2025 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CiFer) | 979-8-3315-0831-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/CIFER64978.2025.10975733

Enhancing Cryptocurrency Trading Strategies: A Deep Reinforcement Learning Approach Integrating Multi-Source LLM Sentiment Analysis

Nanjiang Du[1], Yida Zhao[1], Jintao Wang[1], Yicheng Zhu[1], Siyu Xie[1], Luyao Yang[1], Yiru Tong[1], Shengzhe Xu[1], Wangying Zhang[1], Zecheng Tang[1], Kai Xu[2], Jianfeng Ren[1], and Tianxiang Cui[1,*]
[1] School of Computer Science, University of Nottingham Ningbo China, Ningbo, China
[2] School of Computer Science, University of Nottingham, Nottingham, UK

***Abstract*—Recent advancements in large language models (LLMs) have demonstrated their potential to significantly impact finance trading, particularly through sentiment analysis. The cryptocurrency market, known for its volatility and unpredictability, often renders price-based trading approaches inadequate. This necessitates the adoption of more sophisticated techniques such as market sentiment analysis, which can benefit from the insights provided by LLMs. This study introduces an innovative method that integrates sentiment analysis derived from five distinct LLMs with deep reinforcement learning to devise a cryptocurrency trading strategy. Recognizing that LLM outputs cannot be guaranteed to be infallibly accurate, which contributing to the LLM hallucinations, this paper details the implementation of a stringent outlier detection and removal process. By adopting a "Trust-The-Majority" strategy, the research aims to ensure that trading decisions are informed by reliable sentiment data. In addition, sentiment scores are traditionally timestamped to the publication of news or social media posts. To more accurately reflect the actual impact of such information on market sentiment, this study applies the Ebbinghaus Forgetting Curve to model the waning influence of information over time. This allows for a more nuanced understanding of how news affects market dynamics. The enhanced sentiment scores, in conjunction with traditional market data such as OHLCV (Open, High, Low, Close, Volume), are utilized by a deep reinforcement learning model to make trading decisions. Experimental results demonstrate that the proposed multi-LLM sentiment-driven framework improves trading performance in the fast-paced cryptocurrency market. The methodology outlined in this paper offers a solid foundation for incorporating real-time market sentiment analysis into financial applications.**

I. Introduction

Cryptocurrencies, as emerging financial assets, have become increasingly prominent in the financial markets, with their market experiencing exponential growth in recent years. Bouri et al. [1] suggest that cryptocurrencies function as hedge assets due to their unique characteristics. As a result, an increasing number of financial institutions are integrating cryptocurrencies into their portfolios [2]. Simultaneously, academic interest in the subject has surged, with numerous researchers exploring the complexities of cryptocurrency trading dynamics [3]–[6].

Cryptocurrency trading research has predominantly relied on technical indicators. For example, strategies such as "Turtle Soup" [7] and "Nem (XEM)" [8] focus on trend prediction using chart patterns, with the former employing the Rate of Change (ROC) indicator and the Relative Strength Index (RSI) to forecast prices. Ha and Moon [9] applied over 12 technical indicators, including the moving average (MA) and stochastic oscillator, to uncover appealing technical patterns in the cryptocurrency market through genetic programming (GP). Similarly, studies by Grobys et al. [10] and Al-Yahyaee et al. [11] underscore potential market inefficiencies and long-term memory effects in cryptocurrency prices, indicating opportunities for optimized trading strategies. However, many technical indicators, such as the moving average, are known to exhibit delays [12], a challenge that is particularly pronounced in the highly volatile cryptocurrency market. As a result, by the time these indicators generate trading signals, significant price movements may have already occurred, leading to suboptimal buy or sell decisions [13]–[15].

The rapid advancement of machine learning algorithms has significantly improved their application in cryptocurrency trading, particularly in the prediction of cryptocurrency returns. Nakano [16] employed an artificial neural network (ANN) to forecast returns in technical Bitcoin trading. Similarly, Sun [17] utilized random forests with factors derived from Alpha01 [18], which captures features from the cryptocurrency market's historical data, to construct a trading prediction model. Slepaczuk and Zenkova [19] investigated the profitability of an algorithmic trading strategy based on a Support Vector Machine (SVM) model to identify cryptocurrencies with high or low predicted returns. Their findings revealed that the SVM-based strategy ranked fourth, only surpassing the S&P buy-and-hold (BAH) strategy, which involves holding the S&P index. Moreover, Singh [20] introduced a method using a Long Short-Term Memory (LSTM) model to predict Bitcoin's closing price. Trading decisions were then made by setting buy and sell thresholds based on the predicted price, with a simple return calculated at hourly intervals. However, research has shown that trend forecasting does not guarantee optimal trading outcomes, as prediction loss diverges from the broader objective of the problem [21]. The success of such algorithms is heavily dependent on prediction accuracy, which, according

*Corresponding author: T. Cui (email: tianxiang.cui@nottingham.edu.cn).

