IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

A Proactive, Sentiment-Augmented Reinforcement Learning Framework for Algorithmic Trading in the XAUUSD Market

Prachi Shinde, Pranav Kharad, Shruti Jamdar, and Hitansh Oswal
MITADT University, Loni Kalbhor, Pune, India
prachi.shinde@mituniversity.edu.in, pranav.kharad@mituniversity.edu.in,
shruti.jamdar@mituniversity.edu.in, hitansh.oswal@mituniversity.edu.in

Abstract: *This study proposes a proactive hybrid artificial- intelligence framework for trading the XAUUSD pair by combin- ing reinforcement learning, predictive modeling, and sentiment analysis. Traditional trading systems typically focus either on price prediction or on reactive decision rules, limiting adapt- ability in complex markets. The proposed architecture integrates three complementary components: (1) a reinforcement-learning engine based on the Advantage Actor–Critic (A2C) algorithm that optimizes trading actions; (2) a Long Short-Term Memory (LSTM) network that forecasts near-term price dynamics to enhance foresight; and (3) a FinBERT-based sentiment module that evaluates real-time market sentiment extracted from finan- cial news. The framework employs a Differential Sharpe-Ratio reward to balance profitability with risk management, while the entire trading process is modeled as a Markov Decision Process encompassing both historical and contextual data. Extensive rolling-window experiments on high-frequency XAUUSD data show that the integrated model achieves superior risk-adjusted returns, delivering a higher Sharpe Ratio and smaller drawdowns compared with standard reinforcement or predictive agents. The findings highlight the importance of combining predictive, behavioral, and sentiment-driven intelligence to navigate the volatility of precious-metal markets.*

Keywords: Algorithmic Trading, Reinforcement Learning, Sentiment Analysis, XAUUSD, Deep Learning, LSTM, FinBERT, Financial Markets

I. INTRODUCTION

The gold–U.S. dollar pair (XAUUSD) occupies a distinct position within global financial markets. Unlike equities, which represent company ownership, or fiat currencies reg- ulated by central banks, gold derives its value from a complex interaction of scarcity, investor behavior, and macroeconomic stability. Historically, gold has functioned as a hedge during inflationary episodes and geopolitical uncertainty, preserving real value when paper assets depreciate. Because its supply changes slowly and is largely independent of short-term pro- duction fluctuations, gold's price dynamics differ substantially from those of many other traded instruments.

Price formation in XAUUSD reflects an interplay of supply- and-demand, central-bank action, inflationary expectations, and investor sentiment. Models that rely exclusively on his- torical prices often neglect behavioral and macroeconomic drivers; as a result, they can miss early-warning signals or misinterpret transient noise as persistent trend. Integrating het- erogeneous sources of information—technical indicators, text-derived sentiment, and short-term forecasts—produces richer state representations and enables more nuanced decisions.

A. Evolution of Algorithmic Trading: From Forecasting to Intelligent Agents

Quantitative trading approaches evolved from statistical forecasting (e.g., ARIMA) to machine-learning predictors and, more recently, to decision-theoretic agents trained with reinforcement learning (RL). Classical forecasting focuses on minimizing a predictive loss (e.g., mean squared error) but does not directly address execution, transaction costs, or

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

sequential risk. RL addresses sequential decision-making by optimizing expected cumulative rewards through interaction with an environment. This makes RL particularly appealing for trading, where each action has delayed and path-dependent consequences.

B. Research Motivation and Contribution

This work presents a hybrid architecture that blends predic- tion, sentiment, and reinforcement learning to trade XAUUSD. The main contributions are:

• A multi-modal state representation that integrates portfo- lio state, historical market features, LSTM-based short- term forecasts, and FinBERT-derived sentiment momen- tum.

• An A2C-based decision core trained with a Differential Sharpe-Ratio reward to prioritize risk-adjusted perfor- mance.

• A rigorous rolling-window evaluation on one-minute XAUUSD data demonstrating substantive improvements in Sharpe Ratio and drawdown metrics relative to baseline strategies.

C. Paper Organization

Section II reviews the methods and related work. Section III details the hybrid architecture and modeling choices. Sec- tion IV explains the experimental setup. Section V presents results and analysis. Section VI discusses implications and future paths. Section VII concludes.

II. FOUNDATIONAL METHODOLOGIES AND RELATED WORK

This section summarizes the theoretical building blocks that inform our design: reinforcement learning paradigms, sentiment augmentation, and sequence forecasting.

A. Reinforcement Learning for Sequential Financial Decision- Making

Reinforcement learning models an agent interacting with an environment, formalized as a Markov Decision Process (MDP) defined by the tuple (S, A, P, R, γ) where S denotes states, A actions, P transition probabilities, R rewards, and γ a discount factor. Policy-based and value-based approaches exist. Value- based methods such as Deep Q-Networks (DQNs) approxi- mate Q(s, a) directly and benefit from experience replay and target networks for stability. Policy-gradient and actor–critic methods instead approximate a parameterized policy $\pi\theta(a|s)$, offering better handling of continuous actions and stochastic policies. Advantage Actor–Critic (A2C) reduces variance by using an advantage function $A(s, a) = Q(s, a) - V(s)$ for policy updates.

