Extracting Alpha from the News Cycle

Phase 1 ML Experiments: GDELT Sentiment & Economic Calendar Integration

Dates:	May 3–17, 2026
Configurations:	72
Data Horizon:	2020 - 2026 (1,663 days)
Top Sharpe:	+2.653

78%

Configurations yielding Positive Sharpe Ratios

+2.653

Max Sharpe Attained (`USDJPY, CatBoost`)

+32%

Performance Edge vs. Vol-Adjusted Baseline

Core Hypotheses Validated

GDELT Predicts Direction

H1-aggregated tone and volume spikes carry forward-looking information.

Drift Horizon Alignment

h=24 captures maximum drift, validating `30m`–4h academic propagation windows.

Categorical ML Dominance

`CatBoost` natively exploits sparse news features better than `XGBoost`.

Instrument Sensitivity

`USDJPY` (BoJ policy) and `XAUUSD` (inflation narratives) yield the strongest signals.

Grounding the Signal in Market Microstructure

Andersen et al. (2003)

NFP surprises explain 50%+ of intraday FX variance.

↓

Design Choice: Incorporate `surprise_magnitude` calendar feature.

Lucca & Moench (2015)

Pre-announcement drift mechanics.

↓

Design Choice: Feature `event_countdown` captures vol compression.

Ehrmann & Fratzscher (2005)

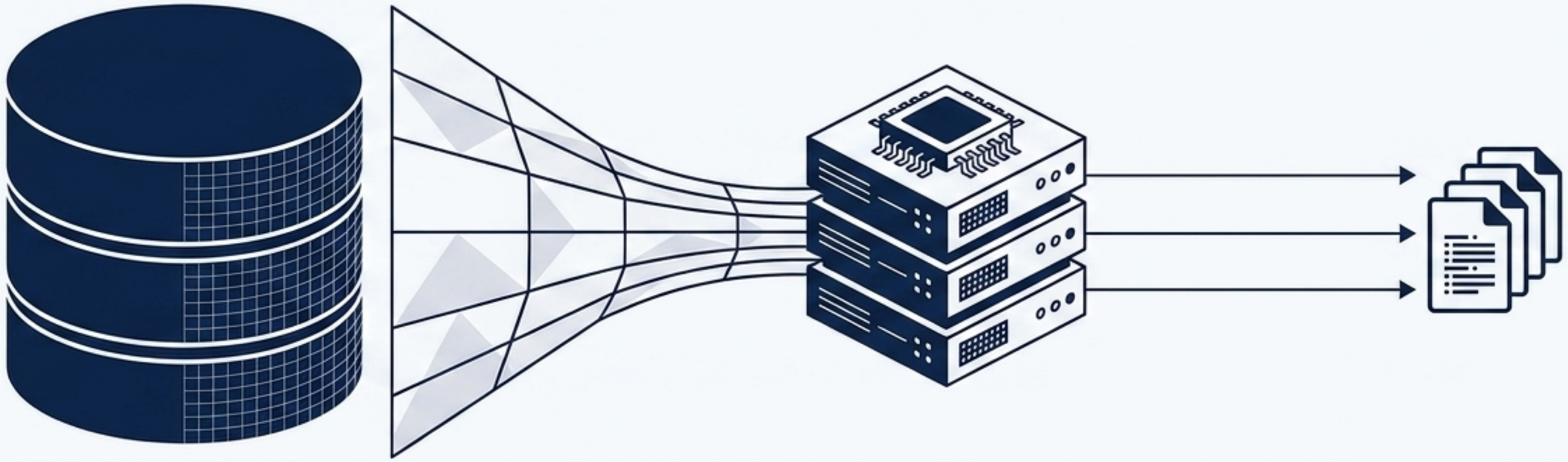
News impact asymmetry.

↓

Design Choice: Instrument-specific theme filtration.

The Academic Consensus: Announcement-driven drift takes **30 minutes to 4 hours to fully propagate**. Spreads widen **3-10x** during the initial window. This informs our **H1** bar resolution and **h=24** forward horizon.

