Transformer-Based Deep Learning for Financial Time-Series Forecasting: A Multi-Horizon Prediction Framework

Julian R. Sterling, PhD
Department of Systems Engineering, Colorado School of Mines
jsterling@mines.edu

Elara V. Montgomery, PhD
School of Computing and Information, University of Pittsburgh
e.montgomery@pitt.edu

Abstract

The rapid evolution of deep learning architectures has fundamentally altered the landscape of financial econometrics and predictive modeling. Traditional linear and autoregressive models, while foundational to financial theory, often fail to capture the high-frequency volatility, non-linear dependencies, and long-range temporal correlations inherent in modern globalized markets. This paper explores the transition toward attention-based mechanisms, specifically focusing on Transformer architectures as a robust framework for multi-horizon financial time-series forecasting. Unlike recurrent structures that suffer from vanishing gradients and sequential processing bottlenecks, the self-attention mechanism enables the simultaneous processing of vast historical datasets, facilitating the identification of structural breaks and regime shifts across multiple temporal scales. This research provides a comprehensive systems-level analysis of the integration of Transformer models within socio-technical financial infrastructures. We examine the architectural trade-offs between computational complexity and predictive accuracy, the role of positional encoding in preserving temporal order, and the systemic implications of deploying such models in high-stakes trading environments. Furthermore, the paper addresses critical dimensions of algorithmic governance, including the interpretability of attention weights, the ethical considerations of market-wide model convergence, and the environmental sustainability of large-scale deep learning deployments. By synthesizing insights from computer science, financial engineering, and public policy, we propose a multi-horizon framework that balances predictive power with systemic stability, offering a roadmap for the next generation of resilient financial AI systems.

Keywords:
Transformer Architectures, Financial Time-Series, Multi-Horizon Forecasting, Socio-Technical Systems, Algorithmic Governance, Deep Learning, Market Infrastructure.

1. Introduction

The conceptualization of financial markets as complex, adaptive systems has long necessitated analytical tools capable of navigating stochastic environments and emergent behaviors. In the contemporary era, the sheer volume and velocity of financial data have rendered classical econometric methods increasingly insufficient for capturing the nuanced dynamics of asset pricing, risk management, and liquidity provision. While the Efficient Market Hypothesis suggests that all known information is already reflected in prices, the reality of market friction, behavioral biases, and information asymmetry creates temporal windows where sophisticated modeling can yield significant predictive advantages. The introduction of deep learning, particularly the emergence of the Transformer architecture, represents a paradigm shift in how researchers and practitioners approach these temporal challenges. By moving beyond the Markovian assumptions of the past, Transformer-based models allow for a more holistic interpretation of historical data, treating time-series not merely as a sequence of independent events but as a rich tapestry of interconnected signals that evolve across varying horizons.

The motivation for this research stems from the growing need to bridge the gap between pure technical performance and the broader socio-technical requirements of financial infrastructures. As AI models become more deeply embedded in the plumbing of global finance, their design must account for more than just a reduction in mean squared error. We must consider how these models interact with existing regulatory frameworks, how they influence market volatility through synchronized behavior, and how they consume energy resources in an increasingly carbon-conscious world. This paper situates Transformer-based forecasting within this wider context, arguing that a multi-horizon framework is not only a technical necessity for capturing different market regimes but also a strategic requirement for building robust financial systems. Through an interdisciplinary lens, we explore the structural trade-offs inherent in these models, emphasizing the tension between the "black box" nature of deep learning and the transparency required by institutional oversight.

Furthermore, the transition to multi-horizon forecasting marks a departure from point-in-time predictions toward a more comprehensive understanding of future trajectories. In financial contexts, knowing the price of an asset one minute from now is of limited utility compared to understanding its potential path over the next hour, day, or week. Transformers, through their ability to weight historical information based on relevance rather than mere proximity, are uniquely suited for this task. They can ignore the "noise" of transient volatility while focusing on the "signal" of underlying structural trends. This capability, however, introduces new challenges regarding data governance and model robustness, as the reliance on vast datasets makes these systems vulnerable to data poisoning, adversarial attacks, and the amplification of historical biases. This introduction sets the stage for a detailed examination of these dynamics, providing a foundation for a rigorous discussion of architecture, deployment, and policy.