to the efficient market hypothesis [22], [23], is inherently limited since all relevant information is already reflected in current stock prices, making accurate future predictions nearly impossible. In contrast, Reinforcement Learning (RL) algorithms have shown considerable efficacy in solving sequential decision-making problems. The integration of deep learning (DL) with RL, referred to as Deep Reinforcement Learning (DRL), has gained significant attention due to its impressive successes in various fields [24]–[33]. RL enables the formulation of an optimal policy that maximizes an agent's expected cumulative profit [34], eliminating the need for direct price movement forecasting in trading decisions [35]. Patel [36] proposed a multi-agent approach to Bitcoin trading, operating on two levels: a macro-agent using a Double Q-learning network with a Multi-Layer Perceptron (MLP) at the minute level and a micro-agent using a Dueling Double Q-learning network at the order book level, with a reward function based on volume-weighted average Bitcoin price. Sattarov et al. [37] applied DRL using historical data from BTC, LTC, and ETH to analyze past price movements and make real-time trading decisions, guiding traders to choose between buying, selling, or holding. However, most existing DRL-based methods are limited to general market price data. According to efficient market theory [23], stock prices already incorporate all available information. While market factors are crucial, they may not fully capture the complexity of non-stationary cryptocurrency markets. Sentiment data from social media and news, which often contain implicit patterns and predictive signals, can potentially enhance decision-making processes and improve trading performance.

Lamon et al. [38] introduced an innovative approach by utilizing daily news and social media data labeled based on actual price changes rather than traditional sentiment analysis reliant on positive or negative sentiment. This method shifts the focus of prediction from price movements to sentiment polarity. However, the manual labeling of data in this approach introduces potential inefficiencies. Li et al. [39] conducted a sentiment analysis of Twitter data using natural language processing (NLP) with the Python package"TextBlob," which assigns polarity values to key words. They integrated this sentiment analysis with trading volume data and employed an Extreme Gradient Boosting Regression Tree Model to predict market trends for the ZClassic (ZCL) cryptocurrency. Sentiment indices, both weighted and unweighted, were computed hourly by aggregating the weights of relevant tweets, allowing direct comparisons with ZCL price data. More recently, large language models (LLMs) have gained prominence in the field of news sentiment analysis. Studies such as [11] and [40] have evaluated the performance of both closed-source LLMs (e.g., GPT-3.5/4) and open-source LLMs (e.g., Qwen [41], Baichuan [42]) for financial sentiment analysis. Despite their powerful capabilities, LLMs are not always accurate; they occasionally produce plausible yet incorrect responses due to hallucinations, and there are currently no reliable methods to verify the accuracy of their outputs. Another notable issue with using LLMs for sentiment analysis in trading strategies is the temporal mismatch between news events and market movements. Not every time stamp corresponds to relevant news, and the impact of news on asset prices is not confined to a specific point in time. This temporal lag poses a significant challenge, as the influence of news can extend across multiple time frames, complicating efforts to precisely align sentiment analysis with immediate market reactions.

To address these gaps, we propose a novel cryptocurrency trading strategy that integrates deep reinforcement learning (DRL) with multi-source large language model (LLM) sentiment analysis. The key contributions are as the follows: 1) We introduce a sentiment analysis-enhanced DRL framework specifically designed for cryptocurrency trading. 2) We employ multi-source LLMs to analyze market sentiment, adopting a "Trust-The-Majority" strategy and incorporate outlier detection process to eliminate anomalous values potentially caused by LLM hallucinations, thereby improving the reliability of the sentiment scores. 3) A forgetting mechanism is applied to the cleaned sentiment scores, allowing for a more accurate reflection of the temporal impact of sentiment information on the market. 4) Our strategy outperforms benchmark strategies, delivering a superior Sharpe ratio and highly competitive annualized returns.

II. PROBLEM DESCRIPTION

This study focuses on the implementation of an automated cryptocurrency trading strategy on an hourly basis using a reinforcement learning algorithm. The strategy involves three possible trading actions: buy, sell, and hold. Let p^t represent the closing price of the cryptocurrency at time t, and C^t denote the available cash at time t. Upon executing a buy action, the trader purchases Θ^t units of the cryptocurrency, where β is the transaction fee, set at 0.1%. The quantity of cryptocurrency purchased is calculated as follows:

$$\Theta^t = \frac{C^t \times (1-\beta)}{p^t}, \tag{1}$$

For the sell action, $\Theta^{t'}$ represents the number of cryptocurrency units held. Since there is no restriction on the number of transactions in the cryptocurrency market, the entire cash balance is used to either buy or sell all available assets. After the sell action, the updated cash value $C^{t'}$ is computed by:

$$C^{t'} = \Theta^{t'} \times p^{t'} \times (1-\beta), \tag{2}$$

When executing a hold action, the total value V^t and the market value M^t at time t are calculated as follows:

$$M^t = \Theta^t \times p^t, \tag{3}$$

$$V^t = C^t + M^t, \tag{4}$$

The objective of this study is to maximize the total returns over the trading period and T is the final time. The cumulative return $\aleph$ is defined as:

$$\aleph = \frac{V^T}{V^0} - 1, \tag{5}$$

III. Methodology

This work is formulated as a Markov decision process (MDP). An MDP is defined by a tuple (S, A, P, R, γ), where:

- S is a finite set of states.
- A is a finite set of actions.
- $P : S \times A \times S \rightarrow [0, 1]$ is the state transition probability function, where $P(s^{t+1} \mid s^t, a^t)$ denotes the probability of transitioning to state s^{t+1} from state s^t after taking action a^t.
- R is the reward function, where $r^t = R(s^t, a^t)$ represents the immediate reward received after taking action a^t in state s^t.
- $\gamma \in [0, 1]$ is the discount factor, representing the importance of future rewards.

