain, such as the need for better model generalizability, evaluation frameworks, data quality, and techniques to enhance transparency and automate fine-tuning for improved financial decision-making.

In summary, Table 7 and Figure 6 together explain the four capability dimensions covering explainability, ability to handle cross market, ability to handle unstructured data and training cost into a comparison of strengths and weaknesses across approaches categories. traditional feature-engineering methods score one high in training cost as they require low computational cost but they have challenges in handling cross-market and unstructured-data (Xu et al., 2023). ML/DL approaches have medium ability in handling unstructured-data (Berradi et al., 2022) however it has low explainability due to black box nature (Beniwal et al., 2024) and it has expensive computational training cost (Md et al., 2023); RL/MARL further improves cross-market robustness (Shavandi & Khedmati, 2022) but still suffers from poor explainability (Wu et al., 2023) and long training periods and costs (He et al., 2023). LLMs—with domain-specific models excels in explainability (Yu et al., 2024) and unstructured text comprehension (Lu et al., 2023) at the expense of huge computational cost for training that reaches multi-million–GPU-hour per training (Wu et

Table 5 Summary of key strengths and challenges of LLMs financial trading agents

LLM Agent	Type	Memory	Explainability	Key Strengths	Challenges and Research Gaps
Wu et al. (2023)	BLOOM	Not explicitly mentioned	Not explicitly mentioned	● Specialized finance LLM with 50B parameters. ● Trained on extensive Bloomberg financial dataset and general datasets. ● Uses unigram tokenization to reduce tokens.	● Limited accessibility as it's not publicly available. ● Potential bias from proprietary data. ● Investigate other tokenization strategies and task fine-tuning.
Xuan Yuan 2.0 (Zhang & Yang, 2023)	BLOOM	Not explicitly mentioned	Not explicitly mentioned	● LLM specifically designed for Chinese finance chat based on BLOOM-176B. ● Utilize novel hybrid-tuning to combine general LLMs domain knowledge with specific new trained domain knowledge.	● Limited to Chinese language and conversational applications. ● Lack solid evaluation and comparisons with other LLMs. ● Needs to consider other financial tasks and other languages.
BBT-FinT5 (Lu et al., 2023)	T5	Not explicitly mentioned	Not explicitly mentioned	● Specialized LLM for Chinese financial domain based on the T5 architecture with one billion parameters. ● Outperforms general-purpose models on Chinese financial tasks.	● Limited to Chinese financial data and Chinese language. ● Limited to textual data source. ● Needs to consider multi-language, multimodal applications and increasing training corpus.
FinGPT (Yang et al., 2023)	Pre-Trained LLMs with LoRA	Not explicitly mentioned	Not explicitly mentioned	● Open-source real-time framework for financial LLMs using pre-trained models, LoRa, and RL. ● LoRa reduces trainable parameters from 6.17B to 3.67M. ● RL enables autonomous model learning during fine-tuning.	● LoRA may risk overfitting or lack generalizability. ● Needs evaluation on diverse datasets and markets compared to other financial agents. ● Inherits any limitation in base pre-trained LLMs model.

Table 5 (continued)

LLM Agent	Type	Memory	Explainability	Key Strengths	Challenges and Research Gaps
Fin-Mem (Yu et al., 2024)	GPT-4-LLM	Layered Hierarchical Memory	Through accessing memory contents	● LLM trading agent with hierarchical memory, dynamic risk adaptation, and enhanced explainability. ● Achieves high returns with limited data, outperforming DRL agents and FinGPT.	● Highly dependent on data quality which complicates training in dynamic stock market. ● Needs considering multi-agent trading for portfolio optimizations. ● Lack cross-market validation for generalizability.
FinAgent (Zhang et al., 2024)	GPT-4 with memory	Multi-component memory (market intelligence, reflection)	Chain-of-Thought (COT) prompting	● First multimodal LLM agent for financial trading with memory and tool-augmented decision-making. ● Uses Dual-level Reflection Module for autonomous learning.	● Limited evaluation on small stocks and datasets. ●Model generalizability may be challenged by cryptocurrencies' lower returns. ● Focused on single stocks, without considering risk types.
Koa et al. (2024)	GPT-3.5 LLM	Not explicitly mentioned	SEP framework	● Self-reflective LLM agent with RL and SEP framework for improved explainability. ● No human intervention needed; RL corrects predictions based on outcomes. ● Outperforms traditional DL and LLM methods in accuracy and correlation. ● Generalizes to other tasks like portfolio construction.	● No framework for evaluating LLM explanation quality and LLM hallucinations. ● Tested on tweets, which may contain fake news affecting predictions. ● Limited to textual data from tweets and historical prices. ● Restricted to specific prompts; further exploration needed.

al., 2023). generic/adapter LLMs that match LLMs- with finance domain specific strengths in explainability and handling unstructured text while reducing cost to a medium level via LoRA or prompt tuning (Yang et al., 2023). General LLMs with finance adaptation emerge as the only sub-family without any low-rated criteria. This comparison underscores the potential of LLMs to address some limitations of other methods. In short, modern LLM pipelines inherit the language-understanding strengths of transformers while adapter techniques mitigate the expensive compute costs that limits domain-specific models.

Table 6 Applications and evaluation of LLM financial agents

LLM Agent	Dataset	Modality	Evaluation Metrics	Financial Application
Bloomberg-GPT (Wu et al., 2023)	Bloomberg's proprietary financial dataset, general-purpose datasets	Textual	Accuracy, F1-score on financial benchmarks	● Sentiment analysis News summarization ● Question answering ● Risk assessment ● Trading strategy generation
BBT-FinT5 (Lu et al., 2023)	300GB corpus of Chinese financial datasets	Textual	Accuracy, F1-score on Chinese financial benchmarks	● Chinese financial NLP Tasks
Xuan Yuan 2.0 (Zhang & Yang, 2023)	Mix of general and Chinese financial datasets	Textual	Accuracy, BLEU score on conversational tasks	● Financial question answering ● Dialogue generation
FinGPT (Yang et al., 2023)	Labeled stock market dataset	Textual	Profitability, Sharpe ratio, other trading performance metrics	● Quantitative trading ● Portfolio management ● Sentiment analysis
FinMem (Yu et al., 2024)	Real-world financial dataset	Textual	Cumulative returns, accuracy, risk measures	● Autonomous stock trading ● Risk-aware portfolio management
FinAgent (Zhang et al., 2024)	Six financial datasets (including numerical, textual, and visual data)	Multimodal (numerical, textual, visual)	Profitability, Sharpe ratio, accuracy	● Multimodal financial analysis ● Trading strategy development
Koa et al. (2024)	Financial Tweets and prices datasets	Textual	Accuracy, Matthews correlation coefficient	● Stock market prediction ● Portfolio optimization