Compressing the **Global Knowledge Graph**



Stage 1:
Raw Acquisition (GDELT Project)

1.3 TB unfiltered archive.
15-minute update frequency.
65+ languages spanning 1,663 days.

Stage 2:
The Filter Engine (6 Parallel Workers)

Download speed: 0.05 days/sec (~14 hrs total).
Applied strict Theme/Geo filters:

```
XAUUSD: ECON_GOLDPRICE, MINING (US, IN, CN)
USDJPY: ECON_INTEREST_RATE, TAX_FNCACT (JA, US)
```

Stage 3:
Parquet Output

36.2 GB Filtered Cache (1,367 files).
Final output: H1 Parquet files yielding 55,873 hourly bars (~2MB/file).

ExpServer Architecture and Execution Stack

Environment: ExpServer 10.10.10.100 | **Directory:** C:\workspace\fx_ml

Module Block 1 (Data Build)
build_gdelt_data.py → src/features/gdelt.py
Feature Engineering & Transformation

Module Block 2 (Dispatch) uso385_gdelt_news_sweep.py
Hyperparameter Grid Management (Manages the 72-config queue)

Module Block 3 (Runner) run_gdelt_news_experiment.py
Model Training & Evaluation (Handles Walk-Forward Validation)

Module Block 4 (Output) **results/gdelt385_*.json → uso385_gdelt_news_report.json**
Performance Metrics Aggregation Report Generation Artefact Storage

Engineering NLP Signals from the News Cycle

Context Panel: Extending enhanced65 to enhanced80 feature set via H1 integration.

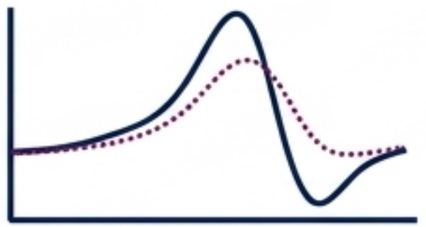
gdelt_tone_rolling

24h trailing average tone capturing persistent bearish/bullish media bias.

gdelt_volume_spike

Article count Z-score vs 7-day rolling mean. Abnormal media attention preceding macro moves.

gdelt_sentiment_momentum



6h tone first difference. Tracks sentiment acceleration.

gdelt_neg_pos_ratio

Absolute negative/total ratio acting as a raw fear gauge.

gdelt_geo_relevance

Tone standard deviation measuring controversy and conflicting narratives.

Quantifying Economic Calendar Events

Pre-Event Features

`event_countdown`: Hours to scheduled release. Tracks pre-event positioning and vol compression.

Event-Trigger Features

`surprise_magnitude`: Actual minus consensus. The primary driver of post-event drift.

`high_impact_flag`: Binary filter isolating signal events from calendar noise.

Post-Event Context Features

`vol_regime`: Volatility context for dynamic position sizing.

`consecutive_miss`: Streak of consensus misses identifying trending fundamental forecast errors.

`forecast_delta`: Consensus revision momentum.

Signal Logic Framework and Feature Importance

Tier 1: Directional Drivers Directional Drivers (High Signal)

`surprise_magnitude`, `gdelt_tone_rolling`, `gdelt_sentiment_momentum`

Function: Determines the vector (UP/DOWN) of the trade prediction.

Tier 2: Volatility Context (Position Modifiers)

`event_countdown`, `gdelt_volume_spike`, `vol_regime`

Function: Identifies when a move is likely to occur and with what intensity.

Tier 3: Noise Filters (Classification Boundaries)

`gdelt_geo_relevance`, `cross_event_corr`, `high_impact_flag`

Function: Helps tree algorithms segment out false-positive environments.

Categorical Advantage. ML trees natively exploit these sparse event flags without complex one-hot encoding.

