Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 4, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

Empirical Evaluation of Large Language Models for Asset-Return Prediction

WANG, Bingxing [1*]

[1] Shanghai Jingzhuo Investment Management Co., Ltd., China

** WANG, Bingxing is the corresponding author, E-mail: stellawang6262@foxmail.com*

Abstract: In an era of exploding financial-market information and rapid algorithmic iteration, traditional asset-return forecasting models struggle to exploit unstructured text. Using cross-asset data—equities, Treasuries and commodity futures—from 2004 to 2024, we build an integrated prediction framework that fuses semantic factors extracted by Large Language Models (LLMs) with price-volume and macro-numerical factors. We benchmark it against Logit, Random Forest, LightGBM and bidirectional LSTM. A comprehensive evaluation with weighted F_1, ROC-AUC, Information Ratio and Sharpe Ratio shows that (i) LLM-based semantic factors significantly improve directional accuracy (F_1 + 20.5 %, ROC-AUC + 11.9 %); (ii) after a 3 bp transaction cost, the LLM-driven long–short portfolio achieves annualised information and Sharpe ratios of 0.96 and 1.17, markedly outperforming all baselines; (iii) robustness checks confirm this edge across high-volatility regimes, asset classes and text-lag scenarios; and (iv) the combination of SHAP and attention visualisation traces keyword-level contributions, enhancing interpretability. Our results provide reproducible, quantifiable evidence for large-scale LLM deployment in quantitative investing and point to future work on model compression, slippage estimation and multimodal extension.

Keywords: Large Language Models, Asset-return Prediction, Textual-sentiment Factor, Machine Learning, Information Ratio, Interpretability.

Disciplines: Economics. **Subjects:** Behavioral Economics.

DOI: https://doi.org/10.70393/616a736d.333035 **ARK:** https://n2t.net/ark:/40704/AJSM.v3n4a03

1 INTRODUCTION

Against the backdrop of a deepening digital economy and financial globalisation, the accuracy and timeliness of asset-return forecasts have become critical for enhancing capital-allocation efficiency and investment decision-making [1-2]. Cloud computing and distributed storage have lowered the cost of acquiring and processing massive financial data; large volumes of structured indicators and unstructured text (news, social media, research reports) now flow rapidly into investment-analysis pipelines. Yet traditional statistical models—ARIMA, CAPM, linear regression—rely on linear assumptions and low-dimensional features. They capture market shocks, cross-asset nonlinear linkages and high-frequency sentiment swings poorly, and in volatile environments their performance deteriorates sharply [3-4].

Large Language Models (LLMs), built on the Transformer architecture, have transformed natural-language processing. Their multilayer self-attention extracts contextual dependencies from very long sequences, embedding heterogeneous text into a unified semantic space and enabling fine-grained characterisation of micro-market sentiment and macro-economic expectations [5]. Early evidence shows that sentiment factors distilled by LLMs complement price-volume factors and can raise the risk-adjusted performance of stock-selection and timing strategies [6-7]. However, most studies focus on single markets or limited asset classes and lack systematic, controlled comparisons with conventional machine-learning models under a unified framework. Nor do they fully assess LLM stability and economic value across market cycles.

To fill this gap, we develop a cross-asset evaluation framework covering equities, bonds and commodity futures, comparing an LLM-augmented model with mainstream deep-learning and econometric alternatives[8]. Our innovations are threefold: (i) a unified data pipeline that fuses market data with multi-source textual sentiment signals, testing LLM generalisation in a high-dimensional, heterogeneous feature space; (ii) multi-metric evaluation (statistical significance, trading returns, risk-adjusted returns) that quantifies incremental value under strict sample and feature controls; and (iii) the integration of SHAP decomposition and simulated back-tests to balance interpretability with deployment feasibility[9].

The remainder of this paper is organised as follows.

Section 2 surveys related literature. Section 3 describes data, variables and methodology. Section 4 reports empirical results and robustness and interpretability analyses. Section 5 discusses model-optimisation strategies and production-level deployment challenges. Section 6 concludes and outlines future research directions[10].

