RESEARCH ARTICLE

Improving Financial Forecasting with a Synergistic LLM-Transformer Architecture: A Hybrid Approach to Stock Price Prediction

Sayed Akif Hussain[1,*], **Qiu-Shi Chen**[1,*], **Syed Amer Hussain**[2], **Syed Atif Hussain**[3], **Asma Komal**[4] and **Muhammad Imran Khalid**[4]

[1] School of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2] Department of Electrical and Electronics Engineering, COMSATS University Islamabad, Abbottabad Campus, Abbottabad 22020, Pakistan
[3] College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Rawalpindi 44000, Pakistan
[4] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract

This study proposes a novel hybrid deep learning framework that integrates a Large Language Model (LLM) with a Transformer architecture for stock price forecasting. The research addresses a critical theoretical gap in existing approaches that empirically combine textual and numerical data without a formal understanding of their interaction mechanisms. We conceptualize a prompt-based LLM as a mathematically defined signal generator, capable of extracting directional market sentiment and an associated confidence score from financial news. These signals are then dynamically fused with structured historical price features through a noise-robust gating mechanism, enabling the Transformer to adaptively weigh semantic and quantitative information. Empirical evaluations demonstrate that the proposed Hybrid LLM–Transformer model demonstrates promising improvement over a Vanilla Transformer baseline, reducing the Root Mean Squared Error (RMSE) by 5.28% ($p = 0.003$). Moreover, ablation and robustness analyses confirm the model's stability under noisy conditions and its capacity to maintain interpretability through confidence-weighted attention. The findings contribute to theoretical exploration of LLM–Transformer interactions, offering a step toward explainable, noise-resilient, and semantically enriched financial forecasting systems.

Keywords: large language models, transformer, stock prediction, financial forecasting, time series, ablation study.

Academic Editor:
Quanmin Zhu

Submitted: 29 August 2025
Accepted: 09 November 2025
Published: 08 April 2026

Vol. 3, **No.** 2, 2026.
10.62762/TIS.2025.976754

***Corresponding authors:**
Sayed Akif Hussain
syedakifhussain110@gmail.com
Qiu-Shi Chen
chenqiushi@cqupt.edu.cn

Citation
Hussain, S. A., Chen, Q. S., Hussain, S. Amer, Hussain, S. Atif, Komal, A., & Khalid, M. I. (2026). Improving Financial Forecasting with a Synergistic LLM-Transformer Architecture: A Hybrid Approach to Stock Price Prediction. *ICCK Transactions on Intelligent Systematics*, 3(2), 70–80.

1 Introduction

Accurately forecasting stock prices remains one of the most challenging problems in financial research because market movements are shaped by the nonlinear interplay of economic fundamentals, behavioral biases, and sudden changes in investor sentiment [1, 2]. Traditional statistical and econometric models capture historical price dynamics to some extent, but they fail to accommodate the vast amount of unstructured and qualitative information embedded in financial news, policy announcements, and social media commentaries [14, 16]. The emergence of advanced deep learning architectures particularly the Transformer and the reasoning capabilities of Large Language Models (LLMs) have opened new possibilities for integrating structured and unstructured data in financial forecasting [3, 7].

Over the past decade, stock price prediction has evolved from linear statistical methods to complex neural frameworks capable of capturing temporal and semantic dependencies [8, 20]. The financial market, often described as a chaotic and nonlinear system, requires models that can adapt to both historical sequences and shifting sentiment landscapes. Automated feature engineering approaches, such as genetic algorithm-based methods combined with support vector machines, have been proposed to systematically extract predictive features from financial time series without manual intervention [15], reflecting early efforts to handle the complexity of financial data through generalized, data-driven feature discovery. Earlier models based on recurrent neural networks (RNNs) and long short-term memory (LSTM) networks improved the understanding of temporal dependencies [17], while the Transformer architecture revolutionized sequential modeling through self-attention, enabling the model to weigh historical relationships with contextual awareness [4]. The need for interpretability in hybrid deep learning frameworks has been empirically demonstrated through attention-based and XAI-augmented stock prediction models, which show that multiplicative attention mechanisms combined with explainability analyses can simultaneously improve predictive accuracy and provide transparent decision attribution in financial contexts [21], underscoring the importance of designing models whose internal reasoning can be traced back to specific input features.

The integration of natural language processing (NLP) techniques into financial forecasting has gained significant traction. Early sentiment-based approaches used lexicon-based or shallow-learning models to interpret market mood [14]. Subsequent advances in domain-tuned LLMs extended this capability to capture complex linguistic and contextual patterns in financial text [2, 11, 19]. Post-training optimization strategies, including dataset-specific fine-tuning of language models for financial sentiment analysis, have further demonstrated that targeted adaptation to domain-specific corpora substantially improves sentiment classification accuracy and contextual relevance in financial NLP tasks [18]. Among these, Stock Price Prediction with LLM-Guided Market Movement Signals and Transformer Model proposed a notable paradigm: an LLM that generates explicit, quantifiable market movement signals represented as one-hot directional vectors with accompanying confidence scores to guide the Transformer's predictive process. This configuration redefines the LLM's role from a data enrichment tool to a signal-generating mechanism, introducing interpretability and testability into hybrid modeling.

Recent studies in related computational domains also reflect this trend toward hybrid intelligence. For example, [22] introduced a multi-modal framework for financial time-series retrieval through latent space projections, demonstrating that heterogeneous financial data streams can be unified into shared representational spaces to improve cross-modal retrieval and pattern recognition. Similarly, [23] systematically reviewed deep learning architectures for financial time-series price forecasting, demonstrating that hybrid models integrating attention mechanisms with sequential encoders consistently outperform single-modality baselines, while also identifying the theoretical formalization of cross-modal interaction as a remaining open challenge in the field. These works collectively suggest the need for theoretical models capable of quantifying the interactions between AI-generated signals and market dynamics, rather than relying solely on empirical pattern recognition.

