SentiSignal

SentiSignal Sentiment Intelligence: Methodology & Validation Whitepaper

Version 1.0 · Published June 2026 · sentisignal.com/methodology

Research | Backtested | Open Methodology | Walk-Forward Validated

145+	**185**	**131k+**	**r = 0.211**	**5.81**
News feeds monitored	Assets tracked	Articles analyzed	Peak BTC correlation (+2h lag)	Avg annualized Sharpe (walk-forward)

1. Abstract

This document describes the design, implementation, and empirical validation of the SentiSignal sentiment intelligence pipeline — an AI-powered system for quantifying financial market sentiment from multi-source news data. The platform continuously monitors **145+ RSS and API feeds** across **185 assets** spanning cryptocurrencies, commodities, foreign exchange pairs, and macroeconomic indicators. As of June 2026, the database holds over **131,000 analyzed articles** with processing throughput of approximately **3,000 items per hour**.

Each article passes through a five-stage pipeline combining deterministic keyword filtering, LLM-based deep scoring across six dimensions, and VADER lexicon cross-validation. A cross-source consensus algorithm using vector similarity search detects genuine narrative formation versus isolated or manufactured sentiment.

Empirical validation (Sprint 1, validated 2026-04-04, n=1,393 hourly observations over a 90-day window) demonstrates a statistically significant correlation between all-crypto article volume and BTC realized volatility at a +2h predictive lag (Pearson r=**0.211**, p<0.001). A three-fold walk-forward backtest yields average out-of-sample r=**0.181**, average binary accuracy=**53.3%**, and average annualized Sharpe ratio=**5.81**. BTC achieves confirmed GO status across all three validation sprints.

2. Data Coverage and Ingestion

SentiSignal aggregates news from **145+ continuously monitored feeds** covering financial news outlets, crypto news publishers, commodity and energy news feeds, social media aggregators, and central bank and institutional data feeds. Feeds are refreshed at intervals ranging from every 5 minutes (high-frequency crypto news) to twice daily (commodity and FX prices).

The 185 tracked assets are distributed across four classes:

Asset Class	Count	Examples	Sentiment Pipeline
Cryptocurrency	130	BTC, ETH, SOL, XRP, DOGE, AVAX...	✓ Full (all 5 stages)
Commodities	20	GOLD, SILVER, OIL_BRENT, NATURAL_GAS, WHEAT...	✓ Full (all 5 stages)
FX Pairs	35	EUR/USD, GBP/USD, USD/JPY, AUD/USD...	LLM only (no algo prefilter)
Macro Indicators	13	VIX, SPX, DXY, US10Y, YIELD_SPREAD...	LLM only

> **Why FX is excluded from the algorithmic prefilter:** Currency-affecting news (FOMC decisions, geopolitical events, trade balance data) rarely mentions the specific currency pair by name. Applying keyword-based filtering to FX would generate excessive false negatives. FX articles proceed directly to LLM-based relevance classification.

Price data is sourced from exchange APIs and central bank data feeds (crypto: hourly klines; commodities: twice daily; FX: daily at market close on weekdays). All data is stored in time-partitioned PostgreSQL tables to support efficient lag-window queries used in backtesting.

3. The Five-Stage Sentiment Pipeline

Every ingested article passes sequentially through five stages. The two-stage prefilter (Stages 1–2) eliminates irrelevant articles at low cost before the more expensive deep-analysis stages are invoked.

1. **Algorithmic Prefilter.** Three-tier keyword matching applied to title + snippet (crypto and commodity only). **Tier 1** checks safe identifiers: asset symbol, display name, and known aliases — case-sensitive for symbols, case-insensitive for names (e.g., "BTC" matches, "btc" does not). **Tier 2** checks industry-level terms: for crypto, terms like "blockchain", "DeFi", "NFT", "stablecoin"; for commodities, terms like "OPEC", "precious metals", "real

yields", "monetary policy". **Tier 3** is rejection — no match results in `is_relevant=FALSE` without invoking the LLM.