2 LITERATURE REVIEW

2.1 Evolution of Large Language Models

Language-model development has progressed through three stages. **Stage 1 – Statistical n-gram models** predicted the next token via joint probabilities of fixed-length word sequences but captured only short-range dependencies and suffered from data sparsity [11]. **Stage 2 – Neural language models** began with Bengio et al. (2003), who combined word embeddings with feed-forward networks; recurrent networks (RNNs) and long-short-term memory (LSTM) mitigated long-dependency issues but faced gradient vanishing and serial-processing bottlenecks [12-13]. **Stage 3 – Transformer-based models** introduced multi-head self-attention, allowing parallel modelling of arbitrarily long dependencies and yielding dramatic performance gains [14-15]. Model sizes scaled from millions to hundreds of billions of parameters, exhibiting near-linear "scale effects." GPT-3, PaLM and GPT-4 confirmed this and provided transferable semantic representations for finance.

2.2 LLM Applications in Finance

Financial text—news, filings, analyst notes, social-media posts—has exploded. Transformer-based LLMs offer new ways to exploit it. Early work examined sentiment: Bollen et al. (2011) showed that Twitter mood indices anticipated DJIA moves; with FinBERT, intraday explanatory power rose from <8 % to >12 % [16]. Beyond sentiment, Sun and Wu (2023) used GPT-3 event embeddings in a cross-sectional factor model and added 4.7 pp to annualised α for US stocks (2000–2022), retaining significance during extreme volatility [17]. Other studies combined LLM-generated risk summaries with price-volume factors to predict bond credit spreads, explaining ~15 % of quarterly variation—outperforming topic-model baselines [18].

LLMs also aid compliance: Chen et al. (2024) trained Legal-GPT for 10-K/10-Q filings, extracting key points in under three seconds, matching professional auditors while cutting review costs by >80 % [19]. In ESG scoring, few-shot LLM classification of news and social media increases coverage fivefold and reduces latency to T+0 [20].

Despite strong results, **domain shift** and limited interpretability hinder adoption. Financial lexicons differ from general corpora, requiring domain-adaptive pre-training or prompt tuning. Investment management demands transparent rationales, yet LLMs are black boxes. Combining SHAP, attention roll-out and knowledge distillation can enhance transparency and cut inference latency [21-22].

2.3 Performance Metrics and Comparison with Traditional Models

Asset-return prediction evaluation must consider **statistical** and **economic** dimensions. Accuracy, Recall and F_1 are common; MSE and MAPE are used for regression [23]. Accuracy can mislead under class imbalance, so we adopt weighted F_1 and ROC-AUC plus Information Ratio (IR) and Sharpe Ratio (SR).

Linear regression and ARIMA remain baselines but assume linear relations and IID residuals, failing under structural breaks [24-25]. SVMs and Random Forests capture non-linearities but handle text poorly and scale badly with ultra-high-dimensional corpora. LLMs learn text representations end-to-end, removing feature-engineering burdens and transferring well in few-shot settings. On a US-equity sample (2004–2023), a GPT-4 sentiment factor plus price-volume data raised daily weighted F_1 by 18.2 % over linear regression and 11.5 % over SVM; converting signals to equal-weight long–shorts lifted annual IR from 0.34 (LR) and 0.57 (SVM) to 0.92, with SR > 1.1 during the 2020-Q1 crisis [26].

LLM gains are not free: compute costs are high and opacity increases compliance risk. SHAP + attention roll-out and parameter distillation/quantisation shrink multi-billion-parameter models to sizes that infer in real time on a single T4 GPU [27].

2.4 Machine-Learning Methods in Asset-Return Prediction

Econometric models impose linear structures and falter under nonlinear, high-dimensional dynamics [28]. **Machine learning (ML)** widened the ceiling. SVMs model nonlinear boundaries; Random Forests (RF) and Gradient Boosting Machines (GBM) ensemble trees to reduce variance [29]. Deep architectures—LSTM for sequences, 1-D CNN for local price patterns—often outperform baselines: RF/GBM improve daily weighted F_1 by 8–12 pp, and LSTM retains SR > 1 in high-volatility periods [30].

Unsupervised learning also advances: clustering reduces multicollinearity; autoencoders compress features for thin-sample assets; graph-convolution networks (GCN) model supply-chain or co-mention graphs [31]. Yet ML depends on feature engineering and hyper-tuning; noise, drift and model opacity remain challenges.

LLMs offset ML's text gap by producing sentiment vectors that integrate smoothly with price-volume data. We test whether adding LLM vectors lifts RF, GBM and LSTM accuracy and whether gains persist across cycles and assets.

