NarrativeRL: Narrative-Conditioned Reinforcement Learning for Anticipatory Commodity Hedging

Romina Hashami
University of Essex
Colchester, Essex, United Kingdom
rh23592@essex.ac.uk

Felipe Maldonado
University of Essex
Colchester, United Kingdom
felipe.maldonado@essex.ac.uk

Abstract

Commodity markets react rapidly to evolving geopolitical, macroeconomic, and supply-chain narratives, yet traditional volatility forecasting frameworks often fail to capture the semantic structure of financial news. This paper proposes NarrativeRL, an interpretable generative-AI framework for commodity market risk forecasting and adaptive hedging. Using more than 645,000 Reuters oil-market headlines from 2014–2024, we construct a Narrative Signal Index (NSI) combining attention intensity, tone imbalance, semantic novelty, narrative disagreement, geopolitical risk, and supply-shock signals extracted via sentence-transformer embeddings, topic-aware decomposition, and a large language model interpretation layer, and embed these representations into a reinforcement-learning-based hedging module that adapts hedge ratios under transaction costs and changing market regimes. Empirical results show that narrative-aware representations improve realised volatility forecasting relative to HAR-based benchmarks, most clearly at the 22-day horizon (Diebold–Mariano $p = 0.039$), with narrative effects becoming larger in high-volatility regimes. In out-of-sample hedging experiments, NarrativeRL attains the highest Sharpe ratio, narrowly ahead of volatility targeting, and reduces $CVaR_{95}$ by 41% relative to the unhedged position, while volatility-targeting and HAR-based hedges attain lower tail risk at the cost of lower returns. NarrativeRL remains ahead of the unhedged benchmark across all tested transaction-cost settings. These findings suggest that interpretable generative AI can support adaptive financial decision systems by transforming evolving market narratives into deployable risk-monitoring and hedging signals.

CCS Concepts

• **Computing methodologies** → **Machine learning**; *Natural language processing*; • **Applied computing** → *Economics*.

Keywords

reinforcement learning, financial markets, hedging, large language models, text-based signals, volatility, risk management

ACM Reference Format:
Romina Hashami and Felipe Maldonado. 2026. NarrativeRL: Narrative-Conditioned Reinforcement Learning for Anticipatory Commodity Hedging. In *7th ACM International Conference on AI in Finance (ICAIF '26), November 14–17, 2026, Milan, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3858063.3858751



This work is licensed under a Creative Commons Attribution 4.0 International License.
ICAIF '26, Milan, Italy
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-3087-0/2026/11
https://doi.org/10.1145/3858063.3858751

1 Introduction

Crude oil markets are highly sensitive to geopolitical conflict, supply disruptions, macroeconomic uncertainty, and rapidly evolving financial narratives. Major market events such as the COVID-19 pandemic and the Russia–Ukraine war illustrate how large volatility shocks can be preceded by narrative escalation in financial news. This is consistent with evidence that news-based geopolitical risk measures foreshadow adverse economic and financial outcomes more broadly [5], a pattern we document specifically for oil market volatility in Section 4.1 and the case study in Section 4.6. Traditional hedging systems, however, remain primarily reactive because they rely predominantly on historical price dynamics and volatility persistence.

Existing volatility forecasting and hedging approaches are largely built upon econometric frameworks such as GARCH and HAR-RV models, which use historical return information to estimate future risk [7]. Although these methods perform well under stable conditions, they struggle to adapt to rapidly changing narrative environments driven by geopolitical stress, supply-chain disruptions, and uncertainty shocks. Reinforcement learning (RL) has recently emerged as a promising alternative for dynamic hedging and portfolio optimisation because it enables adaptive sequential decision-making under uncertainty [4]. Nevertheless, existing RL-based financial systems typically define the state space using price-derived variables alone, making them inherently backward-looking and reactive to realised volatility.

At the same time, advances in natural language processing and large language models (LLMs) have enabled richer extraction of information from financial news. Recent studies show that textual information can improve financial forecasting performance, particularly in commodity markets where geopolitical narratives strongly influence volatility dynamics [14, 16]. However, most existing applications use text only as an exogenous predictor in forecasting models – for example, adding a news-sentiment or uncertainty index as an extra regressor in an otherwise standard HAR-RV specification [14, 16]. Relatively little work investigates whether narrative information extracted from large-scale financial news can directly guide sequential hedging decisions in adaptive financial systems. This paper proposes *NarrativeRL*, a narrative-conditioned reinforcement learning framework for dynamic hedging in crude oil markets. The framework integrates large-scale Reuters financial news with realised-volatility dynamics to construct a multi-dimensional Narrative Signal Index (NSI) capturing attention intensity, tone, semantic

novelty, disagreement, geopolitical stress, and supply-shock narratives. These narrative representations are then embedded directly into the state space of a reinforcement-learning hedging agent.

Unlike traditional price-only hedging systems, the proposed framework enables anticipatory risk management by conditioning decisions on latent narrative stress before volatility fully materialises in market prices. In addition, we incorporate a generative-AI interpretation layer based on large language models to generate structured narrative explanations associated with periods of elevated market stress, improving interpretability and transparency.

Using Reuters oil-market headlines from 2014–2024 and high-frequency Brent crude oil futures data, we evaluate the proposed framework across volatility forecasting, regime analysis, and out-of-sample hedging tasks. Results show that narrative signals contain economically meaningful information beyond traditional realised-volatility predictors. The reinforcement-learning agent learns substantially more defensive hedge allocations during periods of elevated narrative stress, particularly surrounding the Russia–Ukraine conflict. The proposed framework achieves a stronger risk-return trade-off than the unhedged and classical minimum-variance positions under realistic transaction costs, including a 41% reduction in CVaR_{95} relative to the unhedged benchmark, while simpler volatility-based hedges attain lower tail risk at the cost of lower returns.

The contributions of this paper are fourfold:

(1) We construct a multi-dimensional narrative risk representation from large-scale financial news using semantic, topical, and LLM-derived features.
(2) We introduce narrative-conditioned state representations for reinforcement-learning-based hedging in commodity markets.
(3) We show, through a policy analysis and an event study, that narrative information enables anticipatory hedging behaviour before realised volatility peaks.
(4) We provide an interpretable generative-AI framework linking financial narratives with adaptive risk-management decisions.

The remainder of the paper is organised as follows: Section 2 reviews the related literature; Section 3 describes the model and methodology. Section 4 presents the results and benchmark comparisons. Section 5 discusses the results and limitations, and Section 6 concludes the paper.