Table 7 Qualitative capability ratings (Low = 0, Med = 1, High = 2) with supporting references

Approach Category	Explainability	Cross-Market	Unstructured	Training Cost
Feature Engineering	Med[14]	Low[12]	Low[12]	**High**[12]
ML / DL	Low[31]	Low[9]	Med[60]	**Low**[40]
RL / MARL	Low[41]	Med[42]	Low[41]	**Low**[43]
LLM – Domain-Specific	High[94]	Med[94]	High[96]	**Low**[94]
LLM – General with Adaptation	High[102]	Med[73]	High[70]	**Med**[99]

Ratings are ordinal: **Low** = weakness, **Med** = neutral, **High** = strength. In the "Training Cost" column specifically, a **High** rating denotes a favorable (low) compute expense (e.g., LoRA-style fine-tuning), whereas a **Low** rating denotes an unfavorable (high) expense (e.g., multi-million–GPU-hour per training)

6 Research Directions

This section discusses potential directions for future research according to the discovered gaps in the reviewed literature as summarized in tables from 1 to 7.

6.1 Cross-Market Generalization & Evaluation

A recurring limitation across the surveyed studies (Tables 2, 3, 4 and 5) is that nearly every model is trained and tested on a single market or stock usually US or chinese exchanges which limits the generalizability of research findings (Jing et al., 2021;

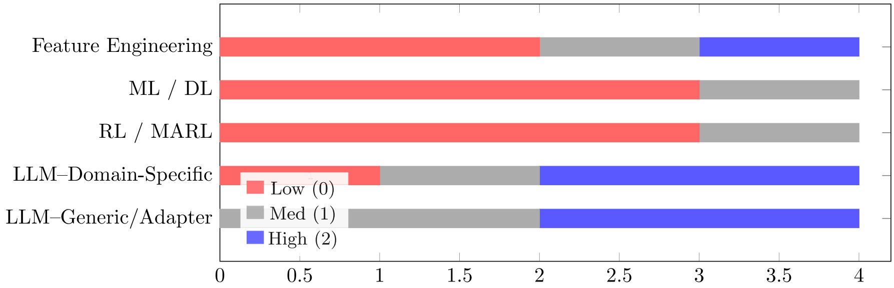


Fig. 6 Stacked count of Low, Medium, and High ratings for each technique category across the four capability criteria. Data drawn directly from Table 7

Kumbure et al., 2022; Kehinde et al., 2023; Ashtiani & Raahemi, 2023; Xu et al., 2023; Patil et al., 2024; Su et al., 2023; Li et al., 2024; Nti et al., 2020; Zhang & Chen, 2024; Guennioui et al., 2024; Yeo et al., 2023; Beniwal et al., 2024; Safari & Badamchizadeh, 2024; Ma et al., 2024; Bagalkot et al., 2024; Mulla et al., 2024; Phuoc et al., 2024; Tatane et al., 2024; Md et al., 2023; Wu et al., 2023; Khan et al., 2022; Wu et al., 2023; Lu et al., 2023; Zhang & Yang, 2023; Yang et al., 2023; Yu et al., 2024; Zhang et al., 2024; Koa et al., 2024). This approach limits models ability to have same performance under different stock market conditions. To address this problem, future research should prioritize the development and evaluation of cross-market prediction models through below :

1. Build multi-geography benchmarks: create datasets spanning North America, Europe, emerging markets, and alternative assets (commodities, FX, crypto) with consistent train/test splits.
2. Design cross-market benchmarks with standard train/test splits covering multiple stock market exchanges.
3. Define zero- and few-shot protocols that explicitly measure transfer learning. for example train on stock market and test on another stock market while report changes in accuracy, Sharpe ratio, drawdown and other evaluation metrics.
4. Adopt domain-adaptation strategies including adversarial alignment and meta-learning to learn market-invariant features and reduce the need for re-training.
5. Utilize LLM few-shot learning and in-context adaptation: pre-train on multi-market financial text so that feature extractors generalize across languages, market conventions, and news sources (Lopez-Lira & Tang, 2023; Yang et al., 2023).

6.2 Mitigating Noisy Real-Time Data Streams

Stock market forecasting faces a critical challenge in handling noise and invalid information in real-time financial data streams including stocks prices, order books, financial news which would lead to invalid model outcome and inaccurate predictions leading to wrong trading decisions (Picasso et al., 2019; Jing et al., 2021; Berradi et al., 2022). future research should consider handling noise in input stock data by focusing on the following areas

1. Denoising Pre-processing: Integrate neural denoisers for example denoising autoencoders (Berradi et al., 2022)) or GAN-based anomaly detectors (He et al., 2023) upstream of prediction models to filter noisy data and handle noisy information.
2. Uncertainty-Aware Prediction: Extend ML/RL/LLMs frameworks with Bayesian components or ensemble-dropout methods to quantify confidence intervals, enabling the system to decline low-certainty inputs.
3. Adversarial Noise Benchmarks: Develop standardized "noise injection" scenarios—simulating noisy data feed or fake-news and evaluate model degradation and recovery times.

6.3 Scalability Across Diverse Financial Tasks

Many RL, MARL and LLM-based approaches require huge compute and memory, which limits their ability to handle latency-sensitive or resource-constrained settings including high-frequency trading, mobile portfolio apps, large-scale backtests (Md et al., 2023; He et al., 2023; Wu et al., 2023). To ensure practical scalability, future work should:

1. utilize adapter-based fine-tuning: Utilize parameter-efficient methods like LoRA (Yang et al., 2023) and prefix-tuning to adapt large LLMs or RL policies to new tasks with minimal re-training.
2. Design distributed and edge pipelines: Develop split-compute frameworks that offload heavy processing to the cloud while enabling low-latency on-device predictions for time-critical financial operations.
3. Apply model compression techniques: Explore model pruning and quantization (Bagalkot et al., 2024; Yang et al., 2023) to shrink models without significant accuracy loss.