At each time step t, the agent observes the current state $s_t \in S$, selects an action $a^t \in A$, receives a reward $r^t = R(s^t, a^t)$, and transitions to the next state s^{t+1} according to the transition probability $P(s^{t+1} \mid s^t, a^t)$. The objective of the agent is to learn a policy $\pi : S \rightarrow A$ that maximizes the expected cumulative reward G:

$$G = \sum_{k=0}^{\infty} \gamma^k r^{t+k+1}. \tag{6}$$

Algorithm 1 Integrating Multi-Source LLM Sentiment Analysis DRL Algorithm

Input: Number of Episode EP, Number of Step per Episode T, news data $\mathbb{D}_{news}$, social media data $\mathbb{D}_{twitter}$, and time-series data $\mathbb{D}_{time}$,
Initialize: Actor Network Ψ, Critic Network $\varPhi$, Buffer $\mathbb{B}$, Anomalous Tolerance φ, Holding State H^0.
Output:
for $ep \leftarrow 1$ **to** EP **do**
  for $t \leftarrow 1$ **to** T **do**
    $\mathbb{E}^t_{news} \leftarrow LLMs(\mathbb{D}^t_{news})$
    $\mathbb{E}^t_{social} \leftarrow LLMs(\mathbb{D}^t_{twitter})$
    $\mathbb{E}^t_{market} \leftarrow Concat(\mathbb{E}^t_{news}, \mathbb{E}^t_{social})$
    $\mathbb{E}^t_{cleaned} \leftarrow OutlierDetection(\mathbb{E}^t_{market}, \varphi)$ (Alg.2)
    Compute Ω^t in Eq. 7
    $s^t = \text{Concat}(H, \mathbb{D}^t_{time}, \Omega^t)$
    Get $prob$ form $\Psi(s^t)$
    Sample $prob$ get a^t
    Update Holding State H^t
    Calculate the immediate reward r^t by $R(s^t, a^t)$
    Store the transition $\xi = \{s^t, a^t, r^t, prob, s^{t+1})\}$ in memory buffer $\mathbb{B}$
  end for
  Update the Actor Ψ and the Critic $\varPhi$
end for

At each time step t, the price movement is captured by five key indicators: Open (o^t), High (h^t), Low (l^t), Close (c^t), and Volume (v^t) which reflects trading activity. The state at time t is then represented as $s^t = [H^t, o^t, h^t, l^t, c^t, v^t, \Omega^t]$, where H^t indicates the holding status of the cryptocurrency, with 0 representing no holdings and 1 indicating holdings. The vector Ω^t includes sentiment scores derived from news and social media data. The general framework of the proposed approach is illustrated in Figure 1. News data $\mathbb{D}news^t$ and social media data $\mathbb{D}twitter^t$ are processed through different-source LLMs, followed by outlier detection. A forgetting mechanism is then applied to obtain the final sentiment data, represented by Ω^t. This sentiment data Ω^t is subsequently concatenated with the time series data $\mathbb{D}^t_{time}$ and the holding state H^t to form the state vector s^t. The PPO agent receives s^t and outputs an action a^t, after which the environment provides feedback in the form of a reward, enabling the PPO agent to update its policy. Detailed steps are provided in Algorithm 1.

News and social media data are processed using five different LLMs, each designed with specific prompts to rate the news based on five criteria: Regulatory Impact, Technological Impact, Market Adoption, Macroeconomic Implications, and Overall Sentiment. Social media data are evaluated using additional criteria, such as virality potential, informative value, sentiment polarity, and duration of impact, reflecting the transient nature of social media posts.

These criteria are selected by advanced LLMs, including GPT-4o, Claude 3.5, and Google Gemini. Sentiment scores are assigned on a scale of 0 to 10, where 0 represents extremely negative sentiment, 10 indicates extremely positive sentiment, and 5 denotes neutrality or irrelevance to that criterion. Each score is explained with examples in the prompts to provide context. An outlier detection process is subsequently applied to the sentiment scores to eliminate anomalous values potentially resulting from LLM hallucinations, thereby enhancing the reliability of the sentiment analysis. The outlier detection procedure is detailed in Algorithm 2, with φ set to 1.5.

Algorithm 2 Outlier Detection Process

Input: Sentiment scores from different LLMs: ϱ^t_1, ϱ^t_2, ϱ^t_3, ϱ^t_4, ϱ^t_5. Parameter to control the tolerance of the anomalous values: φ.
Output: Processed average sentiment score ϱ^t_{avg}
$Q^t_1 = Quartile_{0.25}\ (\varrho^t_1, \varrho^t_2, \varrho^t_3, \varrho^t_4, \varrho^t_5)$
$Q^t_3 = Quartile_{0.75}\ (\varrho^t_1, \varrho^t_2, \varrho^t_3, \varrho^t_4, \varrho^t_5)$
$IQR^t = Q^t_3 - Q^t_1$
$LowerBound = Q^t_1 - \varphi \times IQR^t$
$UpperBound = Q^t_3 + \varphi \times IQR^t$
for $i \leftarrow 1$ **to** 5 **do**
  if $\varrho^t_i < LowerBound$ **or** $\varrho^t_i > UpperBound$ **then**
    $\varrho^t_i = 5$
  else
    $\varrho^t_i = \varrho^t_i$
  end if
end for
ϱ^t_{avg} = Average(ϱ^t_1, ϱ^t_2, ϱ^t_3, ϱ^t_4, ϱ^t_5)