2 Related Work

Hedging and Risk Management. Hedging has traditionally been addressed using statistical and econometric approaches, with minimum-variance strategies forming the standard benchmark [8], typically estimating time-varying covariances and volatility dynamics via GARCH and HAR-RV models [1, 7]. While effective under relatively stable market conditions, such approaches remain largely reactive because they rely exclusively on historical price information, and consequently may fail to capture emerging risks associated with geopolitical tensions, supply disruptions, or rapidly evolving market narratives before these are fully reflected in realised volatility. This limitation is especially important in commodity markets, and particularly crude oil, given the strong sensitivity of prices to geopolitical uncertainty, macroeconomic expectations, and supply-chain shocks.

Reinforcement Learning for Hedging. Reinforcement learning (RL) offers a flexible alternative by formulating hedging as a sequential decision-making problem. The deep hedging framework of [4] demonstrated that RL-based agents can outperform classical hedging methods by directly optimising risk-sensitive objectives under transaction costs, later extended via distributional reinforcement learning for tail-risk management [6, 13] and applied to commodity and energy markets, for example to natural gas storage optimisation under uncertain market conditions [2]. Despite these advances, existing RL-based hedging systems define the state space primarily using price-derived variables such as returns, realised volatility, or technical indicators, remaining reactive to realised market stress rather than anticipatory to evolving narrative conditions.

Text-Based Signals in Financial Markets. Financial news contains predictive information relevant for volatility forecasting and risk assessment: newspaper-based uncertainty measures capture dimensions of economic and financial stress beyond conventional market variables [3], and subsequent work incorporated textual features into volatility forecasting via sentiment indicators, topic models, and news-count variables [12, 15], including applications to crude oil markets specifically [14, 16]. However, these approaches typically integrate text as an exogenous regression predictor and do not connect narrative information to adaptive decision-making systems such as dynamic hedging agents.

Large Language Models in Finance. Recent advances in LLMs have enabled richer extraction of contextual and semantic information from unstructured financial text, with emerging applications including financial sentiment analysis [9], knowledge graph construction [10], and market prediction using generative AI [11]. Nevertheless, most existing financial applications of LLMs remain focused on prediction or classification tasks, and the integration of LLM-derived narrative intelligence into reinforcement-learning-based hedging frameworks remains largely unexplored.

Positioning of This Work. This paper bridges these strands by introducing narrative-conditioned state representations for reinforcement-learning-based hedging in crude oil markets. Unlike prior studies that use textual information solely for prediction, we embed LLM-extracted narrative signals directly into the state space of an RL hedging agent.

Our framework combines attention, tone, semantic novelty, disagreement, geopolitical risk, and supply-shock signals into a unified Narrative Signal Index (NSI). These narrative representations allow the agent to condition hedging decisions on latent market stress before it becomes fully observable in realised volatility. Consequently, the proposed system enables anticipatory hedging behaviour rather than purely reactive risk management.

To the best of our knowledge, this is the first study integrating large-scale narrative representations extracted from financial news with reinforcement-learning-based hedging in commodity markets.

3 Methodology

This section presents the proposed *NarrativeRL* framework for dynamic hedging in crude oil markets. The framework combines high-frequency realised-volatility dynamics with large-scale narrative information extracted from financial news to construct an interpretable reinforcement-learning system for adaptive risk management.

3.1 Market Environment and Problem Setup

Let $\{P_t\}_{t=1}^{T}$ denote the daily crude oil futures price process observed over a discrete trading horizon. Daily log returns are defined as

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right), \tag{1}$$

where r_t denotes the close-to-close daily return computed from the observed price process $\{P_t\}_{t=1}^{T}$. This is distinct from the 5-minute intraday return $r_{t,i}$ introduced in Section 3.2 and the hedged portfolio return r_t^h defined in Eq. (3) below.

At each trading day t, the agent selects a hedge ratio a_t from a finite action space $\mathcal{A}$:

$$a_t \in \mathcal{A} := \{0, 0.25, 0.50, 0.75, 1.00\}. \tag{2}$$

This five-level discretisation was chosen to balance interpretability of the resulting Q-table against the curse of dimensionality in tabular Q-learning: finer discretisation (e.g., 10 or 20 levels) increases the action space linearly but becomes sample-inefficient given the sparsity of extreme-narrative states in the sample (Section 4.3). Section 4.3 confirms that all five actions are actively selected out-of-sample (Figure 6), indicating this granularity is sufficient to support a non-degenerate policy while remaining tractable for the state-action space defined in Section 3.4.1.

The hedge ratio represents the proportion of exposure hedged at time t. A value of $a_t = 0$ corresponds to a fully unhedged position, while $a_t = 1$ denotes a fully hedged allocation.

The resulting hedged portfolio return is

$$r_t^h = (1 - a_t) r_{t+1} - c|a_t - a_{t-1}|, \tag{3}$$

where $c > 0$ denotes proportional transaction costs associated with hedge rebalancing.

The dynamic hedging problem is formulated as a finite-state Markov decision process (MDP) $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ is the state space (constructed from realised volatility RV_t, Section 3.2 Eq. (4), and the Narrative Signal Index NSI_t, full construction in Section 3.4.1 Eq. (13)), $\mathcal{A}$ the action space (Eq. (2)), $\mathcal{P}$ the transition dynamics, $\mathcal{R}$ the reward function, and $\gamma \in (0, 1)$ the discount factor.

3.2 Realised Volatility and Jump Dynamics

Market volatility is estimated using 5-minute Brent crude oil futures returns. Let $r_{t,i}$ denote the i-th intraday return on day t, and M the number of intraday observations. Table 1 presents the realised-volatility, jump-detection, and HAR-component quantities used throughout the framework.

To separate continuous volatility variation from discontinuous jump behaviour, the jump ratio J_t, whose denominator is the bipower-variation estimate of the continuous component, approximates abnormal jump intensity. Jump days and high-volatility regimes are identified using expanding-window quantile thresholds computed over $\{1, \ldots, t-1\}$ only, ensuring regime classification at day t uses no forward-looking information. To capture heterogeneous volatility persistence, we additionally construct the daily, weekly, and monthly HAR-style realised-volatility components shown in Table 1.

Table 1: Realised volatility, jump-detection, HAR components.