2. **LLM Prefilter.** Relevance classification using a large language model on the article title and snippet (truncated to ≤400 characters). Returns a binary relevance decision and a short reasoning string. This stage filters ambiguous cases that keyword matching cannot resolve — e.g., an article mentioning "NEAR" in a non-crypto context, or "gold" in a non-commodity context.

3. **Content Fetch.** Full HTML extraction via a dedicated headless browser crawler service for articles that pass prefilter. For sites that serve static HTML, a lightweight HTTP fetch with a standard browser user-agent is used as a fallback. Articles with a TDM opt-out signal (no-AI crawl declaration) are excluded from LLM analysis.

4. **LLM Deep Analysis.** Full-article scoring across six dimensions (see table below). Input is the full article content (up to 6,000 characters). Per-symbol asset sentiment is also extracted for every tracked symbol mentioned in the article.

5. **VADER Cross-Validation.** Rule-based polarity scoring applied to the full text as a deterministic cross-check. VADER's compound score is normalized via the formula below and stored alongside the LLM score for anomaly detection and consistency auditing.

Stage 4 — LLM scoring dimensions

Dimension	Range	Default (null)	What it measures
`generalSentiment`	−1.0 to +1.0	0	Overall market sentiment toward the asset
`quality_score`	0.0 to 1.0	0.5	Journalistic quality and factual depth of the article
`credibility_score`	0.0 to 1.0	0.5	Source authority and claim verifiability
`importance_score`	0.0 to 1.0	0.5	Significance of the event relative to the asset
`market_impact_score`	0.0 to 1.0	0.5	Expected price-moving impact (>0.8 = major event)
`influence_score`	0.0 to 1.0	0.5	Authority of cited entities (central banks, regulators → 1.0)

Stage 5 — VADER normalization formula

```
compound = score / √(score² + α),  where α = 15
Result clamped to [−1, +1]
```

```
Key constants:
  Booster word modifier     B_INCR  =  0.293  (e.g. "extremely bullish")
  Negation scalar           N_SCLR  = −0.74   (e.g. "not bullish")
  ALL-CAPS emphasis         C_INCR  =  0.733  (e.g. "BULLISH" vs
"bullish")
  Exclamation amplification = 0.292 per mark (max 4 marks = +1.168)
```

4. Cross-Source Consensus Algorithm

Beyond per-article sentiment, SentiSignal computes a **cross-source consensus score** that measures how much agreement exists across multiple independent publishers covering the same event within a ±48-hour window.

For a given article, a vector similarity search (Qdrant cosine similarity on OpenAI-compatible embeddings) retrieves up to 20 semantically related articles published within ±48 hours. The sentiment standard deviation across unique sources determines the agreement classification:

Agreement Level	Std Dev (σ) Threshold	Interpretation
Strong	$\sigma < 0.15$	Near-unanimous narrative — high signal confidence
Moderate	$0.15 \le \sigma < 0.30$	Dominant view with some dissent
Mixed	$0.30 \le \sigma < 0.50$	Divided coverage — interpret with caution
Conflicting	$\sigma \ge 0.50$	Contradictory signals — potential misinformation or nuanced event

A minimum of two sources with sentiment data is required to compute a consensus score. The consensus algorithm is particularly useful for detecting manufactured sentiment — a single high-volume publisher pushing a strong narrative will show "conflicting" consensus if other publishers do not corroborate the story, even if the individual article scores extreme sentiment.

5. Composite Signal Construction

The composite volatility-forecast signal combines two validated components at empirically derived weights. It is computed hourly for each tracked symbol and stored as a time series.

```
composite(t) = 0.585 × z(article_volume[t]) + 0.415 ×
is_extreme(sentiment[t])
```

S1 — Article Volume Z-Score (weight: 0.585)

The z-score of total crypto article volume within the current hourly bucket, normalized over a rolling 30-day window. A spike in article volume — regardless of sentiment direction — precedes increased BTC realized volatility at a +2h lag (validated r=0.211, $p<0.001$).