2. Theoretical Foundations and the Evolution of Temporal Modeling

To understand the significance of Transformer-based models in finance, one must first trace the evolution of temporal modeling from its roots in linear statistics to the current state of deep learning. For decades, the Box-Jenkins methodology and the family of Autoregressive Integrated Moving Average models dominated the field. These models relied on the assumption of stationarity or near-stationarity, focusing on the relationship between a variable and its past values. While effective for stable, low-frequency data, they struggled with the heteroscedasticity and "fat tails" characteristic of financial returns. The subsequent introduction of Generalized Autoregressive Conditional Heteroscedasticity models addressed some of these limitations by accounting for volatility clustering, yet they remained constrained by their linear functional forms and inability to capture complex, non-linear interactions between disparate market features.

The advent of Recurrent Neural Networks and Long Short-Term Memory units marked the first significant wave of the "neural revolution" in finance. These architectures introduced the concept of "memory" through hidden states, allowing models to maintain information over time. However, these systems were inherently limited by their sequential nature. In a financial time-series, the most relevant information for a future prediction might have occurred thousands of steps in the past—perhaps during a similar market regime years prior. Recurrent models often struggled to preserve this long-range information due to the vanishing gradient problem, where the influence of distant data points fades as they are processed through successive layers. This sequential processing also created a computational bottleneck, as it prevented the parallelization of training tasks, making it difficult to scale models to the massive datasets generated by modern high-frequency trading and alternative data sources.

The Transformer architecture, introduced by Vaswani and colleagues, fundamentally reconsidered the role of sequence in data processing. By replacing recurrence with self-attention, the model treats every point in a time-series as being potentially related to every other point, regardless of their distance in time. This allows the model to learn a global representation of the data structure, identifying patterns that transcend simple linear decay. In a financial context, this means the model can simultaneously consider the immediate impact of a news release and the long-term cycle of interest rate changes. The use of multi-head attention further enhances this by allowing the system to attend to different types of information in parallel—for example, one "head" might focus on price momentum while another focuses on volume-based indicators. This theoretical shift from "remembering the past" to "attending to the relevant" is what empowers the multi-horizon framework discussed in this paper.

3. Architectural Trade-offs and System-Level Design

In designing a Transformer-based framework for multi-horizon forecasting, several critical architectural trade-offs must be addressed. The primary challenge lies in the balance between model capacity and computational efficiency. While larger models with more attention heads and deeper layers generally exhibit superior predictive performance, they also require significantly more memory and processing power. In the high-stakes world of finance, where latency can be the difference between profit and loss, the time it takes to generate a prediction is as important as the prediction itself. Therefore, a system-level design must optimize the "inference-latency" curve, potentially employing techniques like distillation or pruning to maintain accuracy while reducing the model's footprint for real-time deployment.

Another significant trade-off involves the method of positional encoding. Since Transformers do not inherently understand the order of data points, researchers must inject temporal information back into the model. In financial time-series, time is not just a sequence; it is a multi-layered construct involving daily, weekly, and seasonal cycles, as well as irregular intervals dictated by market holidays or trading hours. Choosing between fixed sinusoidal encodings and learnable embeddings involves a trade-off between model flexibility and the risk of overfitting. A robust framework must ensure that the encoding captures the hierarchical nature of time without introducing spurious correlations that might lead the model to "memorize" specific dates rather than learning generalizable temporal patterns.

Furthermore, the design of the loss function in a multi-horizon context presents a complex optimization problem. Unlike single-step prediction, where a model aims to minimize the error of a specific future value, multi-horizon forecasting requires the model to maintain accuracy across a range of future steps. This often leads to a conflict where improving the accuracy of long-term forecasts degrades the precision of short-term ones. A sophisticated system-level approach might involve dynamic weighting of the loss components, allowing the model to prioritize different horizons based on the current market state or the specific needs of the end-user. For instance, during periods of high volatility, the system might prioritize short-term risk mitigation, while in stable periods, it may focus on long-term trend following. This adaptability is crucial for the long-term sustainability and utility of the model in a dynamic financial ecosystem.

4. Deployment Infrastructure and Real-Time Integration

The transition from a laboratory-trained Transformer model to a production-ready financial forecasting system requires a robust and scalable infrastructure. Financial data is notorious for its "drift"—the statistical properties of the data change over time as market participants react to new information and to the models themselves. Consequently, a deployment framework must include continuous monitoring and automated retraining pipelines. This infrastructure, often referred to as MLOps, ensures that the model remains calibrated to current market conditions. The integration of high-performance computing clusters and low-latency data feeds is essential, as the self-attention mechanism, while parallelizable during training, still imposes a significant computational load during the forward pass when processing long historical windows.