Quantity	Definition	Eq.
Realised volatility	$RV_t = \sum_{i=1}^{M} r_{t,i}^2$	(4)
Jump ratio	$J_t = \dfrac{RV_t}{\frac{\pi}{2}\sum_{i=2}^{M} \lvert r_{t,i}\rvert\lvert r_{t,i-1}\rvert}$	(5)
Jump indicator	$\mathbf{1}_{jump,t} = \mathbf{1}\left(J_t > Q_{0.90}(J_{1:t-1})\right)$	(6)
High-volatility regime	$\mathbf{1}_{HV,t} = \mathbf{1}\left(RV_t > Q_{0.90}(RV_{1:t-1})\right)$	(7)
Daily HAR component	$RV_t^{(d)} = \log(RV_t)$	(8)
Weekly HAR component	$RV_t^{(w)} = \frac{1}{5}\sum_{i=1}^{5} \log(RV_{t-i})$	(9)
Monthly HAR component	$RV_t^{(m)} = \frac{1}{22}\sum_{i=1}^{22} \log(RV_{t-i})$	(10)

3.3 Narrative Signal Extraction

Lexical and Topic-Based Features. Reuters oil-market headlines are processed daily to extract narrative characteristics associated with financial stress, geopolitical uncertainty, and supply disruptions, using lexicon-based measures of negative tone, uncertainty, and attention intensity. For each day t, headlines are classified into predefined thematic categories $\mathcal{T} = \{\text{supply, demand, geopolitics, policy, inventories}\}$, and narrative concentration is measured through topic entropy $\mathcal{E}_t = -\sum_{k=1}^{K} p_{k,t} \log(p_{k,t})$, where $p_{k,t}$ denotes the proportion of headlines on day t associated with topic k.

3.3.1 Semantic Novelty and Narrative Disagreement. To capture semantic shifts in financial narratives, headlines are embedded using Sentence-BERT representations. Daily semantic representations are aggregated as

$$\bar{e}_t = \frac{1}{N_t}\sum_{j=1}^{N_t} e_{t,j}, \tag{11}$$

where N_t denotes the number of headlines observed on day t, and $e_{t,j} \in \mathbb{R}^d$ the embedding vector associated with headline j on day t.

Narrative novelty is defined as the cosine distance between the current narrative embedding and a rolling historical centroid, $\nu_t = 1 - \cos(\bar{e}_t, \frac{1}{w}\sum_{i=1}^{w} \bar{e}_{t-i})$, distinct from N_t (the headline count in Eq. (11).

Narrative disagreement is measured through intra-day embedding dispersion, $\mathcal{D}_t = \frac{1}{N_t}\sum_{j=1}^{N_t} \|e_{t,j} - \bar{e}_t\|_2^2$, capturing abrupt semantic transitions and fragmentation in market narratives.

3.3.2 LLM-Based Narrative Interpretation. A locally deployed language model (Qwen2.5-1.5B-Instruct) transforms daily collections of Reuters oil-market headlines into structured narrative representations, using constrained few-shot prompting with deterministic

decoding to extract five variables per day: narrative risk intensity, dominant narrative channel (restricted to the ontology $\mathcal{T}$ = {supply, demand, geopolitical, policy, inventory, mixed}), uncertainty level, market pressure direction, and narrative disagreement intensity, with risk and disagreement mapped onto fixed ordinal scales using few-shot examples from major historical oil-market events (e.g., OPEC supply adjustments, COVID-19 demand collapses, geopolitical escalation). Outputs were validated via rule-based JSON parsing and schema-constrained extraction; the LLM layer was used exclusively for structured narrative-state extraction and post-hoc interpretability, and did not directly optimise the RL policy or generate trading decisions. To evaluate prompt robustness and reproducibility, we tested alternative few-shot prompt formulations across a representative headline sample, observing high inter-prompt agreement ($\kappa > 0.90$), which indicates that the extracted narrative categories are stable across prompt formulations; agreement with human-annotated labels was not assessed.

Temporal Alignment and Look-Ahead Prevention. To avoid look-ahead bias, narrative variables at day t use only headlines available prior to the forecasting horizon for RV_{t+1}, with out-of-sample forecasts and hedging decisions evaluated using strictly forward-looking windows, consistent with the expanding/backward-looking convention used throughout (Section 3.2).

Narrative Signal Index Construction. The Narrative Signal Index (NSI) is constructed as $NSI_t = \frac{1}{K}\sum_{k=1}^{K} z_{k,t}$, where the components $z_{k,t}$ are rolling backward-looking z-scores (trailing window $w = 30$) for attention intensity, tone imbalance, semantic novelty, narrative disagreement, geopolitical risk, and supply-shock narratives. We additionally estimate a PCA-based narrative factor, $NSI_t^{PCA} = \mathbf{w}^\top \mathbf{z}_t$, where $\mathbf{w}$ is the first principal-component loading vector.

3.4 Regime-Dependent Narrative Effects

To evaluate whether narrative information becomes more informative during stressed market conditions, we estimate regime-interaction specifications of the form:

$$\begin{aligned}\log RV_{t+1} =& \beta_0 + \beta_1 RV_t^{(d)} + \beta_2 RV_t^{(w)} + \beta_3 RV_t^{(m)} \\ &+ \beta_4 NSI_t + \beta_5 (NSI_t \times HV_t) + \varepsilon_t,\end{aligned} \tag{12}$$

where HV_t denotes the high-volatility regime indicator.

This specification allows narrative effects to vary conditionally across volatility environments.

3.4.1 Narrative-Conditioned State Space. The RL state is constructed by jointly discretising narrative stress and realised-volatility conditions,

$$s_t = (q_t^{NSI}, q_t^{RV}) \in \mathcal{S}, \tag{13}$$

where $q_t^{NSI} \in \{Q1, Q2, Q3, Q4, Q5\}$, discretise the NSI into quintiles, computed using expanding-window quantile boundaries over $NSI_{1:t-1}$, while realised volatility is analogously discretised into three empirical regimes, $q_t^{RV} \in \{Low, Medium, High\}$.

This formulation enables the framework to distinguish between environments characterised by elevated narrative stress despite currently low realised volatility.