Next, the cleaned sentiment scores are processed using a forgetting mechanism, which applies the Ebbinghaus Forgetting Curve to model the diminishing impact of news or social

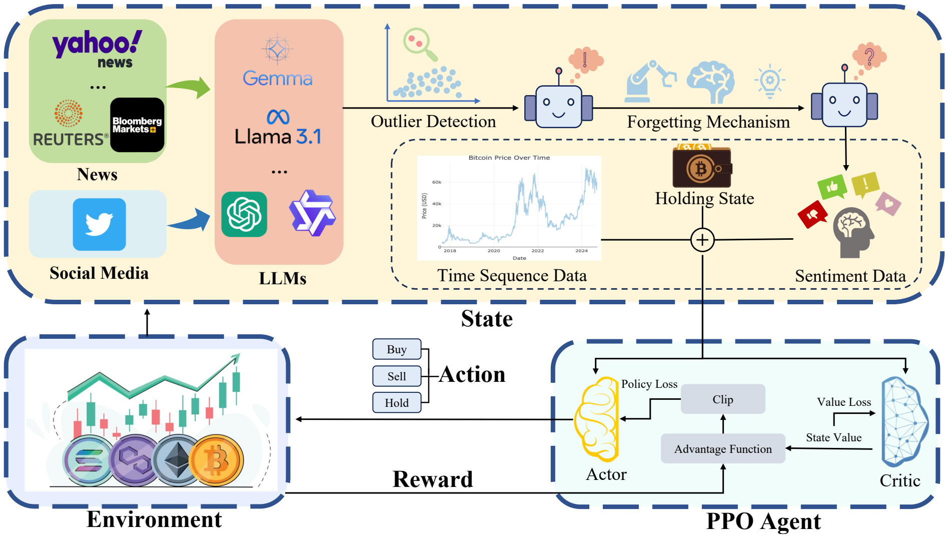

Fig. 1. Block diagram of proposed approach using proximal policy optimization algorithms(PPO) which integrating multi-source LLM sentiment analysis.

media posts on market sentiment over time. The forgetting mechanism is defined as follows:

$$\Omega^t = e^{-\frac{t}{\digamma}} \times \varrho^t_{avg}, \tag{7}$$

where Ω^t represents the residual impact on market sentiment, and $\digamma$ is a factor controlling the rate of decay over time t. When there are no new updates, the influence of the sentiment decays, with $\digamma$ set to 6.6 as a hyper-parameter, ensuring that the sentiment impact diminishes to neutral within 24 hours. This decay ensures that the relevance of each sentiment gradually fades with the passage of time.

A. Action

In typical single-asset trading, there are three possible actions: buy, sell, and hold. We define the agent's action at time t as a^t. If the agent executes consecutive **Buy** or **Sell** actions, the action a^t must transition to **Hold**. This constraint arises from the "Buy-All" and"Sell-All" strategies employed, where the agent commits all available resources to either buy or sell, leaving no assets for further action until the holding state changes. As a result, consecutive **Buy** or **Sell** actions are considered invalid, and the agent is required to **Hold** when there are no remaining assets to execute the original action.

B. State transition

The state transition form s^t to s^{t+1} is followed by the state transition function:

$$s^{t+1} = \omega(s^t) = \omega([H^t, o^t, h^t, l^t, c^t, v^t, \Omega^t]), \tag{8}$$

The o^{t+1}, h^{t+1}, l^{t+1}, c^{t+1}, v^{t+1}, Ω^t are basically affected by the market uncertainties but the holding state H^{t+1} is only affected by a^t and H^t .

C. Reward

The reward function is formulated as follows:

$$R(\mathbf{s}^t, \mathbf{a}^t) = \begin{cases} \frac{c^{t+1}}{c^t} - \beta, & \text{if } a^t == \text{Buy and } H^t == 0, \\ -\frac{c^{t+1}}{c^t} - \beta, & \text{if } a^t == \text{Sell and } H^t == 1, \\ \frac{c^{t+1}}{c^t}, & \text{if } a^t == \text{Hold and } H^t == 1, \\ \frac{c^{t+1}}{c^t} \times 0.1, & \text{else .} \end{cases} \tag{9}$$

IV. Experimental Results

A. Datasets and Experimental Settings

1) Time Sequence Data: For this study, we selected three widely traded cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and Dogecoin (DOGE). The time series data consists of hourly Open, High, Low, Close, and Volume (OHLCV) metrics. The training dataset spans from July 5, 2019, at 12:00 PM to May 6, 2023, at 8:00 PM, while the testing period extends from May 6, 2023, at 9:00 PM to October 20, 2023, at 12:00 PM. At the start of this study, there was no pre-existing dataset fitting the research objective. Thus, Yahoo Finance was chosen as a data source because it aggregates news articles from several reputable outlets such as Reuters, Bloomberg, and The Wall Street Journal. These sources were picked based on their article volume and credibility, and many archived links are available on the Internet Archive due to their popularity.