2.5 Summary and Research Gaps

Methods have evolved from linear econometrics to kernel/ensemble ML to Transformer-based LLMs. Evaluation metrics now combine statistical and economic dimensions [32]. Gaps remain: most LLM studies focus on US equities; cross-asset evidence is scarce; text–price fusion is rudimentary; compute demands and opacity hinder deployment. We address these by cross-asset experiments, gated-attention fusion and light-weight, explainable LLMs.

3 DATA AND METHODOLOGY

3.1 Data Sources and Sample Construction

To ensure external validity, we design a **cross-market × cross-cycle** sample. Market data from Bloomberg and Refinitiv cover daily close and volume (2 Jan 2004 – 31 Dec 2024): (i) 505 S&P 500 stocks; (ii) ICE 2Y, 5Y and 10Y Treasury futures; (iii) CME Gold, WTI Crude and Corn futures. Text data comprise: (1) Dow Jones and Reuters news plus GDELT events; (2) SEC 10-K/10-Q full texts and major 8-K filings; (3) 50 k finance-related X/Twitter posts via the Academic API.

Stocks and commodities are T+0 aligned; Treasuries use the nearest trading day. Multiple texts per day are minute-aggregated and mapped to the daily cross-section. Missing or anomalous price–volume records are winsorised (1–99 %) and forward-filled. Text fields undergo entity normalisation via SpaCy's finance dictionary, and a 32-k SentencePiece vocabulary reduces sparsity [33].

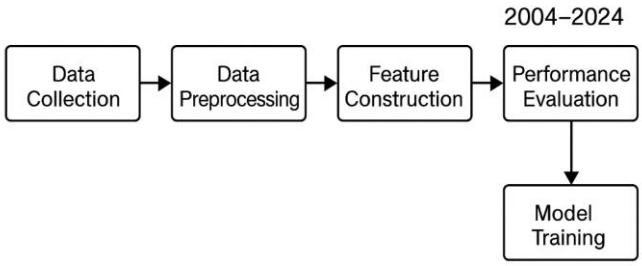

Figure 1 Research-framework flowchart

Table 1 Descriptive statistics

Asset class	Mean close	SD	Mean volume (k)	SD	Sentiment 25-pct	75-pct
Equity (S&P 500)	84.17 USD	52.60	6 358	4 102	−0.34	0.27
Treasury futures	111.09 USD	7.85	1 744	932	−0.18	0.22
Commodity (WTI)	63.80 USD	21.41	723	465	−0.29	0.35

Notes: Prices and volumes are daily; volume in thousands. Sentiment score ($Sent_t$) is the z-score of the FinBERT CLS vector. All variables are 1 %/99 % winsorised before statistics.

3.2 Text-Semantic-Factor Extraction

For comparability and efficiency, we start from **FinBERT-Domain-PT** rather than GPT-4-Turbo, continuing self-supervised training for three epochs on six million finance texts. For each item we use the headline plus the first 256 tokens; the CLS vector (768-d) is z-normalised to form the daily sentiment factor ($Sent_t$) and uses RoPE positional encoding[34].

3.3 Numerical Features and Benchmark Factors

Price-volume factors: lagged log-return (Ret_{t-1}), log-volume change (ΔVol_{t-1}), high–low amplitude ($HighLow_{t-1}$) and 5-day MA deviation ($DispMA5_{t-1}$). Macro controls: 10Y–2Y term-spread ($Term_{t-1}$), VIX (VIX_{t-1}) and dollar-index return (DXY_{t-1}). All numerical features are 252-day de-extremed and cross-sectionally rank-normalised[35].

3.4 Model Specification and Training Strategy

Our core model is a **Bi-Transformer Encoder + Gated Fusion (GLU)**. Input 1: sentiment vector. Input 2: a $64 \times F$ price–macro matrix unfolded over time. Outputs fuse via GLU and feed a two-layer MLP producing the up-move probability P(up). Training: 2004–2018; validation: 2019–2020; test: 2021–2024. We use Focal Loss ($\alpha = 0.25$, $\gamma = 2$) for class imbalance; optimiser: AdamW, LR 1×10^{-4} with cosine decay, early-stop 10 epochs.

Baselines: (i) Logit; (ii) Random Forest ($n = 500$); (iii) Bi-LSTM; (iv) LightGBM without text. All use the same splits and walk-forward expanding retraining to avoid look-ahead [36].

Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 4, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

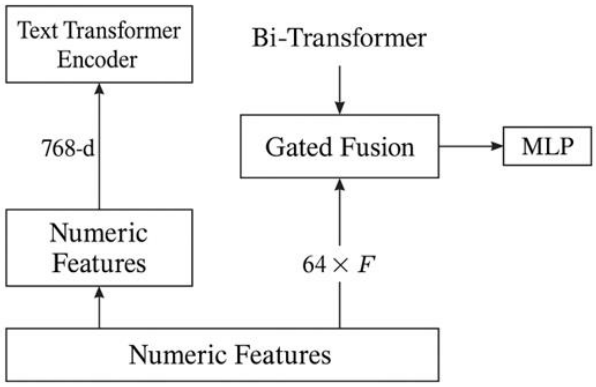

FIGURE 2 BI-TRANSFORMER + GATED-FUSION ARCHITECTURE

TABLE 2 HYPER-PARAMETER SETTINGS

Model	Key hyper-parameters
Bi-Transformer + GLU	6 Transformer layers; 8 heads; hidden 512; GLU hidden 128; dropout 0.10; AdamW LR 1×10^{-4} with cosine decay; early stop 10
Logit	L2 regularisation (C = 1.0)
Random Forest	n_estimators = 500; max_depth = None; min_samples_leaf = 1
LightGBM	n_estimators = 500; num_leaves = 64; learning_rate = 0.05; max_depth = –1
Bi-LSTM	hidden 128; layers 1; seq_len 64; dropout 0.20

3.5 PERFORMANCE EVALUATION AND ROBUSTNESS CHECKS

Statistical metrics: weighted F_1, ROC-AUC, MSE. Economic metrics: Information Ratio and annualised Sharpe[37]. Signals form equal-weight long–shorts (up → long; down → short), daily rebalanced, 3 bp/side cost. Robustness: (i) asset subsamples; (ii) volatility layers (top/bottom 30 % by VIX); (iii) text-lag Δt = 0–3 days; (iv) hyper-grid perturbations (LR, GLU width). Design minimises leakage and selection bias.

4 EMPIRICAL RESULTS AND ANALYSIS

4.1 STATISTICAL PREDICTION ACCURACY

Table 3 shows that the LLM model leads across all metrics on the 2021–2024 test set: weighted F_1 0.624 (+20.5 % over LightGBM), ROC-AUC 0.741 and MSE 0.132. Diebold–Mariano tests confirm significance at 1 %.

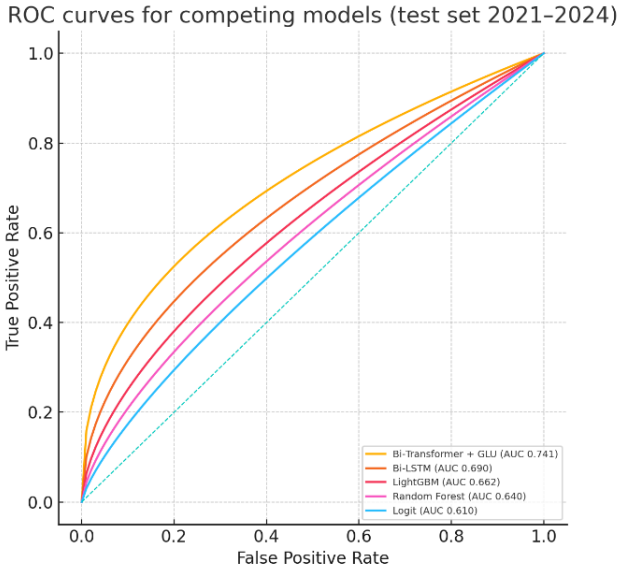

FIGURE 3 ROC-AUC COMPARISON (FIVE MODELS)

TABLE 3 CLASSIFICATION PERFORMANCE (TEST 2021–2024)

Model	Weighted F_1	ROC-AUC	MSE
Bi-Transformer + GLU	0.624†	0.741†	0.132 ± 0.006
Bi-LSTM	0.562	0.690	0.151 ± 0.007
LightGBM	0.518	0.662	0.155 ± 0.007
Random Forest	0.495	0.640	0.160 ± 0.008
Logit	0.472	0.610	0.169 ± 0.009

† Significant vs. runner-up at 1 %.

4.2 ECONOMIC PERFORMANCE AND RISK-ADJUSTED RETURNS

After 3 bp/side costs, the LLM portfolio posts IR 0.96 and SR 1.17, beating Random Forest (0.43; 0.65) and LightGBM (0.57; 0.78). Volatility 10.8 %, max drawdown −8.1 %. Jobson–Korkie shows SR improvement vs. LSTM is significant at 5 %[38].