3.4.2 Risk-Sensitive Reinforcement Learning. Let $\pi : \mathcal{S} \to \mathcal{A}$ denote a hedging policy mapping states to actions. The agent learns a policy by maximising discounted expected utility:

$$\max_{\pi} \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}_t \right], \tag{14}$$

where the expectation $\mathbb{E}_\pi$ is taken over trajectories induced by following π. The state-action value function $Q^\pi(s_t, a_t)$, representing the expected discounted cumulative reward obtained by taking action a_t in state s_t and following policy π thereafter, is defined as

$$Q^\pi(s_t, a_t) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k \mathcal{R}_{t+k} \mid s_t, a_t \right]. \tag{15}$$

We employ tabular Q-learning with Bellman recursion:

$$\begin{aligned}Q(s_t, a_t) \leftarrow & Q(s_t, a_t) + \alpha \Big[\mathcal{R}_t \\ &+ \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \Big],\end{aligned} \tag{16}$$

where α denotes the learning rate and γ the discount factor. The agent is trained using an ϵ-greedy exploration policy with $\alpha = 0.10$ and $\gamma = 0.95$. The agent is trained on 2014–2020, and the resulting greedy policy is evaluated out-of-sample on 2021–2024. The use of reinforcement learning is motivated by the sequential and state-dependent nature of hedging decisions under evolving narrative conditions. Unlike static optimisation frameworks, RL enables hedge allocations to adapt dynamically to transitions between volatility and narrative regimes while accounting for transaction costs, downside-risk asymmetry, and path-dependent market conditions.

The proposed tabular formulation was intentionally selected to preserve interpretability and financial transparency. In contrast to deep neural RL architectures, the finite state-action structure allows direct inspection of the learned policy and reduces the risk of unstable overfitting under relatively sparse extreme-event data.

Because the state contains only the NSI quintile and RV tertile, while the reward depends on the previous hedge ratio and rolling portfolio variance, the process is partially observed from the agent's perspective. We therefore treat tabular Q-learning as a heuristic for learning a memoryless policy rather than as an exact solution of the Bellman optimality equation. Rewards are computed from the actual realised portfolio history.

3.4.3 Risk-Sensitive Reward Function. The reward function jointly penalises volatility exposure, downside tail risk, and excessive turnover:

$$\mathcal{R}_t = r_t^h - \lambda_1 \sigma_t^2 - \lambda_2 \max(0, -r_t^h)^2 - \lambda_3 |a_t - a_{t-1}|, \tag{17}$$

where:

- σ_t^2 denotes rolling portfolio variance,
- $\max(0, -r_t^h)^2$ penalises downside losses,
- $|a_t - a_{t-1}|$ captures hedge turnover.

This specification generates a risk-sensitive utility objective balancing return generation and downside-risk protection.

Benchmark Hedging Strategies. To evaluate the economic value of narrative conditioning, the proposed framework is compared against multiple benchmark hedging strategies: unhedged exposure, minimum-variance hedging, HAR-based dynamic hedging, and volatility-targeting hedging. We additionally construct *threshold-based* hedging rules, which fully hedge ($a_t = 1$) whenever a chosen signal exceeds a fixed historical quantile and remain unhedged ($a_t = 0$) otherwise. Three variants are considered: an NSI-threshold rule (triggered by q_t^{NSI}), an RV-threshold rule (triggered by q_t^{RV}), and a combined rule (triggered by either signal). Unlike NarrativeRL, threshold rules apply a fixed, hand-specified cutoff rather than a learned policy over the full state space.

Forecast Evaluation and Risk Metrics. Forecasting performance is evaluated using out-of-sample Diebold–Mariano tests comparing HAR benchmarks against narrative-augmented specifications, with a heteroskedasticity- and autocorrelation-consistent (HAC, Newey–West) variance estimator. Economic performance is evaluated using Sharpe ratios, maximum drawdown, Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), defined as $CVaR_\alpha = \mathbb{E}[r_t \mid r_t \leq VaR_\alpha]$, and certainty-equivalent utility under quadratic risk aversion, $CE = \mathbb{E}[r_t] - \frac{\eta}{2} Var(r_t)$, where $\eta > 0$ is the risk-aversion coefficient, distinct from the RL discount factor γ.

3.4.4 Event-Based Narrative Stress Analysis. To evaluate whether the framework captures emerging market stress before volatility materialises in prices, we conduct an event-study analysis centred on the Russia–Ukraine conflict between November 2021 and April 2022. For each trading day, the learned RL policy generates a hedge allocation conditional on the joint narrative-volatility state $s_t = (q_t^{NSI}, q_t^{RV})$ defined in Eq. (13).

Interpretability and Narrative Validation. To validate the economic meaning of extreme narrative states, a secondary LLM pass generates a concise natural-language explanation for high-NSI trading days, identifying the dominant geopolitical or supply-driven mechanism associated with elevated market stress. We evaluate the consistency between these LLM-derived channels and rule-based topic classifications from predefined commodity-market lexicons, confirming coherent market structure rather than spurious semantic variation, with JSON parsing and schema-constrained extraction achieving stable output generation across the full sample.

4 Results

4.1 Narrative Dynamics and Market Stress

Figure 1 presents the evolution of the Narrative Signal Index (NSI) together with realised volatility over the full sample period. Several major geopolitical and macroeconomic disruptions coincide with elevated narrative stress, including the COVID-19 oil-market collapse in 2020 and the Russia–Ukraine conflict in 2022. Importantly, spikes in the NSI frequently precede increases in realised volatility rather than occurring simultaneously, suggesting that narrative conditions contain forward-looking information not yet fully reflected in prices.

To better understand the composition of the NSI, Figure 2 visualises the monthly evolution of the underlying narrative sub-indices.

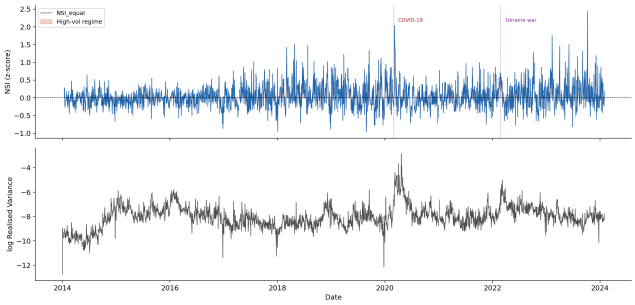


Figure 1: Narrative Signal Index (NSI) and realised volatility over time. Elevated narrative stress aligns with major geopolitical and macroeconomic disruptions, including COVID-19 and the Russia–Ukraine conflict.