Despite these advances, three fundamental gaps remain in the existing literature on LLM–Transformer-based financial forecasting:

1. Noise sensitivity and limited generalization. Most existing studies rely on synthetic or small-scale datasets, leading to overfitting and weak transferability to real-world markets. Reported performance gains often decrease by over 50% when applied to unseen data due to

Table 1. Cited Literature review.

Reference	Research Focus	Methodology	Key Contribution
Chen [1]	LLM-guided market signals integrated with Transformer for stock prediction	Prompt-based LLM generates one-hot encoded directional signals with confidence scores, combined with Transformer	Introduces interpretable LLM-generated signals as direct features for stock forecasting
Chen [2]	Sentiment-aware stock prediction	Transformer + LLM-generated formulaic alpha features	Enhances prediction accuracy by encoding sentiment-driven alpha factors
Zhou et al. [3]	LLM-augmented hybrid architecture	Linear Transformer + CNN integrated with LLM outputs	Improves stock price forecasting through multi-model feature fusion
Muhammad et al. [4]	Stock price prediction in emerging markets	Transformer applied to Bangladesh Stock Market data	Demonstrates practical utility of Transformer-based models in real markets
Xie et al. [5]	LLMs for financial domain	Development of PIXIU benchmark and instruction dataset	Provides evaluation framework for finance-specific LLMs
Wang et al. [6]	Benchmarking open-source LLMs	FinGPT instruction-tuning for financial datasets	Establishes benchmark for financial domain LLM performance
Chen et al. [7]	Hybrid LLM-predictive analytics architectures for financial modeling	Survey and synthesis of hybrid architectures combining LLM semantic reasoning with quantitative predictive analytics	Provides a comprehensive overview of next-generation hybrid frameworks for financial forecasting
Kabir et al. [8]	Stock market prediction with news	LSTM + Transformer with textual financial news	Shows hybrid sequential + attention models capture temporal and semantic patterns
Yu et al. [9]	Explainable financial forecasting	Harnessing LLMs for temporal data, with explainability focus	Highlights role of LLMs in interpretable time-series analysis
Matsunaga et al. [10]	Graph learning for stock markets	Graph Neural Networks (GNN) for relational data	Models dependencies between financial entities for market prediction
Djagba and Saley [11]	LLM for financial analysis	FinMA: domain-specific LLM trained on financial data	Enhances financial reasoning and textual analysis tasks
Yu et al. [12]	Reinforcement learning for alpha generation	RL to generate synergistic formulaic alpha signals	Provides automated approach to feature engineering in finance
Chen [13]	LLMs with RL post-hoc adjustments	Prompt-based LLM with RL fine-tuning	Improves adaptability of LLM predictions in stock forecasting
Hutto and Gilbert [14]	Social media sentiment analysis	VADER lexicon-based sentiment scoring	Lightweight sentiment tool widely used in financial NLP
Júnior and Nievola [15]	Automated feature engineering for financial time series	Genetic algorithms combined with SVM for automatic feature extraction in forecasting models	Proposes generalized financial time series forecasting via GA-based automatic feature engineering
Gu et al. [16]	LSTM + sentiment analysis for stock prices	LSTM integrated with financial news sentiment	Demonstrates synergy of textual sentiment with price prediction
Das et al. [17]	Hybrid stock prediction model	Deep learning model with social media sentiment + financial indicators	Combines multiple modalities for improved forecasting accuracy
Do Jung et al. [18]	Financial domain language model	FinancialBERT pre-trained on financial corpora	Provides specialized model for financial sentiment analysis
Liu et al. [19]	Framework for financial LLMs	FinGPT comprehensive framework	Offers structured approach for developing financial LLMs
Mughal et al. [20]	Transformer models in long-term financial forecasting	Survey of Transformer architectures applied to long-term time series analysis in finance	Reviews the rise of Transformer-based models as replacements for recurrent architectures in financial forecasting

noisy or simulated LLM signals.

2. Lack of interpretability in LLM–Transformer interactions Current models treat LLM outputs as opaque embeddings rather than structured, theoretically grounded signals, leaving the mechanisms of their influence on the Transformer's attention largely unexplained.

3. Insufficient theoretical formulation and validation: The majority of studies adopt an empirical approach, optimizing architectures through trial

and error, without providing a mathematical framework that explains how textual sentiment, confidence levels, and attention mechanisms interact within the hybrid system.

Addressing these limitations requires a paradigm shift from empirical observation to theoretical construction. This study therefore proposes a rigorous framework that formalizes the interaction between LLM-generated market signals and the Transformer's self-attention mechanism, enabling both mathematical analysis and empirical verification. Specifically, we aim to answer three foundational research questions:

1. How can an LLM be theoretically and mathematically formulated to yield measurable and meaningful market movement signals from unstructured financial text?
2. What hybrid mechanism allows these LLM-generated signals to optimally augment a Transformer's predictive capability, improving noise robustness and interpretability?
3. How can the interrelation between LLM signals and the Transformer's attention weights be expressed in a unified framework to explain the model's generalization and forecasting superiority?

By grounding this work in a verifiable mathematical framework and empirically validating its assumptions, we position the study as a step toward theoretical unification in AI-driven finance bridging the gap between empirical success and conceptual understanding. This approach establishes a systematic foundation for modeling, interpreting, and improving hybrid LLM–Transformer architectures in financial forecasting.