The 72-Configuration USO-385 Sweep Matrix

4 Instruments: XAUUSD, GBPUSD, USDJPY, USDCAD

2 Models: CatBoost vs. XGBoost

3 Horizons: h=6, h=12, h=24

3 Thresholds: 8 bps, 10 bps, 15 bps

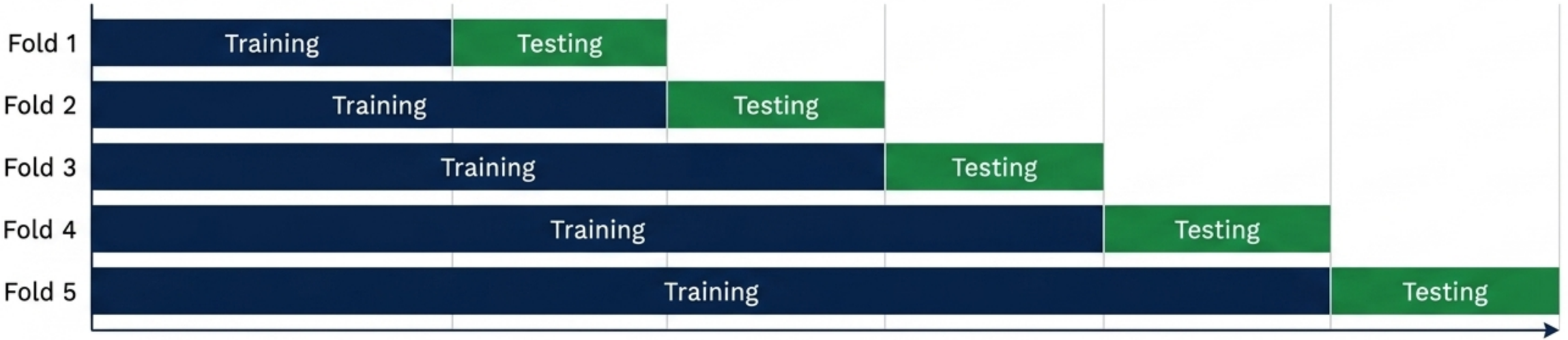
Prediction Task

Binary classification — Will price move UP or DOWN by at least the threshold bps within the h hour window?

Total Output: 72 unique machine learning configurations.

Preventing Look-Ahead Bias via Walk-Forward Validation

Methodology: 5-Fold Expanding Window



The Mechanics

Training: Minimum 2 years of history, continually expanding forward to capture regime shifts.

Testing: Strict 6-month out-of-sample forward
Testing: Strict 6-month out-of-sample forward steps.