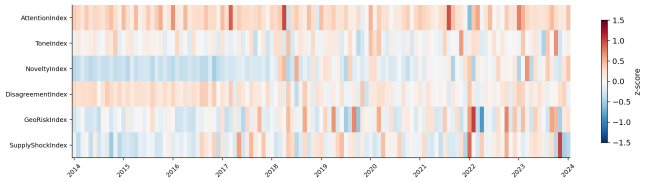


Figure 2: Monthly evolution of standardised narrative sub-indices; each row is z-scored against its own full-sample mean and standard deviation, so color indicates each sub-index's deviation from its own historical norm (red = above average, blue = below), not a comparison across sub-indices. GeoRiskIndex peaks sharply at the 2022 Russia–Ukraine invasion, consistent with the anticipatory hedging behaviour in Section 4.6.

Attention and disagreement components become particularly elevated during crisis periods, while geopolitical-risk and supply-shock signals dominate around the onset of the Russia–Ukraine conflict. The figure highlights that the NSI captures heterogeneous dimensions of market narratives rather than a single sentiment channel.

The unconditional association between NSI and future volatility is positive but modest. The Spearman rank correlation between NSI and subsequent absolute returns is $\rho = 0.085$ ($p < 0.001$, $n = 3{,}126$). A Kruskal–Wallis test confirms that return distributions differ significantly across NSI quintiles ($H = 25.37$, $p < 0.001$).

4.2 Predictive Information Content of Narrative Signals

Figure 3 reports the relationship between NSI quintiles and subsequent realised volatility dynamics. Both the average absolute next-period return and the probability of entering a high-volatility regime increase with narrative stress. Q5 exceeds Q1 in both next-period absolute return (0.0176 vs. 0.0135) and high-volatility entry rate (15.0% vs. 10.2%).

Although the relationship between NSI quintiles and high-volatility entry rate is broadly positive, some non-monotonicity is observed at intermediate quintiles, reflecting noise in quintile boundary estimation over the full sample and consistent with the weak unconditional signal strength of the NSI.

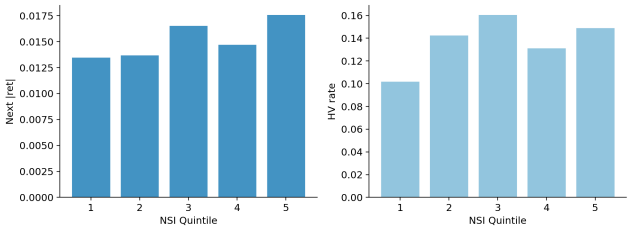


Figure 3: Relationship between NSI quintiles and future market instability. Higher narrative stress is associated with larger future return magnitudes and a higher probability of entering high-volatility (HV) regimes ($H = 25.37$, $p < 0.001$, Kruskal–Wallis).

Table 2 reports HAR-RV regressions predicting $\log \widehat{RV}_{t+1}$. Adding NSI (M2) raises adjusted R^2 from 0.478 to 0.486 ($\Delta = +0.008$; NSI coefficient = 0.273, $t = 7.23$, $p < 0.001$). The interaction with the high-volatility regime (M3) is positive and significant ($\beta = 0.178$, $t = 2.13$, $p = 0.034$), confirming that narrative effects are amplified during stressed market conditions. Regime-split regressions confirm this asymmetry: the NSI coefficient is 0.197 ($t = 4.57$, $p < 0.001$) in low-volatility periods and 0.336 ($t = 4.37$, $p < 0.001$) in high-volatility periods. The Chow test rejects coefficient equality across regimes for all five NSI specifications tested (NSI_{equal}, NSI_{pca}, GeoRiskIndex, NoveltyIndex, DisagreementIndex; $F \geq 9.66$, $p < 0.001$ in every case). The NSI coefficient remains significant across all 12 robustness specifications combining alternative rolling windows (20, 30, 45 days) and sub-index dropouts (t-statistics 2.27–5.17; 12/12 specifications significant at $p < 0.05$). HAR+NSI achieves significantly lower 22-day forecast error than the HAR benchmark (Diebold–Mariano $p = 0.039$; Table 3), with weaker evidence at $h = 1$ ($p = 0.092$) and none at $h = 5$ ($p = 0.190$), so the gains are concentrated at the monthly horizon. A plausible explanation, which we do not test, is that daily narrative signals are noisy and add most when averaged over a longer horizon. The sub-index dropout specifications above indicate that no single component drives these results, although we do not formally test whether the six components capture distinct constructs.

4.3 Economic Performance of NarrativeRL

Table 4 reports out-of-sample performance across the hedging strategies and transaction-cost settings. The unhedged benchmark has $\text{CVaR}_{95} = -4.96\%$ and maximum drawdown of -51.0%. NarrativeRL at 5 basis points (bp) achieves the highest Sharpe ratio (0.465) and a 41% reduction in CVaR_{95} relative to the unhedged benchmark, while retaining most of the unhedged mean return. Its Sharpe advantage over vol-targeting (0.460) is negligible, however, and HAR dynamic hedging and vol-targeting achieve higher CE, lower tail risk, and smaller drawdowns by adopting near-constant low-variance hedging at the cost of mean return. The results therefore reflect a risk-return trade-off rather than uniform dominance: NarrativeRL preserves return-generating capacity while reducing tail risk, whereas the simpler volatility-based hedges prioritise tail protection.

Figure 4 presents cumulative wealth trajectories. The unhedged strategy experiences substantial drawdowns during crisis periods, particularly during the COVID-19 oil-market collapse. Both

Table 2: HAR-RV regressions predicting $\log \widehat{RV}_{t+1}$. HAC t-statistics in brackets (Newey–West, 5 lags).

	M1: HAR	M2: +NSI	M3: +NSI×HV
$RV_{d,t}$	-0.133^{***}	-0.131^{***}	-0.132^{***}
	$[-10.56]$	$[-10.56]$	$[-10.60]$
$RV_{w,t}$	0.839^{***}	0.797^{***}	0.791^{***}
	$[17.64]$	$[16.90]$	$[16.76]$
$RV_{m,t}$	0.105^{**}	0.151^{***}	0.159^{***}
	$[2.35]$	$[3.36]$	$[3.51]$
$\lvert\text{ret}_t\rvert$	10.814^{***}	10.354^{***}	10.211^{***}
	$[7.62]$	$[7.81]$	$[7.80]$
NSI		0.273^{***}	0.202^{***}
		$[7.23]$	$[4.82]$
NSI×HV			0.178^{**}
			$[2.13]$
adj-R^2	0.478	0.486	0.483
Δadj-R^2	—	+0.008	+0.005
N	3,125	3,125	3,106

$^{***}p < 0.01$, $^{**}p < 0.05$. HV = high-vol regime indicator.
Low-vol NSI coef: 0.197 ($t = 4.57$); High-vol: 0.336 ($t = 4.37$). Both $p < 0.001$.