6.4 Adaptive Learning for Sudden Real-Time Market Shifts & Shock Adaptation

We note that memory-augmented agents such as FinMem (Yu et al., 2024) and FinAgent (Zhang et al., 2024) already provide long-term event retention and reflection, which helps with gradual context shifts. However, they do not explicitly detect Sudden shifts in market behavior or adapt in real time to external shocks. Future work should build on these memory modules by integrating:

1. Real-time warning change detection: Implement streaming drift detectors that continuously monitor incoming market data and detect significant changes before the model's internal parameters are updated and This allows for early warning of market changes.
2. Stress scenario simulation : implement shock scenario during training so that shocks become first-class events in the memory hierarchy to update model behavior accordingly.

6.5 Integrating LLMs with Traditional Feature Selection and Predictive Modeling

Table 1 shows that several studies highlight the need for more comprehensive feature engineering that incorporates various data sources beyond technical indicators to improve stock prediction accuracy (Xu et al., 2023; Kumari & Swarnkar, 2023; Chaudhari & Thakkar, 2023; Htun et al., 2023; Patil et al., 2024; Su et al., 2023; Li et al., 2024; Zhao et al., 2024). Future research could explore integrating LLMs with feature selection techniques combined with ML and DL algorithms to enhance stock market predictions. LLMs, when combined with feature selection algorithms like LASSO or ensemble methods, could provide richer, context-aware data inputs by transforming unstructured financial data like news and social media sentiment into features that are more meaningful for ML and DL models. Such hybrid approaches could improve model performance by utilizing LLMs' strengths in NLP to enrich traditional predictive models.

6.6 Improving Prediction Model Explainability using LLMs

Lack of model explainability represents a common challenge across several stock predictions as summarized in Table 2 and Table 4 (Nti et al., 2020; Zhang & Chen, 2024; Guennioui et al., 2024; Kianrad et al., 2024; Zou & Xiao, 2024; Liangchao, 2024; Yeo et al., 2023; Beniwal et al., 2024; Safari & Badamchizadeh, 2024; Ma et al., 2024; Bagalkot et al., 2024; Mulla et al., 2024; Phuoc et al., 2024; Tatane et al., 2024; Md et al., 2023; Wu et al., 2023; Shavandi & Khedmati, 2022; He et al., 2023; Chakole et al., 2021; Jang & Seong, 2023; Gu et al., 2024; Lin et al., 2022; Chen & Tan, 2023; Gronauer & Diepold, 2022; Kapoor, 2018; Yang et al., 2023; Gangopadhyay & Majumder, 2023; Fazlija & Harder, 2022; Li & Pan, 2022; Bouadjenek et al., 2023; Li et al., 2020; Du et al., 2024b; Khan et al., 2022; Mu et al., 2023; Berradi et al., 2022; Li & Zhang, 2023; Abdelfattah et al., 2024; Das et al., 2024; Ko & Lee, 2024; Pelster & Val, 2024). LLMs could play an important role in making ML, DL and some LLM-based models in stock market prediction more interpretable. By utilizing LLMs' strength in generating explanations for prediction outcome, researchers can improve model transparency and reasoning behind predictions (Yu et al., 2024; Park et al., 2023). Some studies utilized ML models to better explain LLMs decision through using topic modeling (Lopez-Lira & Tang, 2023) and some other researchers utilized RL for auto-correction for LLM explanation and prediction decision (Yang et al., 2023; Koa et al., 2024). Further research is needed to enhance model explainability either through utilizing LLMs power and integrating traditional predictive models like ML and DL models with LLMs models to clarify complex stock market forecasts.

6.7 Addressing Text-Based Prediction Challenges with LLMs

Research studies show the potential of text-based data in stock market forecasting as in Table 2. However, there are limitations in addressing issues such as misinformation, sentiment ambiguity and unstructured data (Picasso et al., 2019; Jing et al., 2021; Ashtiani & Raahemi, 2023; Gangopadhyay & Majumder, 2023; Fazlija

& Harder, 2022; Li & Pan, 2022; Bouadjenek et al., 2023; Li et al., 2020; Du et al., 2024b; Khan et al., 2022; Mu et al., 2023; Berradi et al., 2022; Abdelfattah et al., 2024; Das et al., 2024). LLMs research studies show potential improvement in financial sentiment analysis through handling text ambiguity and complex unstructured context (Delgadillo et al., 2024; Yan & Qin, 2024; Fatouros et al., 2023; Lopez-Lira & Tang, 2023; Xing, 2024; Du et al., 2024a). Future research should study integrating LLMs and traditional ML and DL prediction models to properly handle textual unstructured input like news and social media. However, fake information represents a challenge for both LLMs and traditional ML and DL studies. Future studies could examine how LLMs can assess information credibility, identify inconsistencies and integrate textual data from multiple sources to provide a more accurate view of market sentiment, enhancing accurate predictions that depend heavily on textual data sources.

6.8 Optimizing Ensemble and Hybrid Models using LLM

Research on ensemble and hybrid methods highlights the need for further exploration and comparison of different techniques and model combinations (Jing et al., 2021; Patil et al., 2024; Nti et al., 2020; Zhang & Chen, 2024; Kianrad et al., 2024; Zou & Xiao, 2024; Liangchao, 2024; Yeo et al., 2023; Ma et al., 2024; Bagalkot et al., 2024; Mulla et al., 2024; Wu et al., 2023). Future research should systematically evaluate the performance of different ensemble/hybrid learning approaches across different market conditions and datasets to identify optimal combinations of models and optimization algorithms. Furthermore, Integrating LLMs models with traditional ML and DL models presents a novel research direction that allows utilizing LLMs advantages in financial tasks as summarized in Tables 3, 4 and 5. Additionally, Future research should explore LLMs ability to automate selection of ensemble and hybrid models according to market conditions and evaluating LLMs ability to dynamically adjust the weighting of models within an ensemble according to market conditions.