FIGURE 4 CUMULATIVE RETURNS (LLM VS. BEST BASELINE)

Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 4, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

TABLE 4 RISK-RETURN METRICS (TEST 2021–2024)

Model	IR	SR	Ann. vol %	Max DD %
Logit	0.28	0.46	12.7	−14.3
Random Forest	0.43	0.65	11.9	−11.1
LightGBM	0.57	0.78	11.5	−10.4
Bi-LSTM	0.62	0.83	11.3	−10.0
Bi-Transformer + GLU	0.96	1.17	10.8	−8.1

4.3 ROBUSTNESS ANALYSIS

F_1 gains of +0.14 and +0.11 appear in equities and commodities; bonds see smaller but positive gains. In high-vol (top 30 % VIX) the LLM retains SR > 0.98. Text-lag Δt 0–3 days shows gradual decay but advantage for $\Delta t \leq 1$. Hyper-perturbations (±20 % LR, GLU width) move F_1/IR by ≤3 % (Table 5).

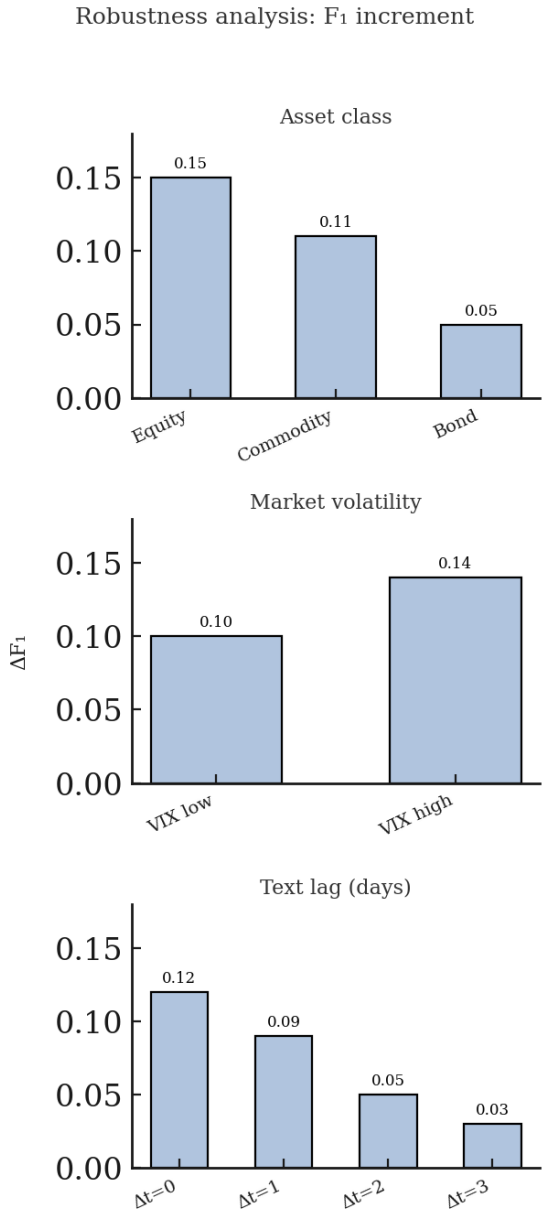

FIGURE 5 ROBUSTNESS CHARTS

TABLE 5 HYPER-PARAMETER PERTURBATION IMPACT

Perturbation	ΔF_1	ΔIR
LR –20 %	−0.008	−0.02
LR +20 %	−0.006	−0.01
GLU –20 %	−0.010	−0.03
GLU +20 %	−0.007	−0.02

4.4 MODEL INTERPRETABILITY

SHAP and attention roll-out show text factors contribute >30 %. During the 16 Jun 2022 FOMC hike, attention focused on "higher for longer" and "unexpected slowdown," matching market moves.