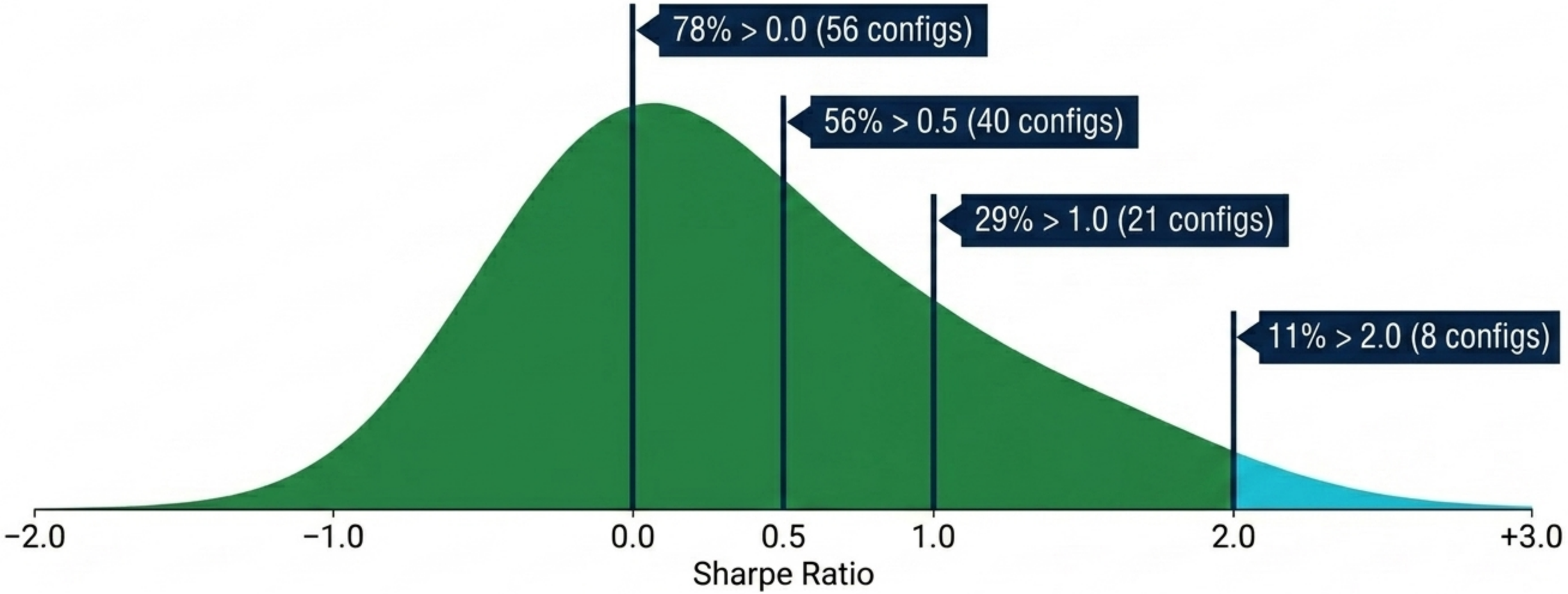
Strategic Rationale

Mimics true production deployment where models retrain on growing historical data and predict blindly into the future.

Eliminates forecast leakage and strictly limits calendar features to pre-release consensus data.

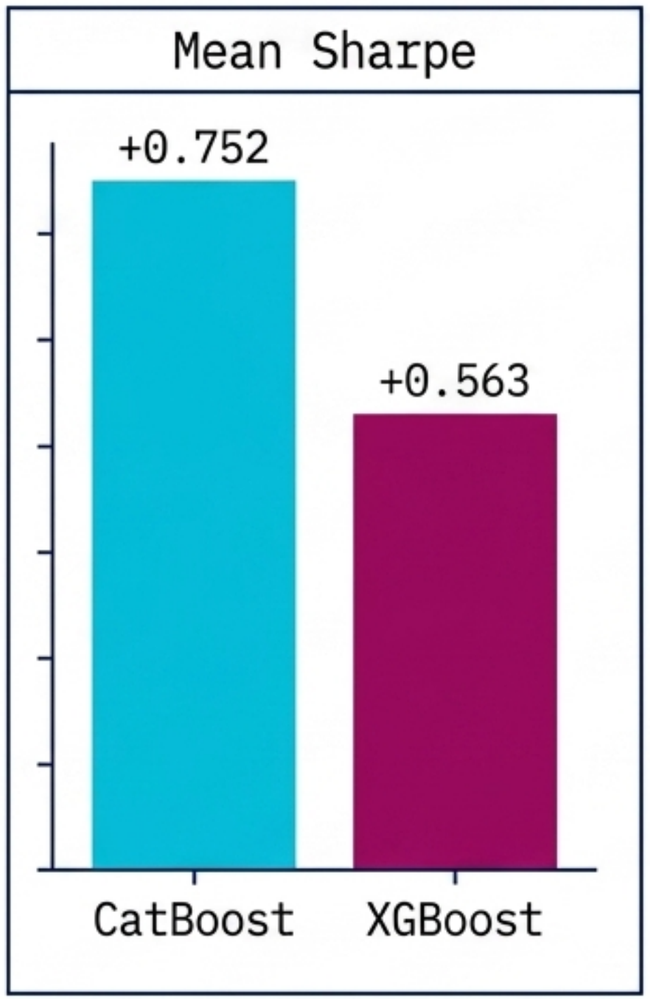
Sweep Results: Widespread Positive Alpha Generation