B. State Augmentation with Exogenous Information

Relying solely on price series constrains an agent's situa- tional awareness. External signals—textual sentiment, macro indicators, and cross-asset relationships—provide orthogonal information. Transformer-based models like FinBERT, pre- trained on financial corpora and fine-tuned for sentiment tasks, produce calibrated sentiment scores that better reflect financial language nuances than generic models. Incorporating senti- ment momentum (aggregated, time-lagged sentiment vectors) captures persistence in market mood and helps disambiguate transient headlines from trend-aligned narratives.

C. Time-Series Forecasting with Recurrent Neural Networks

Long Short-Term Memory (LSTM) networks address vanishing-gradient issues in RNNs and model long-range dependencies in price sequences. In hybrid systems, LSTMs act as predictive modules supplying short-horizon forecasts that an RL agent uses to anticipate near-term dynamics. We treat LSTM outputs as predictive features rather than hard trade signals, allowing the RL policy to integrate predictions probabilistically with other cues.

III. A HYBRID AI ARCHITECTURE FOR XAUUSD TRADING

This section describes the architecture, state design, reward engineering, and network topologies used in the experiments.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

1) FinBERT Sentiment Processor: ingests timestamped finance headlines and emits normalized sentiment scores in [−1, +1].
2) LSTM Predictor: consumes short windows of 1-minute OHLCV series and outputs m-step ahead price esti- mates.
3) A2C Decision Core: accepts a composite state and outputs a stochastic policy over discrete actions {Buy, Sell, Hold}.

A. Multi-Modal State Representation

At time t the agent receives state vector St built from:
1) Internal state: portfolio value Vt, cash Ct, current position Pt.
2) Market history: normalized price window {pt−w+1, . . . , pt} and indicators (EMA, RSI).
3) Sentiment momentum: vector et = [et, et−1, . . . , et−l+1] with aggregated FinBERT scores.
4) Predictive vector: LSTM outputs pˆt+1:t+m providing m short-term forecasts.
Combining these yields a rich representation that spans retro- spective, contemporaneous, and prospective information.

B. Reward Engineering: Differential Sharpe Ratio

We define the stepwise reward using the Differential Sharpe Ratio (DSR) to encourage improvements in risk-adjusted per- formance. Let returns series up to time t be {r1, . . . , rt} with sample mean μt and standard deviation σt. The Sharpe Ratio is SRt = μt/σt (risk-free rate assumed zero for simplicity). The reward is

$$R_t = SR_t - SR_{t-1}. \quad (1)$$

Using DSR guides the agent towards policies that improve volatility-adjusted returns rather than maximizing raw profit alone.

C. A2C Network Architecture and Training

Both Actor and Critic networks share an input layer aligned to St followed by three hidden dense layers (64 units each, ReLU). The Actor emits a softmax over three actions; the Critic outputs a scalar value estimate. Training uses syn- chronous A2C updates with entropy regularization to foster exploration and gradient clipping to stabilize learning.

IV. EMPIRICAL EVALUATION AND METHODOLOGY

This section details data, experimental protocol, baselines, metrics, and implementation choices.

A. Data and Preprocessing

We employ one-minute OHLCV XAUUSD ticks covering January 2018 through December 2023. Headlines from major financial outlets were time-aligned to market ticks and tok- enized for FinBERT inference. Price inputs were normalized using rolling z-score scaling; technical indicators were com- puted on the same rolling windows. To prevent look-ahead, all sentiment and LSTM predictions strictly use data available at the decision time.

B. Training and Rolling-Window Evaluation

To assess temporal generalization, we use a rolling-window scheme:
1) Train on a two-year window (24 months).
2) Test on the following three months.
3) Slide the window forward by three months and repeat. This emulates periodic model re-training used in real deploy- ments and ensures each test fold is strictly out-of-sample.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

TABLE I: Aggregated Performance Metrics Across Rolling- Window Folds

Strategy	CR (%)	AR (%)	Sharpe	MDD (%)	Vol. (%)
Hybrid A2C (Full)	**185.4**	**32.6**	**2.15**	**-12.8**	18.2
A2C-Vanilla	62.1	12.8	0.95	-24.7	25.9
LSTM Bot	35.8	7.9	0.52	-31.0	30.1
Buy-and-Hold	48.5	10.2	0.68	-25.6	22.5

C. Baselines

We compare the proposed Hybrid A2C against:

• Buy-and-Hold (BH): buy at period start, hold until end.
• A2C-Vanilla: same A2C architecture but without senti- ment and LSTM features.
• LSTM-Predictive Bot: rule-based trader acting on im- mediate LSTM direction signals (threshold-based).

D. Transaction Costs and Execution

A transaction cost of 0.05% per trade was included. Slip- page and market impact were approximated by tightening thresholds and increasing effective cost slightly in stress tests. Position sizing is discrete (unit trades) in the current experiments; future work will consider continuous sizing.