Table 3: Diebold–Mariano forecast comparison: HAR vs. HAR+NSI (MSE loss). Positive DM statistic indicates HAR+NSI achieves lower forecast error. OOS period: 2021–2024.

Horizon	DM stat	p-value	Interpretation
$h = 1$	1.684	0.092^{*}	HAR+NSI better
$h = 5$	1.312	0.190	HAR+NSI better
$h = 22$	2.069	0.039^{**}	HAR+NSI better

narrative-based and volatility-based threshold strategies improve downside protection, with the RL-based strategy achieving the strongest out-of-sample Sharpe ratio.

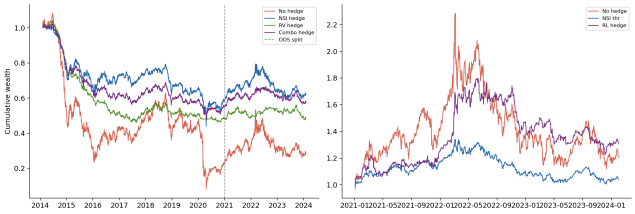


Figure 4: Cumulative wealth comparison across hedging strategies. Left: full-sample (2014–2024) cumulative wealth under the unhedged, NSI-threshold, RV-threshold, and combined-threshold strategies, with the out-of-sample (OOS) split marked. Right: out-of-sample (2021–2024) comparison between the unhedged benchmark and NarrativeRL, showing that the RL-based narrative-conditioned strategy achieves superior downside protection and reduced drawdown severity during the 2022 volatility spike relative to remaining unhedged.

The RL agent does not converge to a trivial always-hedged policy. In fact the agent selects 0% hedge on 23.4% of OOS days, 25% on 6.9%, 50% on 13.9%, 75% on 33.0%, and 100% on 22.8%, confirming the policy adapts dynamically to state conditions.

Table 5 reports an ablation analysis isolating the contribution of each state component. Price-only RL (3 states, RV tertile) reduces CVaR_{95} by 39% from -0.050 to -0.030. NSI-only RL (5 states, NSI

Table 4: Out-of-sample performance of hedging strategies (OOS 2021–2024, $n = 957$ days). NarrativeRL achieves the highest Sharpe ratio; HAR dynamic hedging and vol-targeting achieve higher CE and lower tail risk through near-constant low-variance hedging.

Strategy	Mean P&L	Std	VaR_{95}	$CVaR_{95}$	$CVaR_{99}$	Max DD	CE(η=5)
No Hedge	0.000425	0.021389	−0.0344	−0.0496	−0.0780	−0.5100	−0.000719
Min-Var Hedge (classical)	0.000348	0.020317	−0.0318	−0.0473	−0.0738	−0.5287	−0.000684
HAR Dynamic Hedge	0.000174	0.007341	−0.0124	−0.0195	−0.0318	**−0.1525**	+0.000039
Vol-Targeting Hedge	0.000199	0.006865	**−0.0110**	**−0.0168**	**−0.0254**	−0.1847	**+0.000081**
NSI Threshold	0.000070	0.007710	−0.0128	−0.0186	−0.0280	−0.2505	−0.000078
NarrativeRL (5 bp)	0.000350	0.011953	−0.0172	−0.0291	−0.0456	−0.3184	−0.000007
NarrativeRL (10 bp)	0.000156	0.011956	−0.0174	−0.0293	−0.0459	−0.3766	−0.000202
NarrativeRL (20 bp)	−0.000234	0.011967	−0.0179	−0.0299	−0.0463	−0.4785	−0.000592

Bold = best in column for the risk and utility metrics (VaR, CVaR, Max DD, CE). Annualised Sharpe ratios: NarrativeRL (5 bp) 0.465, Vol-Targeting 0.460, HAR Dynamic 0.376, No Hedge 0.315. CVa
NarrativeRL is preferred over the NSI Threshold rule at low transaction costs; the crossover lies between 5 and 10 bp.

quintile) achieves $CVaR_{95} = -0.009$, an 81% reduction, using a conservative policy that prioritises protection over return (mean P&L = −0.000019, Std = 0.004). The full NarrativeRL model achieves the best certainty equivalent among the RL variants by combining both sources of information across 15 states with all five actions active. Note that no single specification dominates on every metric: NSI-only RL achieves the best tail-risk protection among the RL variants (lowest $CVaR_{95}$ and $CVaR_{99}$) by adopting a highly conservative policy, while the full NarrativeRL model achieves the best certainty-equivalent utility among the RL variants by balancing tail protection against return-generating capacity. This trade-off motivates the joint narrative-volatility state design over either information source in isolation.

4.4 Learned Policy Structure and Anticipatory Hedging

Figure 5 visualises the greedy (highest-Q) hedge ratio and associated Q-values for each of the 15 narrative-volatility states, where Q1–Q5 denote the NSI quintiles defined in (Section 3.4.1) – Q1 the lowest-narrative-stress quintile, Q5 the highest, computed from expanding-window boundaries over past observations only. In low-volatility environments combined with low narrative stress (Q1), the agent selects minimal hedging exposure (0%). When narrative stress becomes elevated (Q4–Q5) despite still-low realised volatility, the agent increases hedge intensity to 75%. This anticipatory behaviour—responding to narrative deterioration before realised volatility fully materialises—is structurally unavailable to price-only systems that observe the same low variance and hedge 0%.

One anomalous cell warrants acknowledgement: state $Q4 \times RV3$ prescribes 0% hedge. This is the rarest joint combination in training data (< 0.5% of days); the agent likely converged to a suboptimal action due to limited observations in that state. In deployment, this state-sparsity issue can be eliminated post-hoc by enforcing a safety floor (e.g., $a_t \geq 0.50$) whenever $q_t^{\text{NSI}} \geq$ Q4, or by applying monotonic Q-table smoothing across adjacent NSI quintiles.

