Use one canonical 4H bar timestamp (UTC or your trading tz). For each feature type, reduce/aggregate into that bar without leaking future info.

Price / OHLCV (XAU, futures, ETFs)

- *Open = first tick after bar start.*
- *High = max price within bar.*
- *Low = min price within bar.*
- *Close = last tick ≤ bar end.*
- *Volume = sum within bar.*

Technical indicators (RSI, EMA, MACD, ATR, Bollinger)

- *Compute on the 4H resampled price series (not per-tick then average). I.e., resample price first, then compute indicators.*

Volatility indices (VIX, MOVE)

- *Use 4H close or 4H mean. Keep both (close for trend, mean for intrabar activity).*

FX / DXY / yields / TIPS / rates

- *Use 4H close. For yields updated frequently, take last quoted value in the bar. For daily/weekly series, stamp at release time then forward-fill only until next release.*

Macro releases (CPI, PCE, Fed rates)

- *Create event features: release_flag (bar contains release), surprise = actual - consensus, lead_indicator = sign(surprise). Never forward-fill the true value into prior bars. For training, use release-time stamping and purge nearby labels to avoid leakage.*

Text / Sentiment (Fed speeches, news, social)

- *Collect all texts within the 4H window. Produce:*
  - *emb_mean = mean embedding (or weighted mean by source reliability).*
  - *sent_mean = average sentiment score.*
  - *shock_score = max(|sentiment|) in window.*
  - *topic_counts or topic_vector (optional).*
- *Compute these from texts whose timestamp ≤ bar end.*

COT, futures open interest, ETF flows, central bank purchases

- Typically lower frequency: forward-fill their latest reported value into

subsequent 4H bars until updated. Also add age_of_report as a feature.

Order-book / microstructure (spread, depth)

- *Aggregate snapshots taken at bar end or average several snapshots within bar: avg_spread, avg_depth_top5, liquidity_score = depth / typical_notional. Also compute spread_99p tail to detect spikes. For small account use liquidity_bucket (high/med/low).*

Time features

- hour_of_day, day_of_week, session_overlap (London/NY), is_holiday.

Alignment & timezone

- Use single timezone; align everything to bar end. For any series with a delayed publish (news wires), use the publish timestamp and stamp to the first bar after that time.