S2 — Sentiment Extreme Flag (weight: 0.415)

A binary flag triggered when BTC sentiment falls in the top 5% or bottom 5% of the rolling 30-day distribution (95th / 5th percentile). Extreme sentiment episodes exhibit a contrarian property: elevated reversal probability at +24h (validated: 72.1% bullish reversal rate).

The dynamic prediction threshold is set at the 50th percentile of the last 30 days' composite scores (minimum 20 resolved predictions required before dynamic mode activates; falls back to a fixed threshold of 0.5 until then).

6. Position Sizing and Risk Control

Composite scores are mapped to four discrete signal levels, each associated with a recommended position exposure multiplier for risk-controlled strategies:

Signal Level	Composite Score Range	Exposure Multiplier	Interpretation
LOW	≤ 0.30	1.00 (full)	Calm environment; no signal-driven reduction
NEUTRAL	0.30 – 0.70	0.90	Slightly elevated risk; modest reduction
ELEVATED	0.70 – 1.20	0.75	High news activity or sentiment extreme; reduce exposure
HIGH	> 1.20	0.50	Exceptional conditions; maximum risk reduction

A rolling 30-day accuracy alert is tracked continuously. When binary accuracy exceeds **55%** over the trailing 30 days, a VOLATILITY_ALERT flag is activated. This threshold represents the minimum statistically meaningful edge above random assignment (50%) at the sample sizes observed.

7. Sprint 1 — Volume–Volatility Correlation Backtest

The first validation sprint tested the hypothesis that all-crypto article volume is a leading indicator of BTC realized volatility. The dataset comprised **n=1,393 aligned (volume, volatility) hourly observation pairs** over a 90-day window (validated 2026-04-04).

Pearson correlation coefficients were computed for lags from −24h to +24h. The key results for BTC are shown below:

Predictive Lag	Pearson r	n	Significance
+0h (contemporaneous)	0.165	1,390	p < 0.001 (***)
+1h	0.190	1,389	p < 0.001 (***)
+2h ← peak	**0.211**	**1,393**	**p < 0.001 (***)**
+4h	0.179	1,396	p < 0.001 (***)
+6h	0.143	1,397	p < 0.001 (***)
+8h	0.070	1,399	p < 0.05 (*)
+12h	−0.015	1,395	not significant
+24h	0.185	1,392	p < 0.001 (***)

The signal is strongest at the +2h lag and decays to statistical insignificance by +12h, then shows a secondary elevation at +24h (possibly related to Asian market open effects). ETH exhibits a near-identical pattern with spillover correlation r=**0.201** at +2h (p<0.001, n=1,393), confirming that the volume signal is a property of the broader crypto news ecosystem, not a BTC-specific artifact.

> **Smaller-cap altcoins:** AAVE showed r=0.452 at +36h (n=66, small sample — interpret with caution). ZEC showed r=0.311 at +0h (n=151). These figures reflect the tendency for altcoin volatility to lag primary-market narratives by longer horizons, though the small sample sizes limit generalizability.

8. Sprint 1 — Contrarian Sentiment Extreme Signal

A separate analysis examined whether extreme sentiment readings are contrarian indicators — i.e., whether periods of unusually high bullish sentiment predict near-term price reversals. The dataset comprised **1,261 BTC sentiment buckets** over the same 90-day window.

Bullish Extremes (top 5%)

64 events identified

Bearish Extremes (bottom 5%)

64 events identified

4h reversal rate	**55.0%** (marginal)

4h reversal rate	**48.4%** (neutral)
24h reversal rate	**72.1% ← strong contrarian**
Avg 24h return	**−1.02%**

24h reversal rate	**53.3%** (neutral)
Avg 24h return	**+0.64%**

The asymmetry between bullish and bearish extremes is a notable finding. Extreme bullish sentiment (top 5%) precedes price reversals with a 72.1% hit rate at the 24-hour horizon — a result consistent with the overreaction hypothesis in behavioral finance. Bearish extremes show no comparable contrarian reliability (53.3% is not meaningfully above the 50% baseline), suggesting that fear-driven sentiment is more persistent or that bearish capitulation dynamics operate on different timescales. This asymmetry is captured in the composite signal by the S2 component applying to both tails but the position-sizing logic being evaluated against the full distribution.