4.5 Robustness: Sub-Periods and Transaction Costs

We re-estimate the relationship between the NSI and next-period realised volatility across three sub-periods. The NSI coefficient is statistically significant in all three periods (Pre-COVID: $\beta = 0.298$, $t = 5.46$, $p < 0.001$; COVID: $\beta = 0.333$, $t = 3.64$, $p < 0.001$; Post-COVID: $\beta = 0.163$, $t = 2.76$, $p = 0.006$), strongest during COVID when narrative uncertainty was most extreme, suggesting the NSI captures a stable relationship rather than a crisis-specific artefact. Separately, Table 4 reports certainty-equivalent utility under transaction costs from 5 to 20 bp: NarrativeRL remains superior to the unhedged benchmark across all cost levels and outperforms the NSI threshold rule at low costs, with the crossover falling between 5 and 10 bp, above which the NSI threshold rule is preferred due to NarrativeRL's higher turnover (0.39). HAR dynamic hedging and vol-targeting retain higher CE than NarrativeRL in Table 4, at the price of lower mean returns.

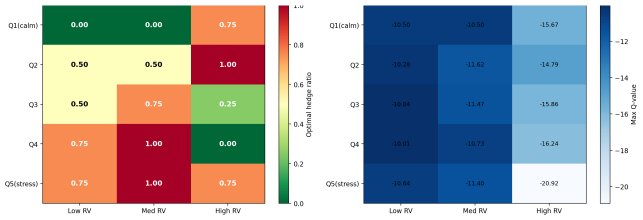


Figure 5: Learned policy structure. Left panel: greedy (highest-Q) hedge ratio in each narrative-volatility state. Right panel: corresponding maximum Q-values. Elevated narrative stress leads the agent to increase hedge intensity before realised volatility peaks (anticipatory behaviour).

4.6 Case Study: Russia–Ukraine Escalation

Figure 6 presents the three-panel timeline around the Russia–Ukraine escalation. On 2022-01-02, the NSI spiked to 2.57 standard deviations on Russia-related crude supply narratives while realised volatility remained in the Low tertile. The Q-table immediately assigned a 75% hedge ratio to the resulting (Q5, Low-RV) state, seven weeks before Russia's invasion on February 24, 2022 triggered a sharp volatility spike and Brent crude prices rose above $130/bbl by March 8. A price-only agent – i.e., the Price-only RL ablation from Table 5 (Section 4.3), conditioned solely on RV_t tertiles with no narrative information – observing the same low realised variance, would have hedged at 0%. This episode illustrates the anticipatory mechanism: narrative stress preceded the price-based signal by

Table 5: Ablation analysis of narrative and volatility state components (OOS 2021–2024). $CVaR_{95}$ reductions vs No Hedge: Price-only 39%, NSI-only 81%. NSI-only conservative policy (Std = 0.004); full model balances tail protection and return.

Strategy	Mean P&L	Std	VaR_{95}	$CVaR_{95}$	$CVaR_{99}$	Max DD	CE(η=5)	States	Actions
No Hedge	0.000425	0.021389	−0.0344	−0.0496	−0.0780	−0.5100	−0.000719	–	–
Price-only RL	0.000095	0.012664	−0.0204	−0.0300	−0.0473	−0.3860	−0.000306	3	2/5
NSI-only RL	−0.000019	0.003651	−0.0062	**−0.0092**	**−0.0136**	**−0.2169**	−0.000052	5	2/5
NarrativeRL (full)	0.000350	0.011953	−0.0172	−0.0291	−0.0456	−0.3184	**−0.000007**	15	5/5

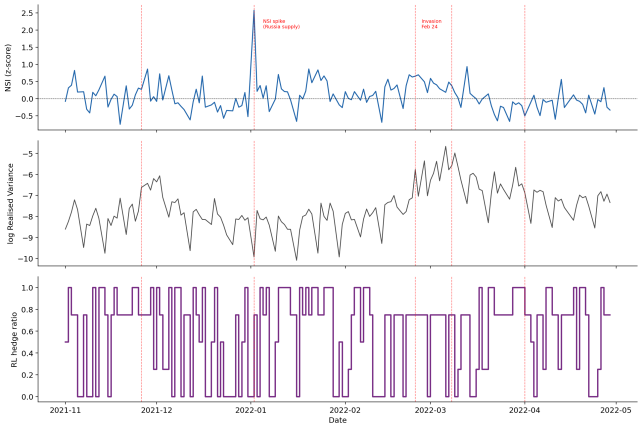


Figure 6: Anticipatory hedging during the Russia–Ukraine escalation (Nov. 2021–Apr. 2022). The NSI spike on 2022-01-02 triggers a 75% hedge seven weeks before the Feb. 24 invasion and volatility peak.

nearly two months, although a single episode cannot establish it as a general property.

4.7 LLM-Based Narrative Explanations

Table 6 reports examples from the post-hoc LLM explanation layer for the highest-NSI episodes. The examples confirm that high-NSI days correspond to identifiable market mechanisms: supply disruptions, geopolitical tensions, and policy-related uncertainty. This step does not affect the learned policy; it provides an audit trail linking the numerical NSI to concrete market narratives, validating that elevated NSI states are not arbitrary statistical artefacts.

5 Discussion

The learned policy allocates approximately 75% hedge ratios in high-narrative-stress states (Q4–Q5) even when realised volatility is still low (Figure 5) – behaviour structurally unavailable to price-only systems, which observe the same low variance and hedge at 0% (Table 5, Price-only RL), as the Russia–Ukraine case study illustrates (Section 4.6). NSI alone achieves near-random classification of high-volatility regimes as a standalone predictor (AUC ≈ 0.50), confirming that narrative value emerges through state-conditioning within the sequential decision framework, not through unconditional prediction.

The LLM's function in this framework is representational rather than predictive: it transforms unstructured Reuters headlines into structured narrative states via constrained few-shot prompting with deterministic decoding, rather than forecasting returns or volatility directly. The ablation in Table 5 is consistent with this: NSI-only states, which use no price information, cut $CVaR_{95}$ by 81% relative to no hedge, versus 39% for price-only states, even though the narrative layer never issues a forecast. The NSI-only policy is conservative (mean P&L = −0.000019, Std = 0.004), while the full model achieves the best certainty-equivalent utility among the RL variants by balancing tail protection against return generation – motivating the 15-state joint design over either information source alone. The NSI-only ablation tests the composite narrative index, which combines lexical, embedding-based and LLM-derived components, rather than the incremental contribution of LLM extraction alone.