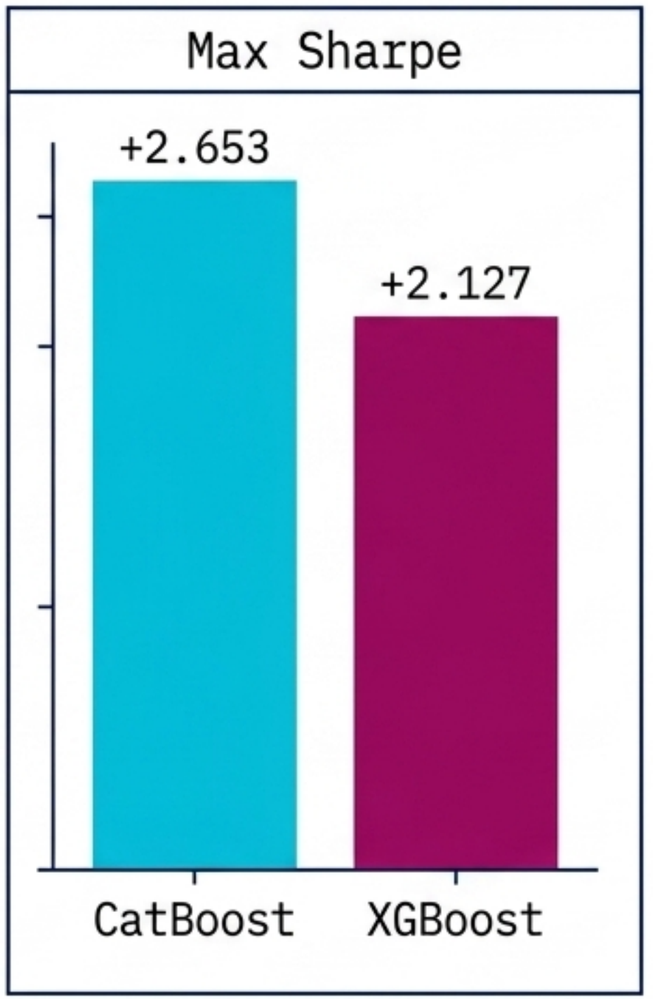
Completion Metrics:
72/72 configurations successfully executed (100% completion rate).



Summary Stats: Mean Sharpe: +0.657 | Max Sharpe: +2.653 | Min Sharpe: -1.172

ML Architecture Battle: CatBoost vs. XGBoost





The Categorical Advantage

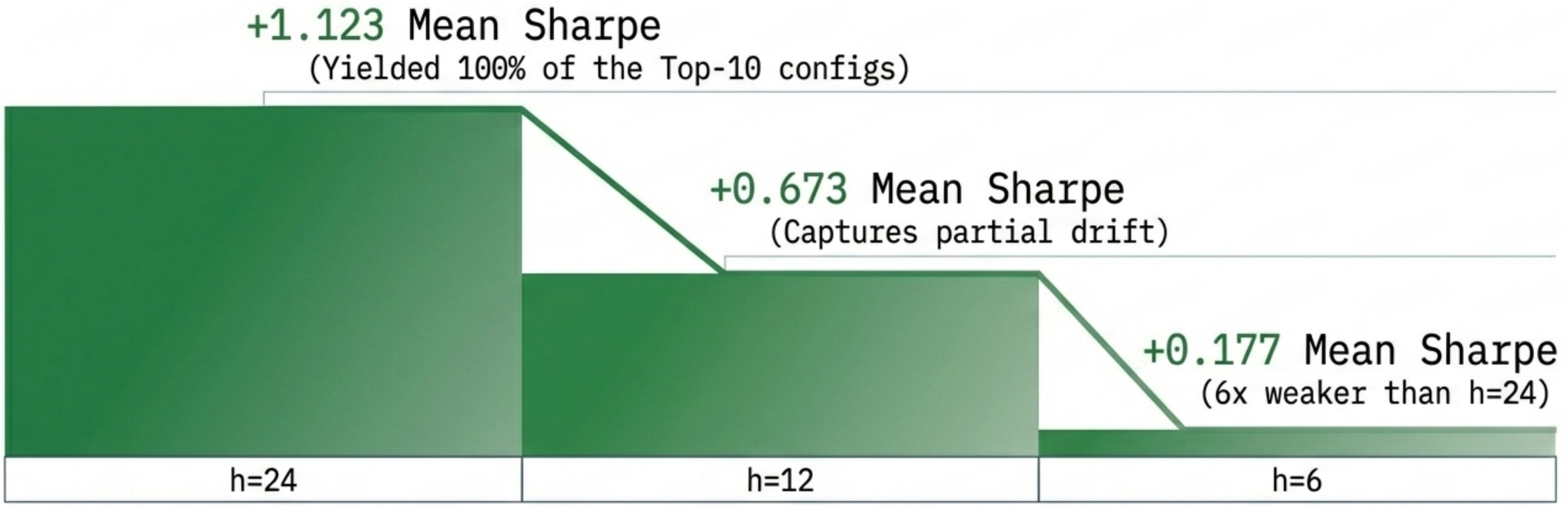
CatBoost delivers a +33% average performance edge over XGBoost.