Crypto-related articles were extracted by fuzzy keyword matching in Python's Rapidfuzz library, targeting on crypto names and symbols. Articles were only included if titles had relevant keywords as bodies often mentioned cryptos peripherally and 23,379 articles matched Bitcoin, 5,063 with Ethereum, 1,305 with Dogecoin. For social media data, X (Twitter) was selected as it's one of the most widely used platforms and plays a pivotal role in sentiment analysis, especially in gauging the influence of public figures on market sentiment. For each cryptocurrency, five prominent influencers were identified, and their tweets from July 5, 2019, to June 30, 2024, were collected. Only tweets that mentioned the respective cryptocurrency and received at least 2,000 likes were included in the dataset. For Bitcoin, influencers were Elon Musk (affecting market), Brian Armstrong (Coinbase CEO), Jack Dorsey (former X CEO), Michael Saylor (MicroStrategy), and Coinbase. In the Ethereum dataset, influencers were Binance, Vitalik Buterin (Ethereum's co-founder), Coinbase, official Ethereum community, and Arbitrum(a layer-two scaling solution improving Ethereum's efficiency). In the Dogecoin dataset, Elon Musk(vocal supporter), Matt Wallace (a major influencer), Mark Cuban (contributing to growth by accepting it as payment), and Snoop Dogg (whose endorsement boosted price), with Coinbase.

2) LLM Selection: Five LLMs were selected for news data analysis: Meta Llama 3.1 8B [43], Google Gemini 2 9B [44], Alibaba Qwen 2.5 7B [45], Mistral 7B v0.3 [46], and Nous Hermes 2 Mistral 7B DPO [47]. Several factors guided this selection. First, benchmark performance was evaluated to ensure robust analytical capabilities. The models were chosen to represent a range of architectures and feature sets, optimizing their complementary strengths in sentiment analysis.

To facilitate real-time hourly sentiment analysis for future trading applications, these models were selected from serverless API platforms such as Together AI and AI/ML API. Model size and computational efficiency were also critical considerations, ensuring timely execution on available hardware within reasonable time constraints. Finally, stability and controllability of the model outputs were prioritized, as some models may deviate from instructions or produce inconsistent results without clear system prompts.

TABLE I
ANNUALIZED RETURN AND SHARPE RATIO OF DIFFERENT METHODS IN THREE CRYPTOCURRENCY INSTANCES.

		BTC	ETH	DOGE
Proposed	AR(%)	**72.34**	**19.91**	27.22
	SR	**2.53**	**2.1**	**2.69**
BAH	AR(%)	14.73	-24.9	-32.91
	SR	0.36	-0.74	-0.63
CCI	AR(%)	22.06	-0.91	**36.08**
	SR	0.84	-0.14	1.14
DIF	AR(%)	-27.92	-24.18	-63.13
	SR	-1.45	-1.05	-2.45
DMI	AR(%)	-33.87	-60.2	-68.78
	SR	-1.51	-2.3	-2.49
KDJ	AR(%)	-52.89	-63.29	-77.26
	SR	-1.79	-2.47	-2.16
MACD	AR(%)	-51.71	-57.62	-70.52
	SR	-2.64	-2.64	-2.82
MA	AR(%)	-65.9	-75.87	-85.92
	SR	-2.57	-3.61	-2.51
RSI	AR(%)	-6.67	-31.77	-51.95
	SR	-0.33	-1.49	-1.62
VOL	AR(%)	-81.32	-85	-89.71
	SR	-3.34	-4.22	-3.64

3) Result Discussion: Performance is evaluated based on annualized return (AR%) and Sharpe ratio (SR), and the proposed method is compared against several traditional strategies. These include the Buy and Hold (BAH) strategy, which involves purchasing and holding an asset long-term irrespective of market fluctuations; the Commodity Channel Index (CCI), which identifies cyclical trends and potential reversal points; the Directional Indicator (DIF) and Directional Movement Index (DMI), which assess trend direction and strength; and the KDJ Indicator (KDJ), a variation of the Stochastic Oscillator that signals overbought or oversold conditions. Other comparison methods include the Moving Average Convergence Divergence (MACD), which shows trend changes via moving averages, the Moving Average (MA), which identifies overall trend direction, the Relative Strength Index (RSI), which evaluates price momentum to indicate overbought or oversold conditions, and Volume (VOL), which tracks trading volume to confirm trends or signal potential reversals. The comparison results are presented in Table I.

Upon reviewing the results, it is evident that the proposed method performs exceptionally well for BTC and ETH in terms of both annualized return (AR%) and Sharpe ratio (SR). However, the method does not achieve the highest annualized return for DOGE, which may be attributed to the relatively limited availability of news and social media data for this asset. Specifically, the CCI method achieves a higher annualized return (36.08%) compared to the proposed method (27.22%) for DOGE. Despite this, the proposed method maintains a significantly higher Sharpe ratio (2.69) compared to CCI (1.14), indicating superior risk-adjusted returns, even with the lower absolute return. Additionally, due to the time limitation, we conducted a half-month live Bitcoin trading which applied our proposed method and achieved a 3% return.