Fixed NSI-threshold rules reduce tail risk even further (Table 4, $CVaR_{95}$ of −0.0186 versus −0.0291) but over-hedge during intermediate conditions, yielding lower CE than NarrativeRL at low transaction costs. The RL framework's advantage – dynamically conditioning on joint narrative-volatility states – comes at the cost of higher turnover (0.39), so its edge narrows as transaction costs rise, with the NSI threshold rule overtaking NarrativeRL at a cost between 5 and 10 bp (Table 4). Threshold rules therefore suit constrained, transparency-first deployments, while RL is preferable where adaptive risk-return optimisation is the priority and rebalancing costs are moderate.

Unlike deep RL, the finite Q-table gives risk managers a fully inspectable state-action mapping, and the post-hoc LLM explanations (Table 6) link high-stress states to concrete geopolitical events, supporting auditability. Limitations include: narrative features are daily-aggregated and do not exploit intraday dynamics; the tabular structure requires discretising both state variables, trading some nonlinearity for transparency; evaluation is limited to crude oil and Reuters headlines, so generalisability to other assets or text sources is untested; results demonstrate exploitable statistical structure, not causal narrative-volatility relationships; hedging results are reported without confidence intervals or multiple training seeds; the five-level hedge-ratio grid was not varied; the LLM layer is validated for prompt consistency but not against human-annotated labels; the six NSI components are not tested for redundancy beyond the sub-index dropout checks; and the policy is estimated from 15 states, some of which have few observations.

6 Conclusion

This paper introduced NarrativeRL, a narrative-conditioned reinforcement learning framework for dynamic hedging in crude oil markets, integrating LLM-based narrative extraction with realised-volatility dynamics to construct interpretable market states for sequential hedging decisions. Using Reuters oil-market news and high-frequency realised volatility data from 2014–2024, the results

Table 6: Examples of LLM-generated explanations for high-NSI episodes.

Date	Channel	LLM-generated explanation
2022-01-02	Supply	Russia-related crude oil and gas-condensate output narratives identified as the dominant source of market stress, indicating supply-side uncertainty before the subsequent escalation in energy-market risk.
2023-10-09	Geopolitics	Surge in oil prices driven by Middle East conflict-related narratives, consistent with geopolitical stress affecting oil-market expectations.
2022-12-04	Geopolitics	Russia's rejection of Western oil price caps and OPEC+ production discussions signal combined policy and geopolitical uncertainty.
2021-07-04	Supply	OPEC+ disagreement around UAE production quotas creates uncertainty over future supply coordination.

show that the proposed Narrative Signal Index (NSI) contains economically meaningful signals beyond price-based volatility alone, improves volatility forecasting, most clearly at the 22-day horizon, and enables the RL agent to implement anticipatory hedging during periods of elevated narrative risk – achieving the highest Sharpe ratio in the comparison and a substantial reduction in tail risk relative to the unhedged position while remaining fully interpretable through its finite Q-table structure, although simpler volatility-based hedges attain lower tail risk at the cost of lower returns. The Russia–Ukraine case study illustrates this anticipatory mechanism directly, with the framework responding to geopolitical escalation before realised volatility fully materialises. More broadly, the study demonstrates how large language models can function within sequential financial decision systems not merely as forecasting tools, but as mechanisms for constructing economically meaningful representations of latent market conditions – suggesting a promising direction for explainable, adaptive AI-driven risk management in high-stakes financial environments.

References

[1] Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71, 2 (2003), 579–625.

[2] Tiziano Bacaloni, Aldo Glielmo, and Marco Taboga. 2025. Natural-Gas Storage Modelling by Deep Reinforcement Learning. In *Proceedings of the ACM International Conference on AI in Finance*.

[3] Scott R. Baker, Nicholas Bloom, and Steven J. Davis. 2016. Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics* 131, 4 (2016), 1593–1636.

[4] Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. 2019. Deep Hedging. *Quantitative Finance* 19, 8 (2019), 1271–1291.

[5] Dario Caldara and Matteo Iacoviello. 2022. Measuring Geopolitical Risk. *American Economic Review* 112, 4 (2022), 1194–1225. doi:10.1257/aer.20191823

[6] Jie Cao, Jin Chen, Soroush Farghadani, John Hull, Zissis Poulos, Zhen Wang, and Jun Yuan. 2023. Gamma and Vega Hedging Using Deep Distributional Reinforcement Learning. *Frontiers in Artificial Intelligence* 6 (2023), 1129370. doi:10.3389/frai.2023.1129370

[7] Fulvio Corsi. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7, 2 (2009), 174–196.

[8] Jeff Fleming, Chris Kirby, and Barbara Ostdiek. 2001. The Economic Value of Volatility Timing. *The Journal of Finance* 56, 1 (2001), 329–352.

[9] Giorgos Iacovides, Thanos Konstantinidis, Mingxue Xu, and Danilo Mandic. 2024. FinLlama: LLM-Based Financial Sentiment Analysis for Algorithmic Trading. In *Proceedings of the 5th ACM International Conference on AI in Finance*. 134–141.

[10] Xiaohui Victor Li and Francesco Sanna Passino. 2024. FinDKG: Dynamic Knowledge Graphs with Large Language Models for Detecting Global Trends in Financial Markets. In *Proceedings of the 5th ACM International Conference on AI in Finance*. 573–581.

[11] Alejandro Lopez-Lira and Yuehua Tang. 2023. Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models. *arXiv preprint arXiv:2304.07619* (2023).

[12] Tim Loughran and Bill McDonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance* 66, 1 (2011), 35–65.

[13] Parvin Malekzadeh, Zissis Poulos, Jacky Chen, Zeyu Wang, and Konstantinos N. Plataniotis. 2024. EX-DRL: Hedging Against Heavy Losses with EXtreme Distributional Reinforcement Learning. In *Proceedings of the 5th ACM International Conference on AI in Finance*. 370–378.

[14] Yixuan Song, Mengxi He, Yudong Wang, and Yaojie Zhang. 2022. Forecasting crude oil market volatility: A newspaper-based predictor regarding petroleum market volatility. *Resources Policy* 79 (2022), 103093.

[15] Paul C. Tetlock. 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* 62, 3 (2007), 1139–1168.

[16] Danyan Wen, Mengxi He, Yudong Wang, and Yaojie Zhang. 2024. Forecasting crude oil market volatility: A comprehensive look at uncertainty variables. *International Journal of Forecasting* 40, 3 (2024), 1022–1041.