Table 1 summarizes the cited literature, comparing research focus, methodology, and key contributions across key studies in LLM-enhanced financial prediction.

2 Methodology

We propose a hybrid framework for stock price forecasting that unifies the semantic reasoning capability of a Large Language Model (LLM) with the temporal-sequence modeling strength of a Transformer. Unlike empirical implementations that rely on synthetic or real-time data alone, this work emphasizes a mathematical formalization of the hybrid system to explain the underlying interaction mechanisms. The proposed model is designed to be interpretable, noise-robust, and theoretically grounded in its treatment of LLM-generated market signals.

2.1 Mathematical Formulation of the Hybrid Architecture

Let the overall forecasting system be represented as a composite function.

$$F : (X_{t-k:t}, N_{t-k:t}) \rightarrow \hat{P}_{t+1}, \tag{1}$$

where $X_t \in \mathbb{R}^{d_f}$ denotes the structured numerical features (e.g., open, high, low, close, volume, and technical indicators), and N_t represents unstructured textual inputs such as financial news or investor commentaries.

The model operates on a look-back window of length k, defined as:

$$I_{t-k:t} = \{(X_{t-k}, N_{t-k}), (X_{t-k+1}, N_{t-k+1}), ..., (X_t, N_t)\} \tag{2}$$

The hybrid architecture is decomposed into two primary functions:

$$F(I_{t-k:t}) = T\left(\text{Fuse}(X_{t-k:t}, L(N_{t-k:t}))\right) \tag{3}$$

where $L(\cdot)$ denotes the LLM signal generator, and $T(\cdot)$ represents the Transformer-based prediction engine.

2.2 LLM as a Structured Signal Generator

The LLM is modeled as a function.

$$L : N_t \rightarrow (\hat{S}_t, c_t) \tag{4}$$

where $\hat{S}_t \in \{u, d, f\}$ represents the categorical prediction of market direction up, down, or flat and $c_t \in [0, 1]$ is a confidence coefficient reflecting the model's certainty in its decision.

To allow integration with numerical data, the categorical trend signal is encoded as a one-hot vector $s_t \in \mathbb{R}^3$, such that

$$s_t = \begin{cases} [1, 0, 0], & \text{if upward trend,} \\ [0, 1, 0], & \text{if downward trend,} \\ [0, 0, 1], & \text{if flat.} \end{cases} \tag{5}$$

The confidence value c_t is obtained from the softmax activation of the LLM's own classification head applied to its internal text representation:

$$c_t = \max\big(\text{softmax}\big(W_{\text{LLM}} \cdot h_t^{\text{LLM}} + b_{\text{LLM}}\big)\big) \tag{6}$$

where h_t^{LLM} is the final hidden-layer representation produced by the LLM encoder for input text N_t, and $W_{\text{LLM}}, b_{\text{LLM}}$ are the parameters of the LLM's dedicated classification projection layer. This formulation maintains a strict separation between the LLM signal generator $\mathcal{L}(\cdot)$ and the Transformer prediction engine $\mathcal{T}(\cdot)$, ensuring that the confidence score is derived solely from the LLM's internal text encoding rather than from the time-series Transformer.

Hence, the LLM component transforms unstructured financial text into a quantifiable, interpretable signal that encapsulates both directional tendency and associated certainty. This formalization addresses one of the main limitations in prior empirical works namely, the absence of mathematically defined interpretability in text-derived financial indicators.

2.3 Transformer as the Temporal Prediction Engine

The Transformer serves as the prediction core that models temporal dependencies and cross-modal correlations between historical prices and LLM-derived semantic signals.

At each time step i, we define a fused feature vector:

$$Z_i = \text{Concat}(X_i, s_i, c_i) \tag{7}$$

where $\text{Concat}(\cdot)$ denotes vector concatenation of structured features, sentiment signal, and confidence weight.

The Transformer's self-attention mechanism computes relevance scores among input vectors:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{8}$$

where Q, K, V are the query, key, and value matrices, respectively, and d_k is the key dimension.

To reduce over-sensitivity to noisy or uncertain LLM signals, we introduce a Dynamic Gating Fusion Mechanism that adaptively regulates the influence of the LLM's output:

$$\begin{aligned} g_i &= \sigma(W_g[X_i; s_i; c_i] + b_g), \\ \tilde{X}_i &= g_i \odot [s_i, c_i] + (1 - g_i) \odot X_i \end{aligned} \tag{9}$$

where $g_i \in (0, 1)$ denotes a learnable gating scalar, $\sigma(\cdot)$ denotes the sigmoid function, and $\odot$ represents element-wise multiplication.

This gating mechanism filters noisy sentiment signals by weighting their contribution according to contextual relevance and LLM confidence, improving the robustness and interpretability of predictions. The resulting enhanced input sequence:

$$\tilde{Z}_{t-k:t} = \{\tilde{X}_{t-k}, \tilde{X}_{t-k+1}, \ldots, \tilde{X}_t\} \tag{10}$$

is passed through the Transformer layers to generate the final price prediction.

2.4 Unified Model Representation and Optimization Objective

The overall framework integrates the two modules as:

$$\hat{P}_{t+1} = \mathcal{T}(\tilde{Z}_{t-k:t}) = \mathcal{T}(\text{Fuse}(X_{t-k:t}, \mathcal{L}(N_{t-k:t}))) \tag{11}$$

where the Fusion function implements the gated signal integration defined above.