It is also crucial to consider the impact of transaction fees on the performance of traditional methods. With a 0.1% transaction fee applied to each buy or sell action, frequent trading can substantially reduce profits, particularly in volatile markets like cryptocurrencies. Traditional strategies such as DIF, DMI, and KDJ tend to trigger more frequent trades, making them especially vulnerable to performance degradation from transaction costs. This may explain why these methods exhibit large negative returns and Sharpe ratios, as the cumulative effect of fees from numerous trades erodes overall profitability. Figure 2 clearly illustrates the evolution of return values over time for BTC, ETH, and DOGE across different methods. The proposed method demonstrates a more stable and gradual increase in return values, while traditional methods show higher volatility. Frequent peaks and troughs in the performance of traditional strategies suggest that they engage in more frequent trades, contributing to lower overall returns when accounting for transaction fees.

Table II illustrates the proportion of sentiment scores from various LLMs identified as outliers among all detected outliers. Notably, Mistral v0.3 accounts for 81.63%, 73.46%, and 67.83% of outliers in BTC, ETH, and DOGE, respectively. This significant deviation may stem from the model's poor calibration in the financial domain, where its prediction confidence often not align well with accuracy. Consequently,

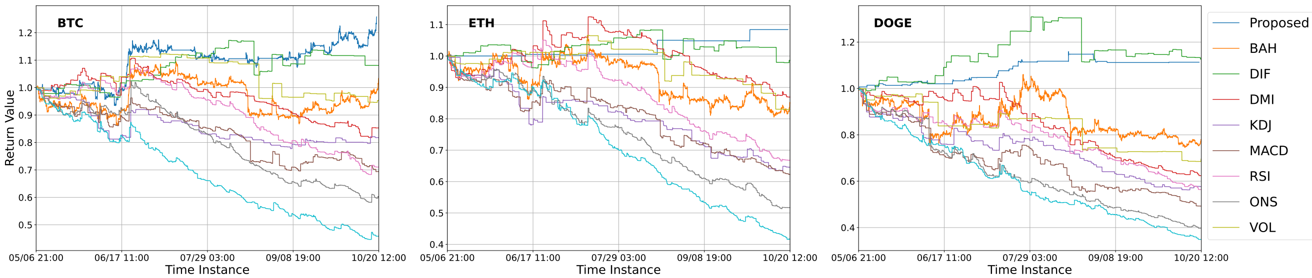

Fig. 2. Test return values of comparing our proposed approach with traditional methods

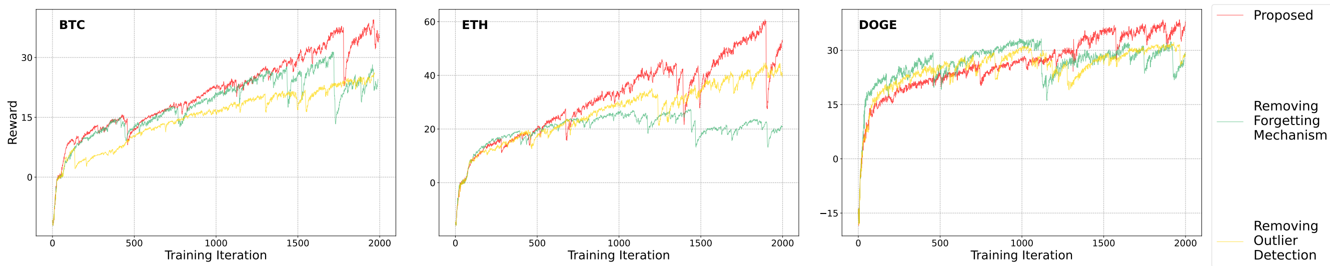

Fig. 3. Training Reward of out proposed approach with removing the outlier detection process approach and the removing forgetting mechanism approach

TABLE II
THE CONTRIBUTION OF DIFFERENT LLMS' SENTIMENT SCORE TO OUTLIERS

	BTC	ETH	DOGE
Meta Llama 3.1	11.48%	13.31%	18.76%
Google Gemini 2	6.11%	9.43%	6.48%
Mistral v0.3	81.63%	73.46%	67.83%
Alibaba Qwen	0.02%	0.54%	4.14%
Nous Hermes 2	0.76%	3.26%	2.79%

it tends to overestimate sentiment scores, leading to more extreme values classified as outliers. In contrast, Nous Hermes 2, which is fine-tuned on part of financial text data, exhibits a significantly lower percentage of outliers, underscoring the benefits of domain-specific adjustments. Similarly, Alibaba Qwen performs well, likely due to substantial improvements in instruction-following and structured data understanding, as well as enhanced robustness to diverse system prompts. This allows it to better grasp score examples at different levels, improving role-play implementation and condition-setting for chatbots.

In the ablation study, the training reward curves are presented in Figure 3. This study evaluates the impact of removing the outlier detection process and the forgetting mechanism, respectively. Notably, the cumulative reward of the proposed approach surpasses that of other methods in the final 500 episodes. The training curves for DOGE are relatively similar, likely due to the limited availability of news and social media data for DOGE, resulting in minimal benefit from the outlier detection and forgetting mechanisms. During the training period, 16.3% of sentiment scores in the BTC dataset and 10.3% in the DOGE dataset underwent outlier detection, compared to only 3.9% in the ETH dataset. Consequently, the training curves for ETH show less divergence between the proposed method and the version without outlier detection, reflecting the lower incidence of outliers in the ETH data.