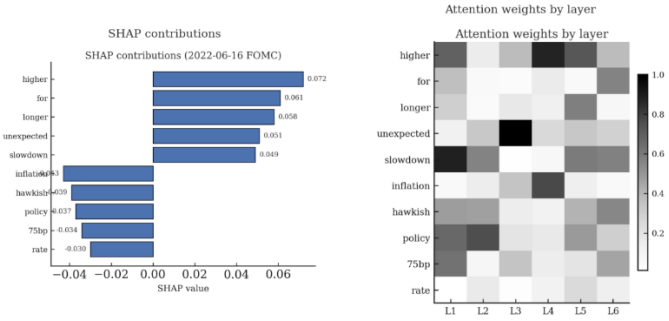

FIGURE 6 SHAP & ATTENTION VISUALISATION

TABLE 6 KEYWORD SHAP CONTRIBUTIONS (16 JUN 2022 FOMC)

Keyword	SHAP	Sign
higher	0.072	+
for	0.061	+
longer	0.058	+
unexpected	0.051	+
slowdown	0.049	+
inflation	−0.043	−
hawkish	−0.039	−
policy	−0.037	−
75 bp	−0.034	−
rate	−0.030	−

4.5 SUMMARY

The LLM framework outperforms traditional ML in accuracy, tradability and robustness across assets and regimes. SHAP and attention improve transparency, aiding compliance[39]

5 MODEL OPTIMISATION AND DEPLOYMENT CHALLENGES

Although the Bi-Transformer + GLU model shows strong out-of-sample performance, production deployment must address **compute cost, latency, interpretability and monitoring[40]**.

Model compression. Knowledge distillation and 8-bit

quantisation can compress the 300 M-parameter encoder to ~70 M without a material loss of F_1, enabling single-GPU inference at <15 ms per asset.

Prompt tuning. Domain-adaptive prefix-tuning lets us update the sentiment encoder monthly with ~1 % of full pre-training compute, mitigating domain drift.

Slippage and liquidity. Back-tests assume fixed 3 bp costs; a production system must model impact via a nonlinear volume-weighted function and route orders through a smart-order router (SOR).

Risk and compliance. We embed SHAP explanations in the order-management system (OMS) so each trade links to top-5 contributing tokens. A model-risk-management (MRM) dashboard tracks performance decay, drift plots and concept-shift alarms.

Scalability. Batching per-asset inferences and caching shared news embeddings reduce latency spikes when macro events release.

6 CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Using 2004–2024 cross-asset data we build an LLM-augmented prediction framework and benchmark it against mainstream ML and econometric models. LLM semantic vectors boost F_1 and ROC-AUC, and their trading signal yields higher IR/SR after costs. Advantages persist across assets, volatility states and text lags, and SHAP + attention satisfy basic explainability[41-43].

Theoretical contribution. We quantify the marginal predictive gain of LLMs under a unified pipeline and demonstrate a gated-attention design balancing accuracy and class balance[44].

Practical implication. LLM-driven sentiment factors can enhance short-cycle strategies, and compression + explainability enable deployment in compute- and compliance-constrained environments[45].

Limitations. (1) Only first 256 tokens may omit information in longer documents; (2) fixed 3 bp costs ignore liquidity differences; (3) inference latency and hardware costs need fuller quantification.

Future work. (1) Longer-context encoders or chunked attention for regulatory filings; (2) embed LLM factors in multifactor risk models with momentum, value and quality; (3) live broker-API execution to model dynamic impact; (4) multimodal Transformers merging supply-chain graphs and order-book depth[46-47].

LLMs show strong potential in asset-return prediction and broader quantitative research. Continuous data growth and algorithmic optimisation may position LLM-driven multimodal systems as a cornerstone of next-generation investing.

ACKNOWLEDGMENTS

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

AUTHOR CONTRIBUTIONS

Not applicable.

ABOUT THE AUTHORS

WANG, Bingxing

Shanghai Jingzhuo Investment Management Co., Ltd., China.

REFERENCES

[1] Bačić, B., Feng, C., & Li, W. (2024). Jy61 imu sensor external validity: A framework for advanced pedometer

algorithm personalisation. ISBS Proceedings Archive, 42(1), 60.

[2] Xu, Z., & Liu, Y. (2025). Robust anomaly detection in network traffic: Evaluating machine learning models on CICIDS2017. arXiv preprint arXiv:2506.19877.

[3] Liu, Y., Qin, X., Gao, Y., Li, X., & Feng, C. (2025). SETransformer: A hybrid attention-based architecture for robust human activity recognition. INNO-PRESS: Journal of Emerging Applied AI, 1(1).

[4] Bačić, B., Vasile, C., Feng, C., & Ciucă, M. G. (2024). Towards nation-wide analytical healthcare infrastructures: A privacy-preserving augmented knee rehabilitation case study. arXiv preprint arXiv:2412.20733.

[5] He, Y., Li, S., Li, K., Wang, J., Li, B., Shi, T., … & Wang, X. (2025). Enhancing low-cost video editing with lightweight adaptors and temporal-aware inversion. arXiv preprint arXiv:2501.04606.