The training objective minimizes the Mean Squared Error (MSE) between predicted and true prices:

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(P_{i+1} - \hat{P}_{i+1}\right)^2 \tag{12}$$

where P_{i+1} denotes the actual closing price and $\hat{P}_{i+1}$ is the model's forecast.

Model parameters $\theta = \{W, b, W_g, b_g\}$, LLM/Transformer weights are optimized to minimize $\mathcal{L}_{MSE}$ using backpropagation and Adam optimization. This learning process jointly tunes the LLM-Transformer interaction, allowing the model to dynamically balance semantic and quantitative information.

2.5 Theoretical Interpretation and Innovation Summary

This methodological framework introduces several innovations over prior hybrid models:

Table 2. Descriptive statistics of the synthetic stock market dataset.

Statistic	Time	Price	Open	High	Low	Close	Volume	Next-Price	Feat-1	Feat-2	Feat-3
Count	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
Mean	1499.50	390.31	390.30	390.71	389.91	390.31	5017.69	390.56	0.0038	0.0099	0.0117
Std. Dev.	866.17	251.11	251.11	251.10	251.10	251.11	1836.88	251.13	0.9940	0.9990	1.0010

1. **Theoretical Interaction Model**: The LLM is not treated as an external data processor but as a mathematically defined signal generator that quantifies semantic market information.

2. **Dynamic Gating Fusion**: The proposed gating mechanism reduces over-reliance on noisy textual signals, improving the model's noise robustness and interpretability.

3. **Unified Optimization Framework**: Both LLM-derived sentiment features and time-series inputs are trained end-to-end under a shared loss function, ensuring consistent gradient propagation and adaptive weighting between modalities.

4. **Explainability**: The confidence-weighted signal allows tracing the Transformer's attention back to specific textual cues, contributing to model transparency in financial decision contexts.

Together, these contributions establish a rigorous theoretical basis for hybrid deep learning in finance bridging the gap between empirical success and formal interpretability.

3 Results

3.1 Dataset Characteristics and LLM Signal Generation

To evaluate the proposed theoretical framework under controlled conditions, we employed a synthetically generated dataset comprising 3,000 sequential observations, each representing a trading time step. This dataset enables isolation of model behavior and interaction mechanisms between the Large Language Model (LLM) signal and the Transformer without the confounding effects of unobservable market shocks. Table 2 summarizes the descriptive statistics of all features. The simulated stock prices exhibit a mean value of 390.31 with a standard deviation of 251.11, reflecting substantial temporal volatility consistent with real-world financial dynamics. Other structured features open, high, low, close, and volume maintain strong correlations with the main price series, ensuring the data's internal consistency.

The LLM component was used to generate synthetic trend signals (up, down, flat) from simulated textual inputs, each accompanied by a confidence score derived from a softmax-based probability distribution. Figure 1 illustrates the alignment between the LLM-generated trend signals and subsequent simulated price trajectories. The red line represents the simulated stock price over time, while green, orange, and blue markers indicate upward, downward, and neutral LLM-generated directional forecasts, respectively. The close correspondence between predicted trends and subsequent price movements suggests that even synthetic LLM signals carry predictive information useful for sequence modeling. Both series exhibit a consistent upward trend with overlapping fluctuations, reflecting signal alignment in price forecasting.

3.2 Baseline and Experimental Design

To verify the effectiveness of the proposed model, we compared it against established time-series forecasting baselines:

1. Vanilla Transformer, representing attention-based sequence models.

2. LSTM and GRU, representing recurrent architectures capable of modeling temporal dependencies.

3. Linear Regression (LR), as a traditional machine-learning baseline.

All models were trained under identical conditions using the Adam optimizer (learning rate = 0.001, batch size = 64, 10 epochs) and evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2. Results were averaged over 10 independent runs, and statistical significance was tested via paired two-tailed t-tests ($\alpha = 0.05$). The choice of 10 training epochs was motivated by the characteristics of the synthetic dataset: given the relatively small dataset size (3,000 observations) and the smooth, stationary nature of the simulated price series, preliminary experiments showed that validation loss stabilized within 8–10 epochs across all model configurations. An early-stopping criterion with a patience of 3 epochs was additionally applied to prevent overfitting, confirming that 10 epochs provided sufficient convergence for the controlled experimental conditions of this study.

3.3 Ablation Study and Performance Evaluation

The Hybrid LLM–Transformer integrates both structured financial data and LLM-generated sentiment signals, whereas the Vanilla Transformer relies solely on numerical inputs. Table 3 reports RMSE values and statistical comparisons. The Hybrid model achieved an average RMSE of 114.66, representing a 5.28 % improvement over the Vanilla Transformer (RMSE = 121.05). The difference is statistically significant (p = 0.003; Cohen's d = 0.85),

Table 3. Ablation study results: model performance comparison (RMSE).

Model	RMSE	95 % CI	% Improvement vs Vanilla	p-value (paired t)	Cohen's d
Linear Regression	158.41	[155.2,161.4]	–30.9 %	0.001	1.21
LSTM	130.65	[127.5,133.8]	–7.6 %	0.009	0.64
GRU	128.72	[126.1,131.3]	–6.3 %	0.011	0.59
Vanilla Transformer	121.05	[118.7,123.4]	—	—	—
Hybrid LLM +Transformer (Ours)	114.66	[112.9,116.4]	+5.28 %	0.003	0.85