6.9 Addressing RL/MARL Challenges with LLMs

RL and MARL show potential, challenges remain in training RL agents, handling multiple assets, incorporating external factors, and evaluating credit assignment in MARL (Sun et al., 2023; Huang et al., 2024; Guennioui et al., 2024; Kianrad et al., 2024; Zou & Xiao, 2024; Liangchao, 2024; Yeo et al., 2023; Beniwal et al., 2024; Safari & Badamchizadeh, 2024; Ma et al., 2024; Srivastava et al., 2022; Low & Sakk, 2023; Bagalkot et al., 2024; Mulla et al., 2024; Achiam et al., 2023; Park et al., 2023). Future research direction should focus on addressing these limitations through developing more efficient training methods, exploring different RL algorithms, evaluating and comparing different solutions to credit assignments in MARL on real-world datasets and stock markets is crucial. Existing research shows that RL integration with LLMs can enhance LLMs ability to correct LLM-Agent's stock trading decisions automatically [98,113]. Extra research is needed to explore the accuracy of different RL techniques in LLMs fine-tunning. In addition to that, LLMs can contribute to enhancing RL agent decision-making in stock market forecasting through providing guidance of RL training and providing RL agent with market knowledge, sentiment analysis, and news interpretation. Finally, combining MARL with LLMs for portfolio management is an unexplored area so future research is needed in this re-search area.

6.10 Prompt Engineering Framework for Financial Applications

Tables 3, 4, and 5 shows several research studies utilizing LLMs to different financial tasks including stock trading, financial sentiment analysis and portfolio management. The quality of LLMs answer depends heavily on the quality of input prompt. Key challenge identified in this survey is the lack of standard framework for prompt engineering in financial domain as each researcher is utilizing his own prompts with no common standard. Suggested Future research is developing a financial tailored prompt engineering framework for different financial tasks that enhances accuracy of LLMs response.

6.11 Finance Tailored LLMs and Fine-tunning LLMs

Although general-purpose LLMs show promising outcomes in financial sentiment analysis, portfolio management and stock trading as summarized in Tables 3, 4, and 5 , some research studies show that performance can be enhanced through fine-tunning LLMs with financial tailored information using LoRA (Zhang & Yang, 2023) or hybrid tuning methodology (Raffel et al., 2020) or building specialized financial LLMs (Li et al., 2023; Le Scao et al., 2022). Future research should explore and compare different fine-tunning approaches to enhance general-purpose LLMs ability to handle financial tasks in addition to comparing tailored general-purpose fine-tuned LLMs with financial specific LLMs.

7 Limitation of the study

The focus of our review was on the current research landscape regarding the application of LLMs and ML in stock market analytics. It is essential to acknowledge study limitations. First, this study doesn't include Ethical considerations of using LLMs and LLMs Hallucinations. Second, evaluating LLMs models requires consideration beyond accuracy, including interpretability, robustness, and ethical implications. This survey does not offer a definitive framework for such an evaluation framework, highlighting the need for standardized benchmarks and metrics. Finally, the accessibility of special-ized financial LLMs, like BloombergGPT, remains limited due to proprietary concerns and computational demands, potentially limiting our ability to validate LLMs outcome and compare results of various specialized LLMs. Overcoming these limitations through collaborative efforts and the development of more accessible frameworks is important for advancing this promising field.

8 Conclusion

This survey provides an overview of recent advancements in stock market fore-casting, focusing on the transformative potential of LLMs in stock market forecasting and financial analytics. Our analysis of traditional predictive models reveals that hybrid and ensemble methods result in superior forecasting accuracy. It also highlights that using proper feature selection techniques to choose relevant features from diverse stock data—such as technical indicators, fundamental factors, and textual data from news and social media—significantly improves prediction accuracy. LSTM networks outperform in stock price prediction, while ensemble methods combining ML and DL approaches enhance accuracy. Text-based predictions using sentiment analysis from social media and news show potential but must address issues like misinformation and sentiment ambiguity. RL-based approaches excel in adapting to dynamic market environments but require extensive training data. MARL performs better in handling multiple stocks and portfolio optimization, though it faces the credit assignment challenge in identifying the best weight distribution across agents. Despite recent advancements in traditional ML, DL, and RL approaches, common issues remain, including the lack of model explainability, difficulty in handling complex unstructured financial news, and limitations in generalizing predictions across diverse datasets and market conditions.

Our analysis shows that LLMs address the limitations of traditional models for stock market analytics by efficiently processing complex unstructured data, such as financial news and social media, and providing improved contextual understanding. This enables LLMs to adapt to dynamic market conditions. LLMs can handle various stock market tasks, including financial sentiment analysis, portfolio management, robo-advising, investment decisions, and financial trading agents. There is a growing need for specialized LLMs tailored to financial tasks, such as BloombergGPT, BBT-FinT5, and FinGPT, which improve performance compared to general-purpose LLMs. The use of memory modules, which store important historical events, is also crucial for improving stock forecasting, as seen in FinMem and FinAgent. Integrating RL with LLMs, as in FinGPT and Koa et al., enables the automation of the outcome correction process without manual intervention. Unlike traditional models that provide black-box predictions, LLMs offer explanations for prediction outcomes; however, validating these explanations is necessary to mitigate the impact of LLM hallucinations. Despite their advantages, LLMs face common challenges, including the lack of a standard prompt framework for financial tasks, limited research on cross-market LLMs, the need to mitigate bias in training data, and the need to explore different fine-tuning techniques.

This survey contributes by analyzing the strengths and weaknesses of both traditional predictive models and LLMs for different financial tasks. A key contribution lies in the exploration of how LLMs can be combined with traditional ML, DL, and RL approaches to overcome their respective limitations. This includes addressing challenges in handling unstructured data, improving model explainability, automatic correction for LLMs outcome through integration with RL, enhancing generalizability across diverse datasets and exploring LLMs capabilities in addressing limitations of MARL, Hybrid models, text-based prediction and other predictive models.

Finally, this survey shows that LLMs have the potential to greatly improve stock market forecasting and by fol-lowing research directions and handling research limitations outlined in previous sections, we can significantly improve the ability of LLMs to handle stock market fore-casting and other financial tasks with higher accuracy.

Author Contributions Conceptualization M.D., A.H. and E.H.; Writing—original draft, M.D. and A.H.; Supervision, E.H. and A.H.; Writing—review & editing, M.D., A.H. and E.H. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). This research received no external funding.

Data Availability Not applicable.

Materials Availability Not applicable.

Code Availability Not applicable.

Declarations

Conflict of interest/Competing interests The authors declare no conflicts of interest.