V. CONCLUSION

In this work, we proposed a novel cryptocurrency trading strategy that integrates multi-source LLM sentiment analysis with a DRL approach. Our method effectively mitigates the impact of LLM hallucinations on market sentiment generation through an outlier detection process. Additionally, by implementing a forgetting mechanism, the sentiment scores more accurately reflect the actual temporal impact of news and social media posts on market sentiment. The backtesting results demonstrate that our approach significantly outperforms traditional indicator-based methods, achieving the highest annualized returns for BTC and ETH, as well as the best Sharpe ratio across all three cryptocurrency instances.

ACKNOWLEDGMENT

This work is supported in part by Ningbo Natural Science Foundation, China (Project ID 2023J194), in part by Ningbo Science and Technology Bureau (2024Z110), in part by the Ningbo Government, China (Project ID 2021B-008-C), in part by University of Nottingham Ningbo China (UNNC) Education Foundation (Project ID LDS202303), in part by the Basic and Commonweal Programme of Zhejiang Natural Science Foundation (Project ID LY24F020006), in part by Ningbo Digital Port Technologies Key Lab and UNNC Computer Vision And Intelligent Perception Laboratory.

REFERENCES

[1] E. Bouri, S. J. H. Shahzad, and D. Roubaud, "Cryptocurrencies as hedges and safe-havens for us equity sectors," *The Quarterly Review of Economics and Finance*, vol. 75, pp. 294–307, 2020.

[2] T. Cui, S. Ding, H. Jin, and Y. Zhang, "Portfolio constructions in cryptocurrency market: A cvar-based deep reinforcement learning approach," *Economic Modelling*, vol. 119, p. 106078, 2023.

[3] S. Ding, T. Cui, X. Wu, and M. Du, "Supply chain management based on volatility clustering: The effect of cbdc volatility," *Research in International Business and Finance*, vol. 62, p. 101690, 2022.

[4] S. Ding, T. Cui, and Y. Zhang, "Futures volatility forecasting based on big data analytics with incorporating an order imbalance effect," *International Review of Financial Analysis*, vol. 83, p. 102255, 2022.

[5] S. Ding, T. Cui, A. G. Bellotti, M. Z. Abedin, and B. Lucey, "The role of feature importance in predicting corporate financial distress in pre and post covid periods: Evidence from china," *International Review of Financial Analysis*, vol. 90, p. 102851, 2023.

[6] S. Ding, A. Wang, T. Cui, A. M. Du, and X. Zhou, "Commodity market stability and sustainable development: The effect of public health policies," *Research in International Business and Finance*, vol. 70, p. 102386, 2024.

[7] TradingstrategyGuides. (2019) Eos cryptocurrency trading strategy. [Online; accessed 29 January 2020]. [Online]. Available: https://tradingstrategyguides.com/eos-cryptocurrency-trading-strategy/

[8] ——. (2019) Free omni cryptocurrency strategy. [Online; accessed 29 January 2020]. [Online]. Available: https://tradingstrategyguides.com/freeomni-cryptocurrency-strategy/

[9] S. Ha and B.-R. Moon, "Finding attractive technical patterns in cryptocurrency markets," *Memetic Computing*, vol. 10, pp. 301–306, 2018.

[10] K. Grobys, S. Ahmed, and N. Sapkota, "Technical trading rules in the cryptocurrency market," *Finance Research Letters*, vol. 32, p. 101396, 2020.

[11] K. H. Al-Yahyaee, W. Mensi, H.-U. Ko, S.-M. Yoon, and S. H. Kang, "Why cryptocurrency markets are inefficient: The impact of liquidity and volatility," *The North American Journal of Economics and Finance*, vol. 52, p. 101168, 2020.

[12] W. Brock, J. Lakonishok, and B. LeBaron, "Simple technical trading rules and the stochastic properties of stock returns," *The Journal of finance*, vol. 47, no. 5, pp. 1731–1764, 1992.

[13] T. Cui, S. Cheng, and R. Bai, "A combinatorial algorithm for the cardinality constrained portfolio optimization problem," in *IEEE Congress on Evolutionary Computation (CEC)*, July 2014, pp. 491–498.

[14] T. Cui, R. Bai, A. J. Parkes, F. He, R. Qu, and J. Li, "A hybrid genetic algorithm for a two-stage stochastic portfolio optimization with uncertain asset prices," in *2015 IEEE Congress on Evolutionary Computation (CEC)*, 2015, pp. 2518–2525.

[15] T. Cui, R. Bai, S. Ding, A. J. Parkes, R. Qu, F. He, and J. Li, "A hybrid combinatorial approach to a two-stage stochastic portfolio optimization model with uncertain asset prices," *Soft Computing*, vol. 24, no. 4, pp. 2809–2831, 2020.

[16] M. Nakano, A. Takahashi, and S. Takahashi, "Bitcoin technical trading with artificial neural network," *Physica A: Statistical Mechanics and its Applications*, vol. 510, pp. 587–609, 2018.

[17] J. Sun, Y. Zhou, and J. Lin, "Using machine learning for cryptocurrency trading," in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*. IEEE, 2019, pp. 647–652.

[18] Z. Kakushadze, "101 formulaic alphas," *Wilmott*, vol. 2016, no. 84, pp. 72–81, 2016.