Hypothesis Validated

News events and calendar releases are highly sparse, categorical features.

CatBoost's native categorical handling algorithm systematically exploits this sparsity more effectively than XGBoost's numeric-centric trees.

Horizon Decay: The Mathematics of News Drift



Market Microstructure Alignment

The dominance of h=24 strictly confirms the Lucca & Moench / Andersen thesis.

News-driven macroeconomic drift requires 30 minutes to 4+ hours to fully propagate through FX order books. H1 bars with a 24-hour target window capture this asymmetric drift optimally.

Instrument Sensitivity Profiles

1. USDJPY (Max Signal)	Best Sharpe +2.653	Claimed 7 of the top 10 configurations. Highly sensitive to BoJ policy surprises and explicit macro news events.
2. GBPUSD (Strong Signal)	Best Sharpe +2.127	Headline-driven behavior confirming the legacy Brexit/BoE sensitivity thesis.
3. XAUUSD (Strong Signal)	Best Sharpe +2.015	Tight coupling with global inflation narratives and US macro data provides clean directional signal.
4. USDCAD (Weak Signal)	Best Sharpe +0.843	General news is uninformative for the oil-CAD nexus; requires specialized oil-specific NLP feeds to extract edge.

Directional Threshold Analysis: Signal vs. Noise

The Top-End Filter

Produced the absolute maximum Sharpe (+2.653).

Forces the model to ignore micro-volatility and only predict high-conviction macro regime shifts.

The Sweet Spot

Produced highly stable, dense trade counts (8,180 - 11,105 trades).

Excellent balance of hit rate (~55%) and frequency.

The Noise Floor

Susceptible to false positives on weaker instruments (like USDCAD).

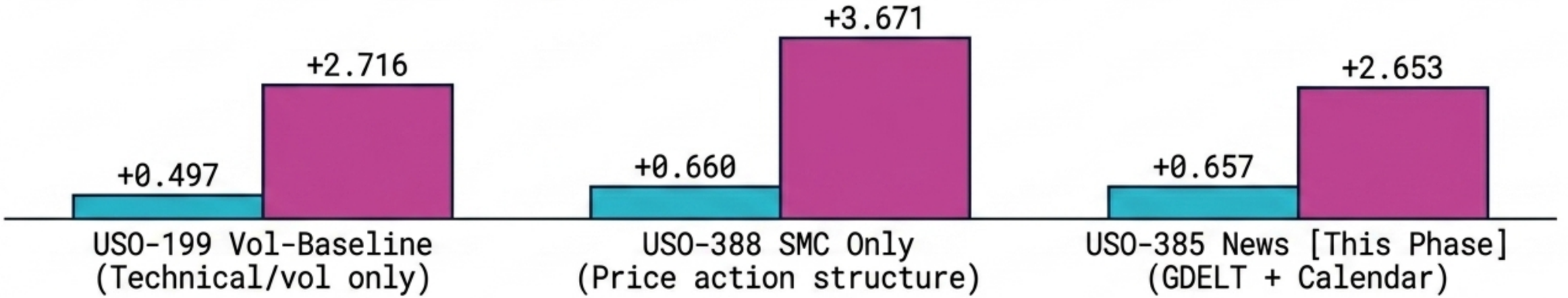
Works well strictly on highly volatile pairs where 8bps is easily cleared during routine drift.