Ethics Approval and Consent to Participate The authors declare that all authors gave their consent to participate.

Consent for Publication The authors declare that all authors gave their consent for publication.

References

Abdelfattah, B. A., Darwish, S. M., & Elkaffas, S. M. (2024). Enhancing the prediction of stock market movement using neutrosophic-logic-based sentiment analysis. *J. Theor. Appl. Electron. Commer. Res., 19*(1), 116–134. https://doi.org/10.3390/jtaer19010007

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L.. others (2023). Gpt-4 technical report. https://doi.org/10.48550/arXiv.2303.08774 (Available online: arXiv:2303.08774)

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. arXiv:1908.10063, https://doi.org/10.48550/arXiv.1908.10063

Ashtiani, M. N., & Raahemi, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Syst. Appl., 217*, Article 119509. https://doi.org/10.1016/j.eswa.2023.119509

Bagalkot, S. S., Naik, N., et al. (2024). Novel grey wolf optimizer based parameters selection for garch and arima models for stock price prediction. *PeerJ Computer Science, 10*, Article e1735. https://doi.org/10.7717/peerj-cs.1735

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B.. others (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. arXiv:2302.04023. https://doi.org/10.18653/v1/2023.ijcnlp-main.45

Beniwal, M., Singh, A., & Kumar, N. (2024). Forecasting multistep daily stock prices for long-term investment decisions: A study of deep learning models on global indices. *Eng. Appl. Artif. Intell., 129*, Article 107617. https://doi.org/10.1016/j.engappai.2023.107617

Berradi, Z., Lazaar, M., Mahboub, O., Berradi, H., & Omara, H. (2022). Combination of deep-learning models to forecast stock price of aapl and tsla. *Jordanian Journal of Computers and Information Technology, 8*(4). https://doi.org/10.5455/jjcit.71-1655723854

Bouadjenek, M. R., Sanner, S., & Wu, G. (2023). A user-centric analysis of social media for stock market prediction. *ACM Trans. Web, 17*(2), 1–22. https://doi.org/10.1145/3532856

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P.. Amodei, D. (2020). Language models are few-shot learners. *Proceedings of the 34th international conference on neural information processing systems.* Red Hook, NY, USA: Curran Associates Inc. https://papers.nips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Cagliero, L., Fior, J., & Garza, P. (2023). Shortlisting machine learning-based stock trading recommendations using candlestick pattern recognition. *Expert Syst. Appl., 216*, Article 119493. https://doi.org/10.1016/j.eswa.2022.119493

Chakole, J. B., Kolhe, M. S., Mahapurush, G. D., Yadav, A., & Kurhekar, M. P. (2021). A q-learning agent for automated trading in equity stock markets. *Expert Syst. Appl., 163*, Article 113761. https://doi.org/10.1016/j.eswa.2020.113761

Chaudhari, K., & Thakkar, A. (2023). Neural network systems with an integrated coefficient of variation-based feature selection for stock price and trend prediction. *Expert Syst. Appl., 219*, Article 119527. https://doi.org/10.1016/j.eswa.2023.119527

Chen, L., Zaharia, M., & Zou, J. (2023). How is chatgpt's behavior changing over time? arXiv:2307.09009. https://doi.org/10.1162/99608f92.5317da47

Chen, R., & Tan, Y. (2023). Credit assignment with predictive contribution measurement in multi-agent reinforcement learning. *Neural Netw., 164*, 681–690. https://doi.org/10.1016/j.neunet.2023.05.021

Chen, S., & Zhou, C. (2020). Stock prediction based on genetic algorithm feature selection and long short-term memory neural network. *IEEE Access, 9*, 9066–9072. https://doi.org/10.1109/access.2020.3047109

Chhikara, G., Sharma, A., Gurucharan, V., Ghosh, K., & Chakraborty, A. (2024). Lamsum: Creating extractive summaries of user generated content using llms. arXiv:2406.15809. https://doi.org/10.48550/arXiv.2406.15809

Chung, H., & Shin, K.-S. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. *Sustainability, 10*(10), 3765. https://doi.org/10.3390/su10103765

Cong, L.W., Liang, T., Zhang, X., & Zhu, W. (2024). *Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information* Tech. rep. National Bureau of Economic Research, 2024. https://doi.org/10.3386/w33168

Dang, Q.-V. (2019). Reinforcement learning in stock trading. *International conference on computer science, applied mathematics and applications Springer., 2019*, 311–322. https://doi.org/10.1007/978-3-030-38364-0_28

Das, N., Sadhukhan, B., Chatterjee, R., & Chakrabarti, S. (2024). Integrating sentiment analysis with graph neural networks for enhanced stock prediction: A comprehensive survey. *Decision Analytics Journal, 100417*. https://doi.org/10.1016/j.dajour.2024.100417

Delgadillo, J., Kinyua, J., & Mutigwe, C. (2024). Finsosent: Advancing financial market sentiment analysis through pretrained large language models. *Big Data and Cognitive Computing, 8*(8), 87. https://doi.org/10.3390/bdcc8080087

Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). Qlora: Efficient finetuning of quantized llms. In: A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), *Advances in neural information processing systems* (Vol. 36, pp. 10088–10115). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2023/file/1feb87871436031bdc0f2beaa62a049b-Paper-Conference.pdf

Du, K., Xing, F., Mao, R., & Cambria, E. (2024a). An evaluation of reasoning capabilities of large language models in financial sentiment analysis. *Ieee conference on artificial intelligence (ieee cai), singapore.* https://doi.org/10.1109/cai59869.2024.00042

Du, K., Xing, F., Mao, R., & Cambria, E. (2024). Financial sentiment analysis: Techniques and applications. *ACM Comput. Surv., 56*(9), 1–42. https://doi.org/10.1145/3649451

Du, Y., Li, S., Torralba, A., Tenenbaum, J.B., & Mordatch, I. (2023). Improving factuality and reasoning in language models through multiagent debate. arXiv:2305.14325. https://doi.org/10.48550/arXiv.2305.14325

Fatouros, G., Soldatos, J., Kouroumali, K., Makridis, G., & Kyriazis, D. (2023). Transforming sentiment analysis in the financial domain with chatgpt. *Machine Learning with Applications, 14*, Article 100508. https://doi.org/10.1016/j.mlwa.2023.100508