[19] R. Ślepaczuk and M. Zenkova, "Robustness of support vector machines in algorithmic trading on cryptocurrency market," *Central European Economic Journal*, vol. 5, no. 52, pp. 186–205, 2018.

[20] J. Singh, R. Thulasiram, and A. Thavaneswaran, "Lstm based algorithmic trading model for bitcoin," in *2022 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 2022, pp. 344–351.

[21] J. Moody and M. Saffell, "Learning to trade via direct reinforcement," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 875–889, 2001.

[22] E. F. Fama, "The behavior of stock-market prices," *The Journal of Business*, vol. 38, no. 1, pp. 34–105, 1965.

[23] ——, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[24] D. Silver, J. Schrittwieser *et al.*, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[25] O. Vinyals, I. Babuschkin *et al.*, "Grandmaster level in starcraft ii using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, 2019.

[26] J. M. Jumper, R. Evans *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, pp. 583 – 589, 2021.

[27] T. Cui, N. Du, X. Yang, and S. Ding, "Multi-period portfolio optimization using a deep reinforcement learning hyper-heuristic approach," *Technological Forecasting and Social Change*, vol. 198, p. 122944, 2024.

[28] Y. Li, N. Du, X. Song, X. Yang, T. Cui, N. Xue, A. Farjudian *et al.*, "Cardinality and bounding constrained portfolio optimization using safe reinforcement learning," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–8.

[29] J. Jin, T. Cui, R. Bai, and R. Qu, "Container port truck dispatching optimization using real2sim based deep reinforcement learning," *European Journal of Operational Research*, vol. 315, no. 1, pp. 161–175, 2024.

[30] X. Yang, T. Cui, H. Wang, and Y. Ye, "Multiagent deep reinforcement learning for electric vehicle fast charging station pricing game in electricity-transportation nexus," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 6345–6355, 2024.

[31] Y. Ye, H. Wang, T. Cui, X. Yang, S. Yang, and M.-L. Zhang, "Identifying generalizable equilibrium pricing strategies for charging service providers in coupled power and transportation networks," *Advances in Applied Energy*, vol. 12, p. 100151, 2023.

[32] T. Cui, X. Yang, F. Jia, J. Jin, Y. Ye, and R. Bai, "Mobile robot sequential decision making using a deep reinforcement learning hyper-heuristic approach," *Expert Systems with Applications*, vol. 257, p. 124959, 2024.

[33] J. Li, W. Jiang, Y. He, Q. Yang, A. Gao, Y. Ha, E. Özcan *et al.*, "Fidrl: Flexible invocation-based deep reinforcement learning for dvfs scheduling in embedded systems," *IEEE Transactions on Computers*, vol. 74, no. 1, pp. 71–85, 2025.

[34] G. Lucarelli and M. Borrotti, "A deep reinforcement learning approach for automated cryptocurrency trading," in *Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings 15*. Springer, 2019, pp. 247–258.

[35] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep reinforcement learning that matters," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[36] Y. Patel, "Optimizing market making using multi-agent reinforcement learning," *arXiv preprint arXiv:1812.10252*, 2018.

[37] O. Sattarov, A. Muminov, C. W. Lee, H. K. Kang, R. Oh, J. Ahn, H. J. Oh, and H. S. Jeon, "Recommending cryptocurrency trading points with deep reinforcement learning approach," *Applied Sciences*, vol. 10, no. 4, p. 1506, 2020.

[38] C. Lamon, E. Nielsen, and E. Redondo, "Cryptocurrency price prediction using news and social media sentiment," *SMU Data Sci. Rev*, vol. 1, no. 3, pp. 1–22, 2017.

[39] T. R. Li, A. S. Chamrajnagar, X. R. Fong, N. R. Rizik, and F. Fu, "Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model," *Frontiers in Physics*, vol. 7, p. 98, 2019.

[40] R. Wu, "Portfolio performance based on llm news scores and related economical analysis," *Available at SSRN 4709617*, 2024.

[41] J. Bai, S. Bai, Y. Chu, Z. Cui, K. Dang, X. Deng, Y. Fan *et al.*, "Qwen technical report," *arXiv preprint arXiv:2309.16609*, 2023.

[42] A. Yang, B. Xiao, B. Wang, B. Zhang, C. Bian, C. Yin, C. Lv *et al.*, "Baichuan 2: Open large-scale language models," *arXiv preprint arXiv:2309.10305*, 2023.

[43] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur *et al.*, "The llama 3 herd of models," 2024. [Online]. Available: https://arxiv.org/abs/2407.21783

[44] G. Team, M. Riviere, S. Pathak, P. G. Sessa, C. Hardin, S. Bhupatiraju, L. Hussenot *et al.*, "Gemma 2: Improving open language models at a practical size," 2024. [Online]. Available: https://arxiv.org/abs/2408.00118

[45] A. Yang, B. Yang, B. Hui, B. Zheng, B. Yu, C. Zhou, C. Li *et al.*, "Qwen2 technical report," *arXiv preprint arXiv:2407.10671*, 2024.

[46] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand *et al.*, "Mistral 7b," 2023. [Online]. Available: https://arxiv.org/abs/2310.06825

[47] NousResearch. (2024) Nous hermes 2 mistral 7b dpo. [Online]. Available: https://huggingface.co/NousResearch/Nous-Hermes-2-Mistral-7B-DPO