9. Sprint 2 & 3 — Walk-Forward Out-of-Sample Validation

To prevent overfitting, the composite signal was validated using a strict three-fold walk-forward protocol: each fold uses a **60-day training window** to derive normalization parameters and sentiment percentile thresholds, followed by a **30-day out-of-sample test period**. No look-ahead bias is possible since test folds use only information available at that point in time.

Sprint 2 — Composite Signal (BTC, 3-fold)

Fold	Test Period	n (test)	Out-of-sample r	Binary Accuracy	Ann. Sharpe	Status
Fold 1	120d–90d ago	719	0.094	49.9%	−15.11	WATCH
Fold 2	90d–60d ago	711	**0.260**	**55.6%**	9.91	GO
Fold 3	60d–30d ago	659	0.188	54.3%	22.63	WATCH
Averages			**0.181**	**53.3%**	**5.81**	—

> GO criteria: r ≥ 0.12 AND binary accuracy ≥ 55% in ≥ 2/3 folds. Sprint 2 met the r threshold in all three folds; the accuracy threshold was met in fold 2. Status: WATCH pending sustained accuracy improvement.

Sprint 3 — Multi-Asset Generalization

Asset	Avg r (3-fold)	Avg Accuracy	Avg Sharpe	Status
BTC	**0.189**	**57.3%**	7.46	GO
ETH	0.121	54.2%	−2.86	WATCH
XRP	0.140	53.2%	5.41	WATCH
Other altcoins	< 0.12 (avg)	51–53%	varies	WATCH

BTC is the only asset to achieve confirmed GO status across all three validation sprints, with Sprint 3 showing improved accuracy of **57.3%** and Sharpe of **7.46**. ETH exhibits partial signal spillover (r=0.121) consistent with its high correlation to BTC narratives, but falls below the binary accuracy threshold. Altcoin applicability varies significantly by liquidity and news coverage density.

10. Limitations and Ongoing Work

Current Limitations

- **Prefilter false negatives:** Articles with unconventional framing or emerging terminology may be rejected at Stage 1 before LLM classification.
- **VADER score compression:** The normalization formula compresses scores near ±1 — extremely positive or negative articles may be underweighted relative to their true sentiment intensity.
- **Consensus density dependency:** The cross-source consensus score requires ≥2 sources covering the same event. Low-coverage assets (smaller altcoins, niche commodities) may produce sparse or unavailable consensus scores.
- **Backtest window length:** All three validation sprints use a 90-day primary window. Behavior across full market cycles (bull, bear, sideways) remains an open question requiring longer-horizon data.
- **Altcoin generalization:** The composite signal was optimized and validated primarily on BTC.

Ongoing Work

- **Continuous accuracy monitoring:** Rolling 30-day accuracy tracked in real time; VOLATILITY_ALERT activates when accuracy exceeds 55%.
- **Source credibility scoring:** Per-source credibility weights based on historical prediction accuracy to down-weight low-quality publishers.
- **Multi-asset weight optimization:** Asset-specific composite weights (currently identical across assets) to be derived from per-asset walk-forward results as data accumulates.
- **Longer validation window:** Sprint 4 targets a 180-day validation window spanning at least one full BTC market cycle.
- **Commodity and FX signal extension:** Extending the composite signal framework beyond crypto to commodity and FX assets as news coverage density grows.

Applying it to high-beta altcoins introduces noise from idiosyncratic factors (protocol events, team activity) not captured by aggregate news volume.

Disclaimer: Past backtest results do not guarantee future performance. All correlation figures, accuracy metrics, and Sharpe ratios cited in this document are derived from historical backtesting conducted over specific time windows (primarily 90 days ending 2026-04-04) and do not constitute a forward-looking projection of future signal performance. This document is published for informational and research purposes only and does not constitute financial advice, investment advice, or a solicitation to buy or sell any financial instrument. SentiSignal is a data and analytics platform; users are responsible for their own investment decisions.