E. Evaluation Metrics

We report cumulative and annualized returns, Sharpe and Sortino ratios, maximum drawdown (MDD), and annualized volatility. Statistical stability is evaluated across multiple rolling folds and by reporting mean and standard error for main metrics.

V. RESULTS AND IN-DEPTH ANALYSIS

This section summarizes the empirical findings and con- ducts ablation and behavioral analyses.

A. Performance Summary

Table I aggregates performance measures across test folds.
The hybrid agent attains substantially higher Sharpe and improved drawdown control relative to baselines.

B. Ablation Study

To quantify component contributions we evaluated:

• Hybrid w/o LSTM: remove predictive vector,
• Hybrid w/o Sentiment: remove sentiment momentum,
• A2C-Vanilla: remove both augmentations.

Sharpe ratios in these settings were 1.58, 1.31, and 0.95 respectively, indicating both augmentations yield material and complementary benefits.

C. Behavioral Analysis: Sentiment Dependence

We measured the hybrid agent's excess return α over A2C- Vanilla as a function of news density and sentiment–price correlation. When news density exceeded 15% and short-term sentiment–price correlation was above 0.1, α rose sharply. This indicates the agent amplifies sentiment influence when the signal is both plentiful and predictive, and attenuates it when noisy.

D. Robustness and Out-of-Sample Checks

Out-of-sample evaluation on early-2024 data produced

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

Strategy CR (%) AR (%) Sharpe MDD (%) Vol. (%) Sharpe ratios above 1.9 and MDD below 15%, demon-rating transferability. Sensitivity analyses over transactionst (0.03%–0.1%) and forecast horizon m confirmed the ybrid's relative resilience: performance degrades gradually der heavier costs but remains favorable versus baselines.

VI. DISCUSSION AND FUTURE DIRECTIONS

This study provides empirical evidence that combining short-term forecasting with sentiment-aware state augmentation yields risk-sensitive, proactive trading policies. The LSTM predictor provides anticipatory cues while FinBERT sentiment supplies behavioral context; the A2C agent reconciles both to produce balanced actions.

A. Practical Implications

For practitioners, the hybrid approach suggests a pragmatic route to enhance automated trading systems: rather than replacing existing signal pipelines, embedding predictions and sentiment as features into a decision-theoretic agent can yield superior risk-adjusted outcomes while retaining interpretability via ablation diagnostics.

B. Limitations

Key limitations include dependence on news feed quality and latency, potential overfitting despite rolling validation, and discrete action/position formulation. Execution-level phenom- ena (order-book depth, market-impact microstructure) are not explicitly modeled; integrating low-latency execution simula- tors remains future work.

C. Future Work

Planned extensions include:

- Implementing continuous control (e.g., SAC) for joint direction and sizing decisions.
- Extending the framework to multi-asset portfolios to leverage cross-asset signals and hedging.
- Incorporating explainability modules (attention, SHAP) to improve transparency for end-users and regulators.
- Deploying an online learning variant that adapts contin- uously to non-stationary regimes.

VII. CONCLUSION

We presented a proactive, sentiment-augmented A2C frame- work for XAUUSD trading that integrates LSTM forecasts and FinBERT sentiment into a comprehensive state representation, trained using a Differential Sharpe-Ratio reward. Rolling- window experiments on one-minute data demonstrate mean- ingful gains in risk-adjusted returns and drawdown control versus standard baselines. The results underscore the value of multi-modal intelligence—prediction, sentiment, and decision- theoretic learning—in addressing the complexity of commod- ity FX markets.

ACKNOWLEDGMENT

The authors acknowledge the computational resources pro- vided by the MITADT University research lab. No external funding was used for this study.

REFERENCES

[1] A. Nan, A. Perumal, and O. R. Zaiane, "Sentiment and Knowledge- Based Algorithmic Trading with Deep Reinforcement Learning," arXiv preprint arXiv:2001.09403, 2020.
[2] F. C. L. Paiva, L. K. Felizardo, R. A. C. Bianchi, and A. H. R. Costa, "Intelligent Trading Systems: A Sentiment-Aware Reinforcement Learn- ing Approach," arXiv preprint arXiv:2112.02095, 2021.
[3] A. Varela, "Achilles: Neural Network to Predict the Gold vs. US Dollar Integration with Trading Bot for Automatic Trading," unpublished manuscript, 2023.
[4] Hugging Face, "ProsusAI/FinBERT," Available: https://huggingface.co/ ProsusAI/finbert.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

[5] MathWorks, "Deep Q-Network (DQN) Agent," Available: https://www. mathworks.com/help/reinforcement-learning/ug/dqn-agents.html.
[6] A. Belantari, "Deep Reinforcement Learning – A2C for Portfolio Opti- mization," Medium, 2025. Available: https://medium.com/@abatrek059/ deep-reinforcement-learning-a2c-portfolio-optimization-347139c7c447.