Security and resilience are paramount in the deployment phase. Because Transformer models are sensitive to the quality of their input data, they are vulnerable to "data poisoning" attacks, where an adversary might purposefully trade in a way that creates misleading patterns in the price history. To mitigate this risk, the infrastructure must include outlier detection and data sanitization layers that filter out anomalous signals before they reach the model. Additionally, the system must be designed with "fail-safe" mechanisms. If the model encounters a market regime it has never seen before—a "black swan" event—the system should be able to detect its own uncertainty and hand off control to a human operator or a more conservative, rules-based algorithm. This hybrid approach ensures that the pursuit of high-alpha predictions does not lead to catastrophic systemic failure.

The geographical distribution of financial infrastructure also plays a role in model performance. To minimize "tick-to-trade" latency, many firms co-locate their servers within the same data centers as the exchange matching engines. Deploying large-scale Transformers in these environments requires specialized hardware, such as Tensor Processing Units or Field-Programmable Gate Arrays, which are optimized for the matrix multiplications at the heart of the attention mechanism. The logistical challenge of maintaining and updating this hardware across global financial hubs adds a layer of physical complexity to what is often perceived as a purely digital endeavor. Thus, the successful deployment of a multi-horizon forecasting framework is as much a feat of systems engineering as it is of mathematical modeling.

5. Robustness, Generalization, and the "Regime Shift" Problem

One of the most persistent challenges in financial forecasting is the problem of regime shifts—sudden changes in the underlying market dynamics caused by geopolitical events, policy changes, or technological shifts. A model trained on a decade of low-interest-rate data may perform poorly when central banks begin aggressive tightening. Transformers, while powerful, are not immune to this issue. Their ability to find complex patterns can actually be a disadvantage if those patterns are idiosyncratic to a specific historical period. Ensuring the robustness and generalization of a Transformer-based framework requires sophisticated regularization techniques and a focus on "out-of-distribution" performance.

To address this, researchers often employ ensemble methods or "mixture-of-experts" architectures, where different sub-models are trained on different market regimes. A gating mechanism then decides which expert to trust based on current market signals. Another approach involves the use of synthetic data generation through Generative Adversarial Networks or Variational Autoencoders to expose the Transformer to a wider variety of "stress test" scenarios than are available in the historical record. By training the model on simulated market crashes, hyper-inflationary periods, and liquidity crises, we can improve its ability to navigate extreme events when they occur in the real world.

The concept of "model fragility" also extends to the hyperparameters of the Transformer itself. Small changes in the dropout rate, the number of layers, or the learning rate schedule can lead to significantly different outcomes in financial applications. This sensitivity necessitates a rigorous cross-validation framework that goes beyond standard "rolling window" approaches. We must test the model's performance across diverse asset classes, geographies, and time periods to ensure that its predictive power is not an artifact of data mining. A truly robust multi-horizon framework is one that maintains a baseline level of performance even when its primary assumptions are challenged, prioritizing the preservation of capital over the maximization of theoretical gains.

6. Interpretability and the "Black Box" Dilemma in Financial Governance

As financial institutions and regulators increasingly rely on deep learning models, the demand for interpretability has become a central pillar of algorithmic governance. Transformer models, with their millions of parameters and complex attention maps, are frequently criticized as being "black boxes." This lack of transparency poses a significant risk: if a model makes a catastrophic trading decision, it is difficult for human supervisors to understand why it did so. In the context of the "Right to Explanation" under regulations like the GDPR and the evolving AI Act, the inability to explain a model's output could lead to legal and reputational consequences for financial firms.

Fortunately, the attention mechanism itself provides a unique window into the model's decision-making process. By visualizing attention weights, researchers can identify which historical events the model prioritized when making a specific prediction. For instance, if the model predicts a sharp decline in an equity index, the attention map might show it is "looking" at a spike in oil prices from two days ago and a certain pattern in the options market from that morning. This level of insight allows for a form of "post-hoc interpretability," where human experts can validate the model's logic against known economic theories. However, caution is required, as attention weights can sometimes be misleading or emphasize correlations that have no causal basis.

The drive for interpretability is also linked to the concept of "fairness" in finance. While less obvious than in credit scoring or hiring, algorithmic bias in trading can manifest as unequal access to liquidity or the systematic disadvantage of certain market participants. If a Transformer model learns to exploit patterns that are the result of market manipulation or structural inequities, it may inadvertently perpetuate those issues. Governance frameworks must therefore include "explainability audits," where models are periodically interrogated to ensure their predictions are grounded in legitimate market signals and do not violate ethical or regulatory standards. Balancing the raw predictive power of Transformers with the need for human-readable explanations remains one of the most critical challenges in the field.