Fazlija, B., & Harder, P. (2022). Using financial news sentiment for stock price direction prediction. *Mathematics, 10*(13), 2156. https://doi.org/10.3390/math10132156

Gangopadhyay, S., & Majumder, P. (2023). Text representation for direction prediction of share market. *Expert Syst. Appl., 211*, Article 118472. https://doi.org/10.1016/j.eswa.2022.118472

Gao, J., Ding, X., Qin, B., & Liu, T. (2023). Is chatgpt a good causal reasoner? a comprehensive evaluation. arXiv:2305.07375. https://doi.org/10.18653/v1/2023.findings-emnlp.743

Gronauer, S., & Diepold, K. (2022). Multi-agent deep reinforcement learning: a survey. *Artif. Intell. Rev., 55*(2), 895–943. https://doi.org/10.1007/s10462-021-09996-w

Gu, Y. H. (2024). Pro trader rl: Reinforcement learning framework for generating trading knowledge by mimicking the decision-making patterns of professional traders. *Expert Systems with Applications* 124465. https://doi.org/10.1016/j.eswa.2024.124465

Guennioui, O., Chiadmi, D., & Amghar, M. (2024). Global stock price forecasting during a period of market stress using lightgbm. *International Journal of Computing and Digital Systems, 15*(1), 19–28. https://doi.org/10.12785/ijcds/150103

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J.. others (2018). Soft actor-critic algorithms and applications. arXiv:1812.05905. https://doi.org/10.48550/arXiv.1812.05905

He, F.-F., Chen, C.-T., & Huang, S.-H. (2023). A multi-agent virtual market model for generalization in reinforcement learning based trading strategies. *Appl. Soft Comput., 134*, Article 109985. https://doi.org/10.1016/j.asoc.2023.109985

Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation, 9*(1), 26. https://doi.org/10.1186/s40854-022-00441-7

Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S.. Chen, W. (2021). Lora: Low-rank adaptation of large language models. arXiv:2106.09685. https://doi.org/10.48550/arXiv.2106.09685

Huang, B., Huan, Y., Xu, L. D., Zheng, L., & Zou, Z. (2019). Automated trading systems statistical and machine learning methods and hardware implementation: a survey. *Enterprise Information Systems, 13*(1), 132–144. https://doi.org/10.1080/17517575.2018.1493145

Huang, Y., Zhou, C., Cui, K., & Lu, X. (2024). A multi-agent reinforcement learning framework for optimizing financial trading strategies based on timesnet. *Expert Syst. Appl., 237*, Article 121502. https://doi.org/10.1016/j.eswa.2023.121502

Jang, J., & Seong, N. (2023). Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. *Expert Syst. Appl., 218*, Article 119556. https://doi.org/10.1016/j.eswa.2023.119556

Jin, H., Zhang, Y., Meng, D., Wang, J., & Tan, J. (2024). A comprehensive survey on process-oriented automatic text summarization with exploration of llm-based methods. arXiv:2403.02901. https://doi.org/10.48550/arXiv.2403.02901

Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Syst. Appl., 178*, Article 115019. https://doi.org/10.1016/j.eswa.2021.115019

Kapoor, S. (2018). *Multi-agent reinforcement learning: A report on challenges and approaches.* https://doi.org/10.48550/arXiv.1807.09427. arXiv:1807.09427 [cs.AI].

Kawamura, K., Li, Z., Lin, C.-K., & McDanel, B. (2024). Revelata at the finllm challenge task: Improving financial text summarization by restricted prompt engineering and fine-tuning. C.-C. Chen et al. (Eds.), In: *Proceedings of the eighth financial technology and natural language processing and the 1st agent ai for scenario planning* (pp. 146–152). Jeju, South Korea: -. https://aclanthology.org/2024.finnlp-2.15

Kehinde, T., Chan, F. T., & Chung, S. H. (2023). Scientometric review and analysis of recent approaches to stock market forecasting: Two decades survey. *Expert Syst. Appl., 213*, Article 119299. https://doi.org/10.1016/j.eswa.2022.119299

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing,* 1–24. https://doi.org/10.1007/s12652-020-01839-w

Kianrad, A., Arani, M. N., Hasani, K., Zargar, M., Erfani, E., & Razmjou, A. (2024). Investigating the impact of company announcements on stock prices: an application of machine learning on australian lithium market. *Miner. Econ., 37*(1), 163–172. https://doi.org/10.1007/s13563-024-00428-z

Kim, J. H. (2023). What if chatgpt were a quant asset manager. *Financ. Res. Lett., 58*, Article 104580. https://doi.org/10.1016/j.frl.2023.104580

Ko, H., & Lee, J. (2024). Can chatgpt improve investment decisions? from a portfolio management perspective. *Financ. Res. Lett., 64*, Article 105433. https://doi.org/10.1016/j.frl.2024.105433

Koa, K. J., Ma, Y., Ng, R., & Chua, T.-S. (2024). Learning to generate explainable stock predictions using self-reflective large language models. In: *Proceedings of the acm on web conference 2024* (pp. 4304–4315). https://doi.org/10.1145/3589334.3645611

Kumari, B., & Swarnkar, T. (2023). Forecasting daily stock movement using a hybrid normalization based intersection feature selection and ann. *Procedia Computer Science, 218*, 1424–1433. https://doi.org/10.1016/j.procs.2023.01.121

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Syst. Appl., 197*, Article 116659. https://doi.org/10.1016/j.eswa.2022.116659

Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D.. others (2022). Bloom: A 176b-parameter open-access multilingual language model. arXiv:2211.05100. https://doi.org/10.48550/arXiv.2211.05100

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N.. Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 9459–9474). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

Li, J., Liu, Y., Gong, H., & Huang, X. (2024). Stock price series forecasting using multi-scale modeling with boruta feature selection and adaptive denoising. *Appl. Soft Comput., 154*, Article 111365. https://doi.org/10.1016/j.asoc.2024.111365

Li, M., & Zhang, Y. (2023). Integrating social media data and historical stock prices for predictive analysis: A reinforcement learning approach. *International Journal of Advanced Computer Science & Applications, 14*(12). https://doi.org/10.14569/ijacsa.2023.0141203

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong. *Information Processing & Management, 57*(5), Article 102212. https://doi.org/10.1016/j.ipm.2020.102212