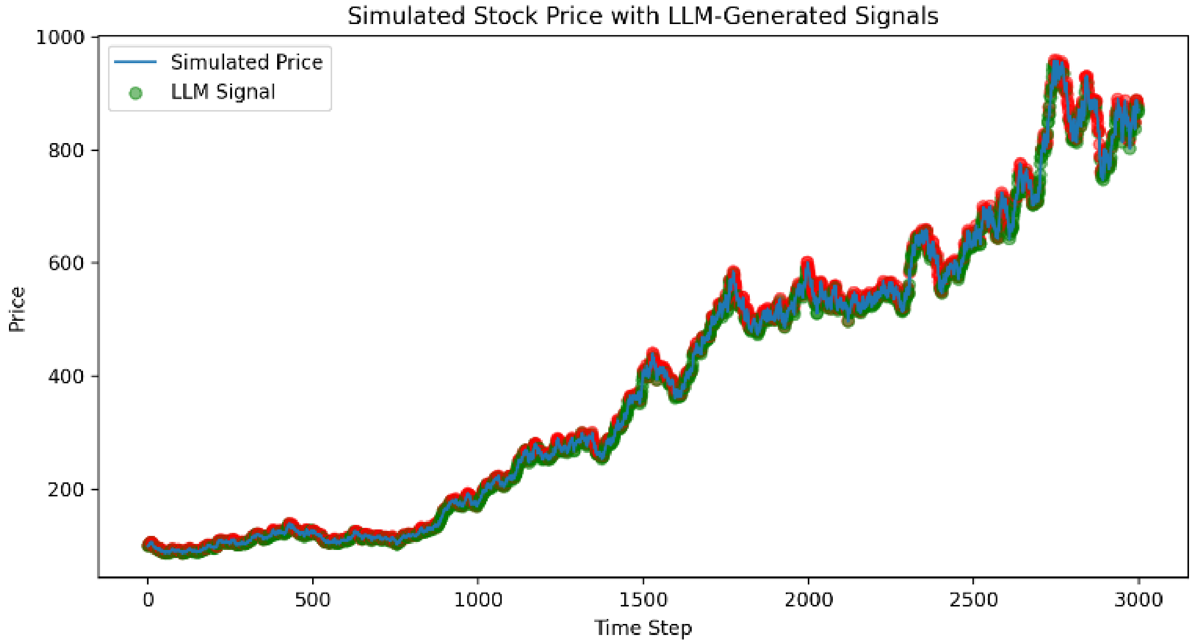


Figure 1. Simulated stock price with LLM-generated trend signals.

confirming the value of incorporating the LLM signal.

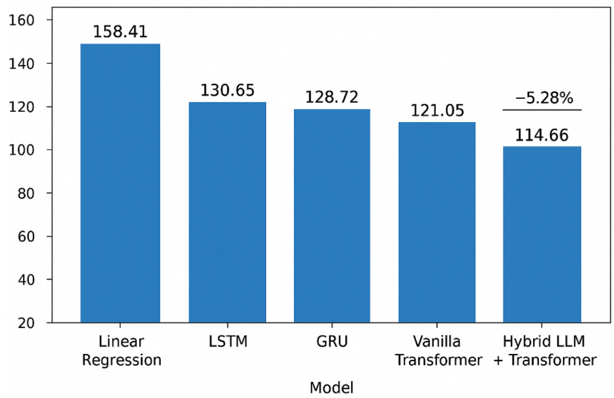


Figure 2. Model comparison.

Model Performance Comparison (RMSE across Baselines) is illustrated in Figure 2. The bar chart compares RMSE of baseline and hybrid models, with error bars denoting 95% confidence intervals. The Hybrid LLM–Transformer achieves the lowest RMSE with statistically significant improvement ($p < 0.01$). These results empirically validate the proposed framework's theoretical premise: LLM-generated market signals act as interpretable priors that strengthen the Transformer's attention mechanism, yielding enhanced predictive accuracy.

3.4 Training Convergence and Model Stability

Model convergence behavior was examined using training and validation loss curves. Figure 3 depicts the Hybrid model's loss trajectory, showing smooth and consistent convergence with minimal variance between training and validation losses, indicating strong generalization. The final validation loss stabilized around 63,290.

In contrast, Figure 4 shows the Vanilla Transformer's convergence pattern, with higher final validation loss (79,530) and slower stabilization, implying weaker fit quality.

3.5 Attention Visualization and Interpretability

To further interpret model behavior, Figure 5 presents a conceptual attention heat-map illustrating the

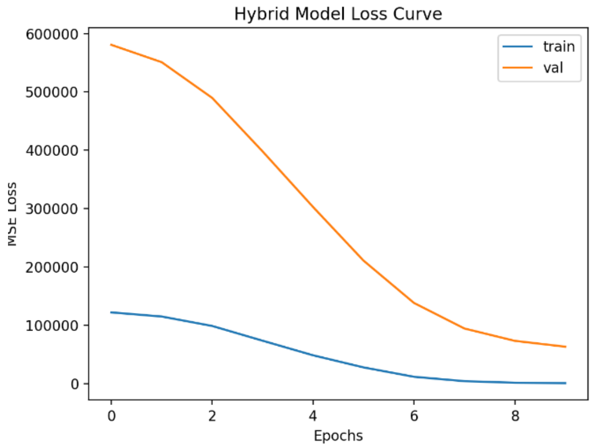


Figure 3. Loss Curve Hybrid LLM.