The Top Performing Signal Configurations

Rank	Symbol	Model	Horizon	Threshold	Sharpe	Hit Rate	Profit Factor
#1	USDJPY	CatBoost	h=24	15 bps	+2.653	55.9%	1.40
#2	USDJPY	CatBoost	h=24	10 bps	+2.397	55.5%	1.28
#3	GBPUSD	XGBoost	h=24	8 bps	+2.127	53.2%	1.20
#8	XAUUSD	CatBoost	h=24	10 bps	+2.015	51.8%	1.19

Insight Note: Top models command healthy asymmetry, achieving 1.19 - 1.40 Profit Factors with mid-50s hit rates.

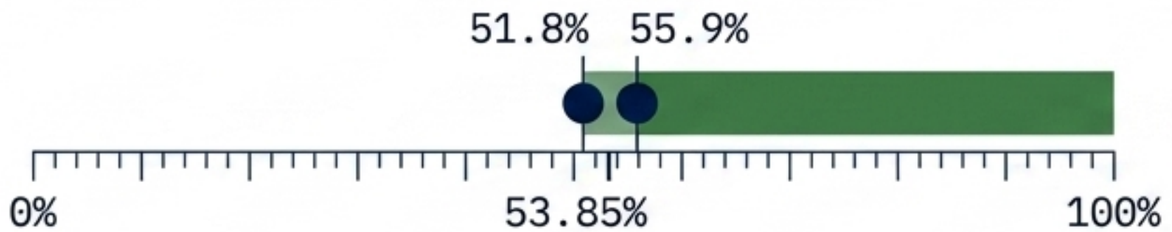
Edge Comparison: News vs. Market Structure



Synthesis

- News features generate a +32% systematic improvement over the vol-adjusted baseline.
- News mathematically matches the mean performance of pure market structure features.
- Because they rely on entirely orthogonal datasets (text/events vs. price action), combining them offers a massive ensemble opportunity.

Risk Diagnostics: Hit Rate & Asymmetry

Hit Rate Dependency	Profit Factor Profile
	
Top 10 models rely on a tightly clustered 51.8% to 55.9% hit rate. Risk: This represents a mathematically meaningful edge over thousands of trades, but is susceptible to short-term variance and flat periods.	PFs range from 1.18 to 1.40 for top configs. Mitigation: The models exhibit strong asymmetry. Even if hit rates decay toward 50%, the size of the winning directional drifts protects the equity curve.

Structural Market Risks and Mitigations

1. Spread Widening (Transaction Costs)	⚠ **Risk:** Spreads widen 3-10x during NFP/CPI announcement windows.	🛡 **Mitigation:** Phase 3 implementation of dynamic spread penalty modeling.
2. Regime Inversion (Macro Crises)	⚠ **Risk:** Models overfit to a specific macro narrative (e.g., rate-hike cycles).	🛡 **Mitigation:** Mitigated by the 5-fold expanding window validation and the dynamic `vol_regime` feature acting as a state-switch.
3. GDELT Publication Lag	⚠ **Risk:** 15-minute raw data update cycle.	🛡 **Mitigation:** H1 bar resolution safely absorbs this lag, preventing intra-bar look-ahead execution failures.

Deployment Roadmap and Next Steps

Phase 1: QC Validation

Independent out-of-sample verification of USDJPY CatBoost & GBPUSD XGBoost candidate models.

Phase 2: Signal Architecture

Ensemble: Merge US0-385 News features with US0-388 SMC features.

Ablation: Isolate GDELT vs. Calendar edge impact.

Phase 3: FinBERT Upgrade

Replace generic tone with GPU-based FinBERT headline NLP scoring (Expected +5-12% edge increase).

Phase 4: Live Deployment

Promote validated, cost-adjusted configurations to the paper trading server environment.