Li, Y., & Pan, Y. (2022). A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics, 13*(2), 139–149. https://doi.org/10.1007/s41060-021-00279-9

Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large language models in finance: A survey. *Proceedings of the fourth acm international conference on ai in finance* (pp. 374–382). https://doi.org/10.1145/3604237.3626869

Liangchao, L. (2024). Presenting an optimized hybrid model for stock price prediction. *International Journal of Advanced Computer Science & Applications, 15*(1). https://doi.org/10.14569/ijacsa.2024.0150174

Lin, Y.-C., Chen, C.-T., Sang, C.-Y., & Huang, S.-H. (2022). Multiagent-based deep reinforcement learning for risk-shifting portfolio management. *Appl. Soft Comput., 123*, Article 108894. https://doi.org/10.1016/j.asoc.2022.108894

Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. arXiv:2304.07619. https://doi.org/10.2139/ssrn.4412788

Low, P. R., & Sakk, E. (2023). Comparison between autoregressive integrated moving average and long short term memory models for stock price prediction. *IAES Int J Artif Intell, 12*(4), 1828–1835. https://doi.org/10.11591/ijai.v12.i4.pp1828-1835

Lu, D., Wu, H., Liang, J., Xu, Y., He, Q., Geng, Y.. Xiao, Y. (2023). Bbt-fin: Comprehensive construction of chinese financial domain pre-trained language model, corpus and benchmark. arXiv:2302.09432. https://doi.org/10.48550/arXiv.2302.09432

Ma, C., Wu, J., Hu, H., Chen, Y.N., Li, J.Y. (2024). Predicting stock prices using hybrid lstm and arima model. *IAENG International Journal of Applied Mathematics, 54*(3). https://www.iaeng.org/IJAM/issues_v54/issue_3/IJAM_54_3_12.pdf

Mao, R., Chen, G., Zhang, X., Guerin, F., & Cambria, E. (2024). Gpteval: A survey on assessments of chatgpt and gpt-4. N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue (Eds.), *Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (lrec-coling 2024)* (pp. 7844–7866). Torino, Italia: ELRA and ICCL. https://aclanthology.org/2024.lrec-main.693

Md, A. Q., Kapoor, S., & AV, C. J., Sivaraman, A. K., Tee, K. F., Sabireen, H., & Janakiraman, N. (2023). Novel optimization approach for stock price forecasting using multi-layered sequential lstm. *Appl. Soft Comput., 134*, Article 109830. https://doi.org/10.1016/j.asoc.2022.109830

Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon and Schuster. isbn: 9780743276641 https://dl.acm.org/doi/book/10.5555/1204056

Mu, G., Gao, N., Wang, Y., & Dai, L. (2023). A stock price prediction model based on investor sentiment and optimized deep learning. *Ieee Access, 11*, 51353–51367. https://doi.org/10.1109/access.2023.3278790

Mulla, R., Saini, S., Mane, P., Balkhande, B., Pawar, M., & Deshmukh, K. (2024). A novel hybridapproach for stock market index forecasting using cnn-lstm fusion modell. *Int J Intell Syst Appl Eng*. https://www.ijisae.org/index.php/IJISAE/article/view/4513

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data, 7*(1), 20. https://doi.org/10.1186/s40537-020-00299-5

Oehler, A., & Horn, M. (2024). Does chatgpt provide better advice than robo-advisors? *Financ. Res. Lett., 60*, Article 104898. https://doi.org/10.1016/j.frl.2023.104898

OpenAI (2023). *Gpt-v system card.* https://cdn.openai.com/papers/GPTV_System_Card.pdf. OpenAI. https://cdn.openai.com/papers/GPTV_System_Card.pdf (Accessed: 2024-09-25)

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th annual acm symposium on user interface software and technology* (pp. 1–22). https://doi.org/10.1145/3586183.3606763

Patil, P. R., Parasar, D., & Charhate, S. (2024). Wrapper-based feature selection and optimization-enabled hybrid deep learning framework for stock market prediction. *International Journal of Information Technology & Decision Making, 23*(01), 475–500. https://doi.org/10.1142/S0219622023500116

Pelster, M., & Val, J. (2024). Can chatgpt assist in picking stocks? *Financ. Res. Lett., 59*, Article 104786. https://doi.org/10.1016/j.frl.2023.104786

Peng, Y., Albuquerque, P., Kimura, H., & Saavedra, C. (2021). Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators. *Machine Learning with Applications, 5*, Article 100060. https://doi.org/10.1016/j.mlwa.2021.100060

Phuoc, T., Anh, P., Tam, P. H., & Nguyen, C. V. (2024). Applying machine learning algorithms to predict the stock price trend in the stock market-the case of vietnam. *Humanities and Social Sciences Communications, 11*(1), 1–18. https://doi.org/10.1057/s41599-024-02807-x

Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Syst. Appl., 135*, 60–70. https://doi.org/10.1016/j.eswa.2019.06.014

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M.. Liu, P.J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research, 21*(140), 1–67. https://dl.acm.org/doi/abs/10.5555/3455716.3455856

Rossi, S., & Tinn, K. (2021). Rational quantitative trading in efficient markets. *Journal of Economic Theory, 191*, Article 105127. https://doi.org/10.1016/j.jet.2020.105127

Safari, A., & Badamchizadeh, M. A. (2024). Deepinvesting: Stock market predictions with a sequence-oriented bilstm stacked model-a dataset case study of amzn. *Intelligent Systems with Applications, 24*, Article 200439. https://doi.org/10.1016/j.iswa.2024.200439

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017a). Proximal policy optimization algorithms. arXiv:1707.06347. https://doi.org/10.48550/arXiv.1707.06347

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017b). Proximal policy optimization algorithms. arXiv:1707.06347. https://doi.org/10.48550/arXiv.1707.06347

Sethia, A., & Raut, P. (2019). Application of lstm, gru and ica for stock price prediction. *Information and communication technology for intelligent systems: Proceedings of ictis 2018, volume 2* (pp. 479–487). https://doi.org/10.1007/978-981-13-1747-7_46

Shavandi, A., & Khedmati, M. (2022). A multi-agent deep reinforcement learning framework for algorithmic trading in financial markets. *Expert Syst. Appl., 208*, Article 118124. https://doi.org/10.1016/j.eswa.2022.118124