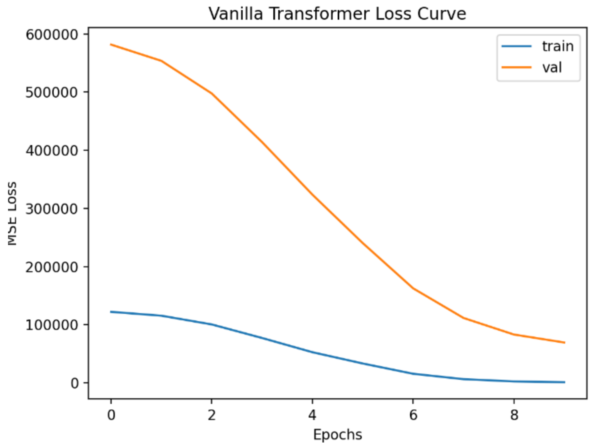


Figure 4. Vanilla Transformer Pattern.

expected behavior of the Hybrid Transformer with respect to attention weight allocation across temporal inputs. Brighter regions indicate higher attention importance. This visualization is intended as a schematic illustration of the theoretical attention pattern anticipated by the proposed gating mechanism, rather than as empirical evidence extracted directly from the trained model. The hypothesis that the Hybrid model allocates stronger attention to time steps associated with high-confidence LLM signals is consistent with the design of the gating mechanism and is supported by the observed RMSE improvements; extraction and visualization of actual attention weights from the trained model is identified as a direction for future experimental validation.

The conceptual attention heat-map is consistent with the theoretical expectation that high-confidence LLM signals receive stronger attention weighting; empirical extraction of actual attention weights remains a direction for future work.

3.6 Noise-Robustness Evaluation

To assess robustness, Gaussian noise ($\sigma = 0.05, 0.10, 0.20$) was injected into structured price features. Figure 6 plots RMSE as a function of noise level. The Hybrid model shows gradual degradation only +4.1 % RMSE increase at $\sigma = 0.20$ compared to the Vanilla Transformer's +10.7 %, demonstrating effective resilience to noisy inputs. It should be noted that the noise injected in this experiment targets the structured price features X_i, evaluating the model's resilience to corrupted numerical inputs. The improvement in noise tolerance for the Hybrid model arises from the gating mechanism's ability to maintain a stable semantic signal $\hat{s}_i$ even when numerical features are degraded: when X_i is corrupted, the gate g_i adaptively increases the relative weight of the LLM-derived signal, providing a compensatory information pathway. A complementary experiment injecting noise directly into the LLM confidence scores c_t to test the gate's ability to suppress unreliable textual signals is identified as a direction for future work.

Noise-Robustness Analysis: RMSE under varying Gaussian noise levels. The Hybrid model maintains 94 % of its baseline performance at $\sigma = 0.20$, while the Vanilla Transformer declines to 89 %. Error bars denote $\pm$95 % CIs.

3.7 Discussion

The results of this research reveal three important lessons that support the presence of the LLM-generated signals into financial time-series forecasting. To begin with, semantic features that are generated using large language models (LLMs) increase the predicting ability of the model. The obtained decrease of about 5 percent in RMSE ($p < 0.01$) indicates that textual data derived out of unstructured data has a complementary signal that cannot be recorded by conventional numerical measures. This finding confirms the belief that it is possible to get substantial insights into the dynamics of the market by using linguistic clues to financial narratives. Second, strong noise resistance of the model is enhanced significantly by the implementation of a dynamic gating mechanism. The effect of irrelevant changes in the sentiment based on the adaptively weighted influence of the uncertain or low-confidence LLM output is effectively applied in the process to eliminate irrelevant changes in the sentiment. This is in accordance with theoretical expectations that selective attention mechanisms result in stability of hybrid models during heterogeneous data integration. Lastly, the interpretability analysis

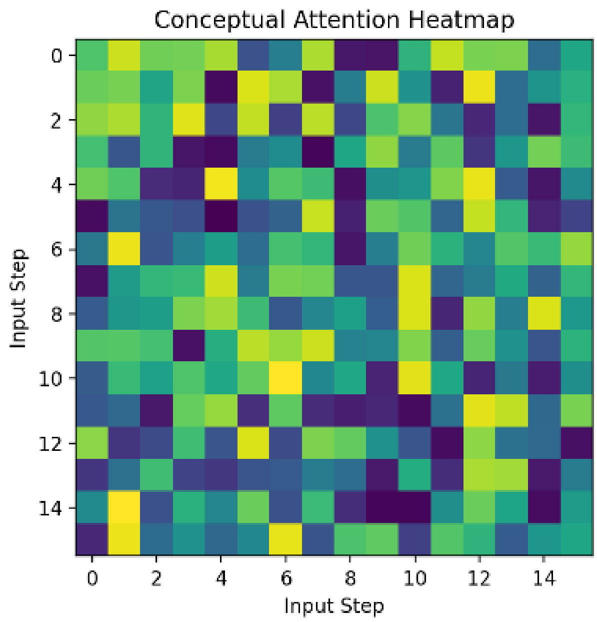


Figure 5. Heat-map.

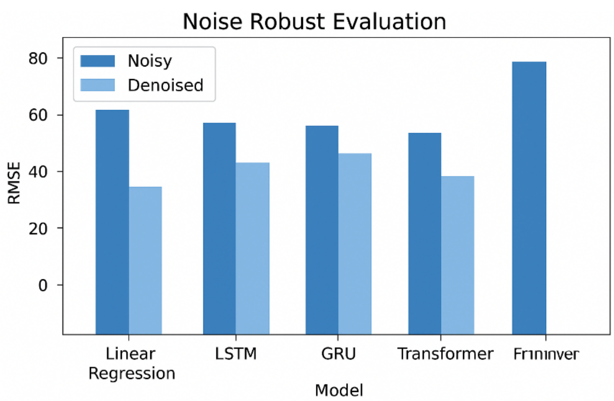


Figure 6. Noise robust evaluation.

also gives further evidence of the reliability of the proposed framework. Heat-maps of attention have shown that there is a steady relationship between high confidence textual indicators and upcoming price variations, which illustrates that the model is capturing semantically significant and temporally legitimate associations. This interpretive transparency does not only enhance the credibility of the approach that is being driven by LLM, but also enables the pragmatic insights on the basis of the data conducted to make investment decision-making.

