newsquawk

Quantitative News Dataset (Real-Time + Historical)

This is not a news feed. It is a pre-filtered dataset built for signal extraction.

Noise, duplication, and low-value content are removed before ingestion. What remains is a structured record of market-moving events, classified using a hybrid of analyst judgement and machine processing developed over two decades in live markets.

What you don't see in the dataset is as important as what you do.

Curation and Signal Quality

- Pre-filtered dataset built for signal extraction, not aggregation
- Noise, duplication, and low value content removed before ingestion
- Time consistent classification of importance, testable against market impact
- Hybrid human + machine curation developed over two decades in live markets
- Focus on material, market-moving events rather than headline volume

Coverage and market Scope

- Global macro: central banks, economic releases, policy shifts
- Corporate events: earnings, filings, M&A, disclosures
- Regulatory and exchange-driven updates
- Unscheduled and fast-breaking developments
- Particular strength in events where aggregated feeds lag or overproduce noise
- Verified social media included as a filtered signal layer

Data Structure & Backtesting Integrity

- Pre-classified, tagged, high-signal event dataset
- Time-consistent definition of "importance" across regimes
- No aggregation layer or post-processing required
- Structured for direct integration into quant workflows

Real-Time Delivery & Deployment

- Continuous stream of curated, tagged headlines via API
- Real-time (T+0) delivery with no delay
- Identical curation logic across historical and live data
- No regime shift between research and production environments
- Structured for direct integration into quant workflows
- Deploy without changing inputs