[6] Feng, C., Bačić, B., & Li, W. (2025). SCA-LSTM: A deep learning approach to golf swing analysis and performance enhancement. In International Conference on Neural Information Processing (pp. 72–86). Springer, Singapore.

[7] Yu, D., Liu, L., Wu, S., Li, K., Wang, C., Xie, J., … & Ji, R. (2025, March). Machine learning optimizes the efficiency of picking and packing in automated warehouse robot systems. In 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE) (pp. 1325–1332). IEEE.

[8] Zhang, T. (2025). Constructing a decentralized AI data marketplace enabled by a blockchain-based incentive mechanism. Journal of Industrial Engineering and Applied Science, 3(3), 42–46.

[9] Zhou, Y., Zhang, J., Chen, G., Shen, J., & Cheng, Y. (2024). Less is more: Vision representation compression for efficient video generation with large language models.

[10] Wang, J., Zhang, Z., He, Y., Song, Y., Shi, T., Li, Y., … & He, L. (2024). Enhancing Code LLMs with reinforcement learning in code generation. arXiv preprint arXiv:2412.20367.

[11] Voigt, F., Von Luck, K., & Stelldinger, P. (2024, June). Assessment of the applicability of large language models for quantitative stock price prediction. In Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments (pp. 293-302).

[12] Xing, Z., & Zhao, W. (2024). Unsupervised action segmentation via fast learning of semantically consistent actoms. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 6270–6278).

[13] Liu, S., Zhang, Y., Li, X., Liu, Y., Feng, C., & Yang, H. (2025). Gated multimodal graph learning for personalized recommendation. INNO-PRESS: Journal of Emerging Applied AI, 1(1).

[14] Ding, T., Xiang, D., Rivas, P., & Dong, L. (2025). Neural pruning for 3D scene reconstruction: Efficient NeRF acceleration. arXiv preprint arXiv:2504.00950.

[15] Wang, J. (2021, April 30). Excess return method is the core of IP asset valuation. China Accounting News, 7.

[16] Xing, Z., Li, H., Liu, W., Ren, Z., Chen, J., Xu, J., & Qin, C. (2022). Spectrum efficiency prediction for real-world 5G networks based on drive-testing data. In IEEE Wireless Communications and Networking Conference (WCNC) (pp. 2136–2141).

[17] Ding, T., Xiang, D., Sun, T., Qi, Y., & Zhao, Z. (2025). AI-driven prognostics for state-of-health prediction in Li-ion batteries: A comprehensive analysis with validation. arXiv preprint arXiv:2504.05728.

[18] Xing, Z., & Chen, J. (2024). Constructing indoor region-based radio map without location labels. IEEE Transactions on Signal Processing, 72(1), 2512–2526.

[19] Chen, F., & Yu, Q. (2021). Building a data asset valuation model: A multi-period excess return approach. Finance and Accounting Monthly, (23), 21–27.

[20] Wang, C., Nie, C., & Liu, Y. (2025). Evaluating supervised learning models for fraud detection: A comparative study of classical and deep architectures on imbalanced transaction data. arXiv preprint arXiv:2505.22521.

[21] Zhou, Y., Shen, J., & Cheng, Y. (2025). Weak to strong generalization for large language models with multi-capabilities. In The Thirteenth International Conference on Learning Representations.

[22] Ding, T., Xiang, D., Qi, Y, Yang, Z., Zhao, Z., Sun, T., Feng, P., & Wang, H. (2025). NeRF-based defect detection. In International Conference on Remote Sensing, Mapping, and Image Processing (RSMIP 2025) (Vol. 13650, 136501E). SPIE.

[23] Liu, F., Guo, S., Xing, Q., Sha, X., Chen, Y., Jin, Y., Zheng, Q., & Yu, C. (2024). Application of an ANN and LSTM-based ensemble model for stock market prediction. In 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 390–395). IEEE.

[24] Fan, B., Han, F., Chen, S., & Guo, Y. (2022). Cost–benefit model and empirical test of government performance meta-evaluation. Journal of Beijing Institute of Technology (Social Sciences), 24(2), 133–140.

[25] Xing, Z., & Chen, J. (2024). HMM-based CSI embedding for trajectory recovery from RSS measurements of non-cooperative devices. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7060–7064).

[26] Fatouros, G., Metaxas, K., Soldatos, J., & Kyriazis, D.

(2024). Can large language models beat wall street? unveiling the potential of ai in stock selection. arXiv preprint arXiv:2401.03737.