7. Systemic Risk, Market Convergence, and Policy Implications

From a macro-prudential perspective, the widespread adoption of similar Transformer-based models across the financial industry introduces the risk of market convergence. If a large number of market participants use similar architectures trained on the same historical data, their models may reach the same conclusions simultaneously. This synchronized behavior can lead to "crowded trades" and "flash crashes," where a collective sell signal triggers a liquidity vacuum as everyone tries to exit the market at once. The very efficiency that makes these models attractive to individual firms can become a source of instability for the system as a whole.

Policymakers and regulators must therefore consider the systemic implications of deep learning in finance. This might involve the implementation of "circuit breakers" specifically designed for algorithmic trading, or requirements for "model diversity" among systemically important financial institutions. There is also a need for new reporting standards that allow regulators to monitor the "algorithmic health" of the market, identifying periods where high-correlation among models might signal an impending volatility spike. The goal is to create a regulatory environment that encourages innovation while protecting the integrity of the financial system from the unintended consequences of technological homogeneity.

Furthermore, the global nature of financial markets means that policy interventions must be coordinated internationally. A Transformer model operating in London may be reacting to signals from Tokyo while executing trades in New York. Disparate regulatory regimes could lead to "algorithmic arbitrage," where firms move their most aggressive or opaque models to jurisdictions with the weakest oversight. Establishing a global framework for the governance of AI in finance—one that addresses robustness, transparency, and systemic risk—is essential for maintaining stability in an increasingly interconnected and automated world. This section emphasizes that the "system" in "systems engineering" must include the legal and social structures that govern human and machine behavior alike.

8. Environmental Sustainability and Computational Ethics

The environmental impact of training large-scale Transformer models is an increasingly prominent concern in the academic community. The computational resources required to process high-dimensional financial data across multiple horizons are immense, translating into significant electricity consumption and carbon emissions. As the financial sector moves toward "Green Finance" and ESG (Environmental, Social, and Governance) goals, the carbon footprint of the models themselves cannot be ignored. A truly sustainable framework for financial forecasting must prioritize computational efficiency alongside predictive accuracy.

Technological solutions such as "Green AI" emphasize the development of models that require fewer floating-point operations to achieve the same results. This can include the use of more efficient attention mechanisms—such as Sparse Attention or Linear Transformers—which reduce the quadratic complexity of the standard self-attention operation. Additionally, firms can choose to locate their training facilities in regions with high proportions of renewable energy or use specialized hardware designed for energy efficiency. Ethical considerations also extend to the "compute divide," where only the largest and wealthiest institutions can afford the hardware and expertise required to build and maintain state-of-the-art Transformer models, potentially exacerbating the wealth gap between dominant firms and smaller market participants.

The ethical dimension also encompasses the responsibility of researchers and developers to consider the long-term impact of their work. While a more accurate forecasting model may increase a firm's profits, its contribution to the overall social good must be evaluated. If the model primarily serves to extract value from the market without contributing to price discovery or capital allocation, its social utility is questionable. By integrating sustainability and ethics into the core of the research process, we can ensure that the advancement of AI in finance serves to build a more resilient and equitable global economy, rather than merely a more profitable one.

9. Future Directions: Towards Adaptive and Self-Correcting Systems

Looking forward, the next frontier in Transformer-based financial modeling lies in the development of adaptive and self-correcting systems. Current frameworks, while powerful, are largely reactive—they learn from the past to predict the future. The future of the field involves models that can actively learn from their own mistakes in real-time, adjusting their internal representations as they receive feedback from the market. This move toward "online learning" or "continual learning" would allow Transformers to stay relevant in the face of rapid structural changes without the need for periodic, manual retraining.

Another promising direction is the integration of multi-modal data. While this paper focuses on time-series data, financial markets are influenced by a vast array of information, from textual data (news, earnings calls, social media) to visual data (satellite imagery of retail parking lots or oil tankers). A truly holistic multi-horizon framework would use Transformers to fuse these disparate data sources into a single, unified representation of the market state. The "Cross-Attention" mechanism, which allows a model to relate information from one domain (e.g., a news headline) to another (e.g., a price chart), will be instrumental in this evolution.

Finally, we anticipate a growing convergence between deep learning and traditional economic theory. Instead of viewing AI as a replacement for human intuition, we should see it as a tool that can augment and refine our understanding of economic systems. By constraining Transformer models with physical or economic "laws"—such as no-arbitrage conditions or supply-demand balance—we can create "Physics-Informed Neural Networks" for finance. These models would be more robust and interpretable, as their predictions would be grounded in both the data and the fundamental principles of the discipline. The journey toward this integrated future will require continued collaboration between computer scientists, engineers, economists, and policymakers.