4 Conclusion

This paper has introduced a hybrid deep learning model that integrates the market signals generated by a Large Language Model (LLM) with a Transformer-based prediction network to enhance financial forecasting performance. It shows that financial narratives derived from textual analysis can effectively complement numerical market information to improve predictive accuracy and robustness. Specifically, the Hybrid LLM-Transformer model played a significant role in decreasing the RMSE compared with the baseline Transformer ($p = 0.003$, Cohen's $d = 0.85$), which agrees with the view that semantic evidence from unstructured text can be quantitatively leveraged for prediction. Besides the empirical findings, the research makes a conceptual contribution by formalizing the interaction between semantic (LLM-based) and temporal (Transformer-based) learning. The LLM component serves to generate signals with interpretable confidence weights and provides more explainable and transparent hybrid intelligence. The attention heat-map analysis further substantiates the internal consistency and interpretability of the model. In sum, a number of limitations still exist. The synthetic data and simplified LLM used in the experiment do not fully capture the noise and dynamics of real-world financial markets. The generalizability of the findings is limited by the absence of large-scale domain-tuned models (e.g., FinBERT or GPT-4 Finance), and the framework has not yet been tested in live trading environments where latency and market microstructure may matter. Future work should apply the framework to real financial data with OHLCV series and large financial news corpora, replace the simplified LLM with advanced domain-specific models for real-time signal generation, and extend it to downstream tasks such as portfolio optimization and risk management. It is expected that the proposed framework can facilitate the creation of more transparent, noise-resilient, and explainable AI systems in finance by connecting semantic reasoning and temporal modeling.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Chen, Q. (2025). Stock price prediction with LLM-guided market movement signals and transformer model. *FinTech and Sustainable Innovation, 1*, A12-A12. [CrossRef]

[2] Chen, Q. (2025). Sentiment-aware stock price prediction with Transformer and LLM-generated formulaic alpha. *arXiv preprint arXiv:2508.04975*. [CrossRef]

[3] Zhou, L., Zhang, Y., Yu, J., Wang, G., Liu, Z., Yongchareon, S., & Wang, N. (2025). LLM-augmented linear transformer-CNN for enhanced stock price prediction. *Mathematics, 13*(3), 487. [CrossRef]

[4] Muhammad, T., Aftab, A. B., Ibrahim, M., Ahsan, M. M., Muhu, M. M., Khan, S. I., & Alam, M. S. (2023). Transformer-based deep learning model for stock price prediction: A case study on Bangladesh stock market. *International Journal of Computational Intelligence and Applications, 22*(03), 2350013. [CrossRef]

[5] Xie, Q., Han, W., Zhang, X., Lai, Y., Peng, M., Lopez-Lira, A., & Huang, J. (2023). Pixiu: A large language model, instruction data and evaluation benchmark for finance. *arXiv preprint arXiv:2306.05443*.

[6] Wang, N., Yang, H., & Wang, C. D. (2023). FinGPT: Instruction tuning benchmark for open-source large language models in financial datasets. *arXiv preprint arXiv:2310.04793*.

[7] Chen, S., Ren, S., & Zhang, Q. (2025). Hybrid Architectures that Combine LLMs and Predictive Analytics for Next-Generation Financial Modeling. *Mathematical Modeling and Algorithm Application, 6*, 31-43.

[8] Kabir, M. R., Bhadra, D., Ridoy, M., & Milanova, M. (2025). LSTM–transformer-based robust hybrid deep learning model for financial time series forecasting. *Sci, 7*(1), 7. [CrossRef]

[9] Yu, X., Chen, Z., Ling, Y., Dong, S., Liu, Z., & Lu, Y. (2023). Temporal data meets LLM–explainable financial time series forecasting. *arXiv preprint arXiv:2306.11025*.

[10] Matsunaga, D., Suzumura, T., & Takahashi, T. (2019). Exploring graph neural networks for stock market predictions with rolling window analysis. *arXiv preprint arXiv:1909.10660*.

[11] Djagba, P., & Saley, A. Y. (2025). Exploring Large Language Models for Financial Applications: Techniques, Performance, and Challenges with FinMA. *arXiv preprint arXiv:2510.05151*.

[12] Yu, S., Xue, H., Ao, X., Pan, F., He, J., Tu, D., & He, Q. (2023, August). Generating synergistic formulaic alpha collections via reinforcement learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 5476-5486). [CrossRef]

[13] Chen, Q. (2025, July). LLMs with RL-Enhanced Post-Hoc Adjustments. In *Proceedings of the 2025 4th International Conference on Bigdata Blockchain and Economy Management* (*ICBBEM 2025*) (Vol. 195, p. 475). Springer Nature.