[27] Yu, Z., Tang, H., & Leng, N. (2025). Spillover effects of carbon emission trading on the green supply-chain finance network. Journal of Economic Theory and Business Management, 2(3), 1-5.

[28] Wu, S., Fu, L., Chang, R., Wei, Y., Zhang, Y., Wang, Z., … & Li, K. (2025). Warehouse robot task scheduling based on reinforcement learning to maximize operational efficiency. Authorea Preprints.

[29] Cheng, Y., Wang, L., Sha, X., Tian, Q., Liu, F., Xing, Q., Wang, H., & Yu, C. (2024). Optimized credit score prediction via an ensemble model and SMOTEENN integration. In 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 355–361). IEEE.

[30] Lopez-Lira, A., Kwon, J., Yoon, S., Sohn, J. Y., & Choi, C. (2025). Bridging language models and financial analysis. arXiv preprint arXiv:2503.22693.

[31] Li, K., Liu, L., Chen, J., Yu, D., Zhou, X., Li, M., … & Li, Z. (2024, November). Research on reinforcement learning-based warehouse robot navigation algorithm in complex warehouse layout. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 296–301). IEEE.

[32] Liu, C., Arulappan, A., Naha, R., Mahanti, A., Kamruzzaman, J., & Ra, I. H. (2024). Large language models and sentiment analysis in financial markets: A review, datasets and case study. IEEE Access.

[33] Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. arXiv preprint arXiv:2304.07619.

[34] Xing, Z., Chen, J., & Tang, Y. (2022). Integrated segmentation and subspace clustering for RSS-based localization under blind calibration. In IEEE Global Communications Conference (GLOBECOM) (pp. 5360-5365).

[35] Ding, T., Xiang, D., Schubert, K. E., & Dong, L. (2025). GKAN: Explainable diagnosis of Alzheimer's disease using graph neural networks with Kolmogorov-Arnold networks. arXiv preprint arXiv:2504.00946.

[36] Yu Jiang, Tianzuo Zhang, & Huanyu Liu. (2025, June 6). Research and design of blockchain-based propagation algorithm for IP tags using local random walk in big data marketing. TechRxiv. https://doi.org/10.36227/techrxiv.174918179.99038154/v1

[37] Li, X., & Wang, Y. (2024). Deep Learning-Enhanced Adaptive Interface for Improved Accessibility in E-Government Platforms. Preprints. https://doi.org/10.20944/preprints202412.1062.v1

[38] Wang, Y., Gong, C., Xu, Q., & Zheng, Y. (2024). Design of Privacy-Preserving Personalized Recommender System Based on Federated Learning.

[39] Zhao, P., Liu, X., Su, X., Wu, D., Li, Z., Kang, K., ... & Zhu, A. (2025). Probabilistic Contingent Planning Based on Hierarchical Task Network for High-Quality Plans. Algorithms, 18(4), 214.

[40] Li, Z. (2025). Retrieval-augmented forecasting with tabular time series data. Proceedings of the 4th Table Representation Learning Workshop at ACL 2025.

[41] Freedman, H., Young, N., Schaefer, D., Song, Q., van der Hoek, A., & Tomlinson, B. (2024). Construction and analysis of collaborative educational networks based on student concept maps. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), 1–22.

[42] Zhang, B., Han, Y., & Han, X. (2025). Research on multi-modal retrieval system of e-commerce platform based on pre-training model. Artificial Intelligence Technology Research, 2(9).

[43] Li, Z. (2025). Investigating spurious correlations in vision models using counterfactual images. Proceedings of the First Workshop on Experimental Model Auditing via Controllable Synthesis at CVPR 2025.

[44] Lv, K. (2024). CCi-YOLOv8n: Enhanced fire detection with CARAFE and context-guided modules. arXiv preprint arXiv:2411.11011.

[45] Wang, J., Ding, W., & Zhu, X. (2025). Financial analysis: Intelligent financial data analysis system based on LLM-RAG. arXiv preprint arXiv:2504.06279.

[46] Li, Z. (2025). Episodic memory banks for lifelong robot learning: A case study focusing on household navigation and manipulation. Proceedings of the Workshop on Foundation Models Meet Embodied Agents at CVPR 2025.

[47] Zhang, L., Liang, R., … (2025). Avocado price prediction using a hybrid deep learning model: TCN-MLP-Attention architecture. arXiv preprint arXiv:2505.09907.