10. Conclusion

The integration of Transformer-based deep learning into financial time-series forecasting represents a significant milestone in the evolution of socio-technical financial infrastructures. Through their ability to capture long-range dependencies and process information across multiple horizons, these models offer a powerful alternative to traditional econometric methods. However, as this paper has demonstrated, the transition to such sophisticated systems is fraught with architectural, operational, and ethical challenges. Success in this domain requires more than just high predictive accuracy; it demands a holistic commitment to robustness, interpretability, and systemic stability.

We have explored the structural trade-offs between computational complexity and performance, the critical role of deployment infrastructure, and the macro-level risks associated with model convergence and regime shifts. Furthermore, we have highlighted the necessity of aligning these powerful technologies with the broader goals of environmental sustainability and fair governance. As we move toward an era of increasingly automated and interconnected markets, the frameworks we build today will determine the resilience of the global financial system tomorrow. By fostering a culture of transparency and interdisciplinary collaboration, we can harness the potential of Transformer architectures to create a more efficient, stable, and equitable financial future.

References

1. Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. Applied Intelligence, 6(3), 205-213.

2. Arratia, A. (2014). Computational Finance: An Introductory Course with R. Atlantis Press.

3. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

4. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327.

5. Zhou, D. (2026). AI-Driven Hybrid SAST–DAST–SCA–IAST Framework for Risk-Based Vulnerability Prioritization in Microservice Architectures.

6. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

7. Brock, W. A., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. The Journal of Finance, 47(5), 1731-1764.

8. Brownlee, J. (2018). Deep Learning for Time Series Forecasting. Machine Learning Mastery.

9. Qi, R. (2025). AUBIQ: A generative AI-powered framework for automating business intelligence requirements in resource-constrained enterprises. *Frontiers in Business and Finance*, *2*(01), 66-86.

10. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

11. Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance, 1(2), 223-236.

12. Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2017). Language modeling with gated convolutional networks. International Conference on Machine Learning.

13. Zhang, T. (2025). A Knowledge Graph-Enhanced Multimodal AI Framework for Intelligent Tax Data Integration and Compliance Enhancement. *Frontiers in Business and Finance*, *2*(02), 247-261.

14. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

15. Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253-263.

16. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654-669.

17. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

18. Qi, R. (2025, August). Interpretable Slow-Moving Inventory Forecasting: A Hybrid Neural Network Approach with Interactive Visualization. In *Proceedings of the 2025 International Conference on Generative Artificial Intelligence for Business* (pp. 41-46).

19. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

20. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

21. Yi, X. (2025, October). Real-Time Fair-Exposure Ad Allocation for SMBs and Underserved Creators via Contextual Bandits-with-Knapsacks. In *Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science* (pp. 1602-1607).

22. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.

23. Kim, S. (2017). Financial series prediction using attention-based LSTM. arXiv preprint arXiv:1701.01887.

24. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

25. Liu, T. (2022, December). Financial Constraint'Impact on Firms' ESG Rating Based on Chinese Stock Market. In *2022 4th International Conference on Economic Management and Cultural Industry (ICEMCI 2022)* (pp. 1085-1095). Atlantis Press.

26. Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A, 379(2194), 20200209.

27. Lopez de Prado, M. (2018). Advances in Financial Machine Learning. John Wiley & Sons.

28. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. International Journal of Forecasting, 34(4), 596-608.

29. Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020, February). InP grating coupler design for vertical coupling of InP and silicon chips. In *Integrated Optics: Devices, Materials, and Technologies XXIV* (Vol. 11283, pp. 33-38). SPIE.

30. Paszke, A., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems.

31. Rangapuram, S. S., et al. (2018). Deep state space models for time series forecasting. Advances in Neural Information Processing Systems.

32. Liu, T. (2026). Volatility Forecasting and Early-Warning Market Stress Detection: A Leakage-Safe Evaluation with Tree Ensembles and Transformers.

33. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 36(3), 1181-1191.

34. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. Communications of the ACM, 63(12), 54-63.

35. Taylor, S. J. (2011). Asset Price Dynamics, Volatility, and Prediction. Princeton University Press.

36. Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

37. Wen, R., et al. (2017). A multi-horizon quantile recurrent forecasting network. arXiv preprint arXiv:1711.11053.

38. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.

39. Yi, X. (2026). A Federated and Differentially Private Incentive–Marketing Framework for Privacy-Preserving Cross-Channel Measurement in AI-Powered Digital Commerce.

40. Zhou, H., et al. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. The Thirty-Fifth AAAI Conference on Artificial Intelligence.

41. Zoran, D., et al. (2020). Towards learning interpretable representations of self-attention. arXiv preprint arXiv:2004.14300.