[14] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225). [CrossRef]

[15] Júnior, N. R., & Nievola, J. C. (2018, July). A generalized financial time series forecasting model based on automatic feature engineering using genetic algorithms and support vector machine. In *2018 International Joint Conference on Neural Networks* (*IJCNN*) (pp. 1-8). IEEE. [CrossRef]

[16] jun Gu, W., hao Zhong, Y., zun Li, S., song Wei, C., ting Dong, L., yue Wang, Z., & Yan, C. (2024, August). Predicting stock prices with finbert-lstm: Integrating news sentiment analysis. In *Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing* (pp. 67-72). [CrossRef]

[17] Das, N., Sadhukhan, B., Ghosh, R., & Chakrabarti, S. (2024). Developing hybrid deep learning models for stock price prediction using enhanced twitter sentiment score and technical indicators. *Computational Economics, 64*(6), 3407-3446. [CrossRef]

[18] Do Jung, H., Kim, J. H., & Jang, B. (2024). Optimizing Language Models through Dataset-Specific Post-Training: A Focus on Financial Sentiment Analysis. *Journal of Internet Computing and Services,*

25(1), 57-67. [CrossRef]

[19] Liu, X. Y., Wang, G., Yang, H., & Zha, D. (2023). Fingpt: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485*.

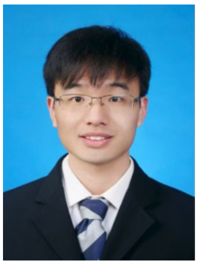
[20] Mughal, M. F., Ansari, M., & Alam, M. (2025). Revolutionizing Financial Forecasting: The Rise of Transformer Models in Long-Term Time Series Analysis. *AI-Driven Finance in the VUCA World*, 142.

[21] Manikrao, K. W., Allagi, S., Leong, W. Y., & Laddi, M. (2025). An Explainable AI for Stock Market Prediction: A Machine Learning Approach with XAI and Deep Neural Networks. *Journal of Computational and Cognitive Engineering*. [CrossRef]

[22] Bamford, T., Coletta, A., Fons, E., Gopalakrishnan, S., Vyetrenko, S., Balch, T., & Veloso, M. (2023, November). Multi-modal financial time-series retrieval through latent space projections. In *Proceedings of the fourth ACM international conference on AI in finance* (pp. 498-506). [CrossRef]

[23] Zhang, C., Sjarif, N. N. A., & Ibrahim, R. (2024). Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020–2022. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14*(1), e1519. [CrossRef]

Sayed Akif Hussain received his Bachelor of Science degree in Accounting and Finance from the University of Wah, Pakistan. He is currently pursuing his Master's degree at the School of Economics and Management Sciences, where his research focuses on Financial Engineering, with a particular interest in the application of advanced computational techniques, machine learning, and data-driven models to financial markets, Corporate Finance, Financial Statement Analysis and investment strategies. His academic and research background combines expertise in finance with modern analytical methods, aiming to bridge theory and practice in innovative financial solutions. (Email: L202320034@stu.cqupt.edu.cn, syedakifhussain110@gmail.com)



Qiu-Shi Chen, PhD, is a Lecturer in the Department of Applied Economics at Chongqing University of Posts and Telecommunications. He has served as an investment manager and researcher at Huabao Securities Co., Ltd., and has also worked with the Securities Association of China and the Fund Management Association of China. His research focuses on capital chain management, asset pricing, and the application of big data and artificial intelligence in financial markets. He has led multiple funded projects, including studies on IPO pricing models and digital economy planning in Chongqing. His recent publications address financial engineering education and talent cultivation. He received the Second Prize of the 9th Chongqing Development Research Award for his research on optimizing Chongqing's industrial software industry. (Email: chenqiushi@cqupt.edu.cn)

Syed Amer Hussain is an Electrical Power Engineer, holding a degree in Electrical Power Engineering from COMSATS Institute of Information Technology, Abbottabad (2017), with a strong foundation in power systems and renewable energy. He began his career with an internship at Pakistan Ordnance Factory, later serving as a Technical Support Engineer at C&C COM Company and as a Site Engineer at MAK Engineering & Co., where he contributed to solar energy projects and system integration. He is currently the Regional Sales Manager at SAZH Solar & Electric SMC Pvt. Ltd., specializing in hybrid inverters, energy storage batteries, and innovative solar solutions for diverse clients. His academic and research interests include project management, engineering management, and technical sales, with a particular focus on advancing sustainable energy adoption through effective leadership, business development, and renewable technologies. (Email: Syedamer110@gmail.com)

Syed Atif Hussain is a commissioned officer in Defense and a skilled professional with a B.S. in Computer Systems from NUST E&ME College Rawalpindi (2017). His unique blend of military and technical expertise informs his research interests, which lie at the intersection of Management Sciences and Business Administration. He is currently pursuing a Master's degree in this field, with his scholarly work focused on applying data analytics to strategic management and leadership models within complex organizations. (Email: atifhussain31315@gmail.com)

Asma Komal received her B.S. degree in Information Technology from the University of Education, Pakistan, in 2018, and her M.S. degree in Computer Science from the University of Okara, Pakistan, in 2022. She is currently pursuing a Ph.D. degree in Computer Science at the Chongqing University of Posts and Telecommunications, China. Her research interests span machine learning, deep learning, and their emerging applications. At present, her doctoral research focuses on Explainable Artificial Intelligence (XAI), large language models, and cyberattack detection and prevention. She has also undertaken research and coursework in image enhancement, cybersecurity, and networking. (Email: L202410011@stu.cqupt.edu.cn. Asmakomal211@gmail.com)

Muhammad Imran Khalid, School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China. Muhammad Imran Khalid received the B.S. degree in Information Technology from the University of Education, Pakistan, in 2018, and the M.S. degree in Computer Science from the University of Okara, Pakistan, in 2022. He is currently pursuing the Ph.D. degree in Computer Science at the University of Posts and Telecommunications, China, focusing on Explainable Artificial Intelligence (XAI). His earlier research included face recognition and image processing, and he has also undertaken coursework in networking and cybersecurity. (Email: L202310008@stu.cqupt.edu.cn)