Srivastava, A. K., Srivastava, A., Singh, S., Sugandha, S., & Tripta, & Gupta, S. (2022). Design of machine-learning classifier for stock market prediction. *SN Computer Science, 3*(1), 88. https://doi.org/10.1007/s42979-021-00970-5

Su, I.-F., Lin, P. L., Chung, Y.-C., & Lee, C. (2023). Forecasting of taiwan's weighted stock price index based on machine learning. *Expert. Syst., 40*(9), Article e13408. https://doi.org/10.1111/exsy.13408

Sun, S., Wang, R., & An, B. (2023). Reinforcement learning for quantitative trading. *ACM Transactions on Intelligent Systems and Technology, 14*(3), 1–29. https://doi.org/10.1145/3582560

Tatane, K., Sahib, M.R., & Zaki, T. (2024). From time series to images: Revolutionizing stock market predictions with convolutional deep neural networks. *International Journal of Advanced Computer Science & Applications, 15*(1). https://doi.org/10.14569/ijacsa.2024.0150144

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N.. Polosukhin, I. (2017). Attention is all you need. I. Guyon et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wan, M., Safavi, T., Jauhar, S.K., Kim, Y., Counts, S., Neville, J.. others (2024). Tnt-llm: Text mining at scale with large language models. *Proceedings of the 30th acm sigkdd conference on knowledge discovery and data mining* (pp. 5836–5847).

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E.. Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Nips '22: Proceedings of the 36th international conference on neural information processing systems* (pp. 24824–24837). Red Hook, NY, USA: Curran Associates Inc. ISBN: 9781713871088. https://dl.acm.org/doi/10.5555/3600270.3602070

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S.. Mann, G. (2023). Bloomberggpt: A large language model for finance. arXiv:2303.17564. https://doi.org/10.48550/arXiv.2303.17564

Wu, Y., Fu, Z., Liu, X., & Bing, Y. (2023). A hybrid stock market prediction model based on gng and reinforcement learning. *Expert Syst. Appl., 228*, Article 120474. https://doi.org/10.1016/j.eswa.2023.120474

Xing, F. (2024). Designing heterogeneous llm agents for financial sentiment analysis. *ACM Trans. Manag. Inf. Syst.* https://doi.org/10.1145/3688399

Xu, X., Zhang, Y., McGrory, C. A., Wu, J., & Wang, Y.-G. (2023). Forecasting stock closing prices with an application to airline company data. *Data Science and Management, 6*(4), 239–246. https://doi.org/10.1016/j.dsm.2023.09.005

Xue, Q. (2024). Unlocking the potential: A comprehensive exploration of large language models in natural language processing. *Applied and Computational Engineering, 57*, 247–252. https://doi.org/10.54254/2755-2721/57/20241341

Yan, Y., & Qin, Q. (2024). Sentiment analysis of financial media commentary text based on finbert-bigru-cnn. *Proceedings of the 2024 international conference on smart city and information system* (p.1). https://doi.org/10.1145/3685088.3685187

Yang, B., Liang, T., Xiong, J., & Zhong, C. (2023). Deep reinforcement learning based on transformer and u-net framework for stock trading. *Knowl.-Based Syst., 262*, Article 110211. https://doi.org/10.1016/j.knosys.2022.110211

Yang, H., Liu, X.-Y., & Wang, C. D. (2023). Fingpt: Open-source financial large language models. arXiv:2306.06031. https://doi.org/10.2139/ssrn.4489826

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., & Hu, X. (2024). Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Trans. Knowl. Discov. Data, 18*(6), 1–32. https://doi.org/10.1145/3649506

Yang, J., Zhang, W., Zhang, X., Zhou, J., & Zhang, P. (2023). Enhancing stock movement prediction with market index and curriculum learning. *Expert Syst. Appl., 213*, Article 118800. https://doi.org/10.1016/j.eswa.2022.118800

Yang, X., Liu, W., Zhou, D., Bian, J., & Liu, T.-Y. (2020). Qlib: An ai-oriented quantitative investment platform. arXiv:2009.11189. https://doi.org/10.48550/arXiv.2009.11189

Yeo, L., Cao, Q., & Quek, C. (2023). Dynamic portfolio rebalancing with lag-optimised trading indicators using serofam and genetic algorithms. *Expert Syst. Appl., 216*, Article 119440. https://doi.org/10.1016/j.eswa.2022.119440

Yu, Y., Li, H., Chen, Z., Jiang, Y., Li, Y., Zhang, D.. Khashanah, K. (2024). Finmem: A performance-enhanced llm trading agent with layered memory and character design. *Proceedings of the aaai symposium series* (Vol. 3, pp. 595–597). https://doi.org/10.1609/aaaiss.v3i1.31290

Zhang, B., Yang, H., Zhou, T., Ali Babar, M., & Liu, X.-Y. (2023). Enhancing financial sentiment analysis via retrieval augmented large language models. *Proceedings of the fourth acm international conference on ai in finance* (pp. 349–356). https://doi.org/10.1145/3604237.3626866

Zhang, J., & Chen, X. (2024). A two-stage model for stock price prediction based on variational mode decomposition and ensemble machine learning method. *Soft. Comput., 28*(3), 2385–2408. https://doi.org/10.1007/s00500-023-08441-0

Zhang, W., Zhao, L., Xia, H., Sun, S., Sun, J., Qin, M.. others (2024). A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. *Proceedings of the 30th acm sigkdd conference on knowledge discovery and data mining* (pp. 4314–4325). https://doi.org/10.1145/3637528.3671801

Zhang, X., & Yang, Q. (2023). Xuanyuan 2.0: A large chinese financial chat model with hundreds of billions parameters. *Proceedings of the 32nd acm international conference on information and knowledge management* (pp. 4435–4439). https://doi.org/10.1145/3583780.3615285

Zhao, R., Lei, Z., & Zhao, Z. (2024). Research on the application of deep learning techniques in stock market prediction and investment decision-making in financial management. *Frontiers in Energy Research, 12*, 1376677. https://doi.org/10.3389/fenrg.2024.1376677

Zou, Z., & Xiao, G. (2024). Presenting a hybrid method to overcome the challenges of determining the uncertainty of future stock price identification. *International Journal of Advanced Computer Science & Applications, 15*(3). https://doi.org/10.14569/ijacsa.2024.0150327