Complete Specification — Gold Trading Bot (full system)

Below is the **single, consolidated blueprint** of the entire bot we designed across this conversation. It preserves your original intent, is structured for implementation, and contains every operational detail you requested: data, agents, models, feature engineering, training/backtesting, decision logic, risk & portfolio rules, manual execution assistant, monitoring, and deployment notes.

---

1. One-line summary

A 4-hour decision-anchor gold trading system that learns from historical data (LightGBM + LSTM + small Transformer), discovers and validates lagged correlations, scores macro/sentiment/technical stacks, fuses them with a meta-model, then produces human-reviewed entry/exit suggestions; a Risk & Portfolio agent enforces safety and reports full performance metrics.

---

2. High-level architecture (agents & data flow)

[Historical Data Store] --> [Feature Builder] --> [Model Training / Backtest]

|

V

[Live Data Ingest] --> [Research Agent] --> [Feature Engine] --> [Analysis Agent (LightGBM)]

| ^

|-> [Transformer Sentiment] ->|

|-> [LSTM Sequence Encoder] ->|

|

[Meta LightGBM (fusion)]

|

[Decision (FinalProb)]

|

----------+----------

| |

[Risk & Portfolio Agent] [Trader Agent]

| |

Approve/adjust & size trades Prepare entry timing hints

| |

[LLM Reasoning] (approve/deny/notes)

|

[Execution / Alerts] -> You (manual execution)

|

[Order Monitor / Logs / PnL]

---

3. Data (full list you will collect — historical & live)

All items are aligned to 4-hour bars. Slow data is stamped at release and forward-filled only until update.

Macro & Rates

- 1Y, 5Y, 10Y TIPS yields (FRED DFII1/DFII5/DFII10)
- Nominal Treasury yields (1Y,2Y,5Y,10Y)
- Fed funds futures curve
- Fed funds rate / FOMC decision timestamps
- CPI (headline & core) MoM and YoY
- PCE (headline & core) MoM and YoY
- TIPS breakevens (5Y, 10Y)
- Yield curve spreads (10Y–2Y, 10Y–3M)

Dollar & FX

- DXY spot and futures
- USD/EUR, USD/JPY, USD/GBP, USD/CHF, USD/AUD, USD/CAD, USD/CNY
- FX volatility indices (as available)

Market Stress / Macro Sentiment

- VIX (CBOE)
- MOVE index (if available)
- Geopolitical Risk Index (GPR) or equivalent
- Central bank gold purchases (WGC monthly/quarterly)
- Gold ETF flows (GLD, IAU, SGOL) – daily AUM/flows
- Physical demand proxies (China SGE premium, India import data)
- Gold futures open interest; COT net positioning (commercials/speculators)

Price & Technical

- XAU/USD spot & futures OHLCV (4H)
- Silver price (for gold/silver ratio)
- RSI (4H), EMA50/EMA200 (4H), MACD (4H), Bollinger Bands (4H), ATR (4H)
- MomentumFactor, VolFactor, TrendState (derived)

Text & Sentiment

- FOMC minutes, Fed speeches (full text)
- News headlines (Reuters, Bloomberg, WSJ — vectorized)
- Social media streams (Twitter/X, Reddit substreams) — vectorized / aggregate
- Google Trends queries (optional)

Other / Alpha Boosters (optional)

- Weather indicators affecting mining regions

- Shipping indices (Baltic Dry)

- Mining production/supply reports

---

4. Feature engineering (per 4H bar)

- Clean & align timestamps. Winsorize extreme tails (1%/99%). Z-score rolling scaling.

- Δ (changes) for rates/indices to reduce non-stationarity.

- MacroFactor PCA: compress correlated macro inputs into MacroFactor■ (real rates), MacroFactor■ (USD), MacroFactor■ (inflation).

- Sentiment vectors → SentScore (mean), ShockScore (magnitude percentiles), TopicVec (8–16 dims) with quality weighting (EWMA of source accuracy).

- Technical compressions → MomentumFactor (weighted z of RSI, MACD hist, EMA slope), VolFactor (z of ATR, BB width), TrendState (EMA50>EMA200 binary).

- Lag map creation: cross-correlation (CCF), Granger tests, mutual information to discover candidate lags (±12 bars). Store accepted lags.

---

5. Model roles & training plan

Primary models

1. **Macro LightGBM** — trained on snapshot tabular features (latest MacroFactors, lagged macro features from lag map) → outputs `MacroScore` (prob up) + SHAP importance.

- Use purged time series CV; use class/probability target (next N 4H returns or direction).

2. **Transformer (small encoder)** — inputs: texts per 4H window → outputs `SentScore`, `ShockScore`, `TopicVec`.

- Pretrained embeddings (e.g., sentence-transformers), then fine-tune a small head.

3. **LSTM (sequence encoder)** — inputs: sequences (K last 4H bars) of curated features (MacroFactors, SentScore, ΔDXY, VIX, Momentum/Vol) → outputs `SeqContext` embedding (16–64 dims).

- Train to predict next-N returns or as autoencoder objective for regime capture.

4. **Meta LightGBM** — inputs: `MacroScore`, `SentScore`/`Shock`, `SeqContext`, MomentumFactor, VolFactor, TrendState → outputs `FinalProb` (main decision probability).

Training order

- Build lag map & prune features by significance (FDR control).

- Train Macro LightGBM (feature selection + SHAP).

- Train Transformer on labeled events (news→price moves) to get SentScore.

- Train LSTM on curated inputs to produce SeqContext.

- Train Meta LightGBM on outputs of above plus technicals.

- Validate via purged, embargoed walk-forward CV.

---

6. Backtesting & validation rules

- Use **purged & embargoed** cross-validation to prevent label leakage.
- Walk-forward: e.g., train 12 months → test 3 months, slide forward.
- Include **realistic execution**: bid/ask spread, commission, latency slippage model.
- Performance metrics to compute:
- CAGR, Total return
- Sharpe, Sortino
- Max Drawdown (MDD) and drawdown duration
- CVaR (expected shortfall) at 95%/99%
- Hit rate, avg win/loss, profit factor, turnover
- R distribution, median R, avg R
- Stability checks: rolling SHAP / PSI (population stability index); regime PnL splits (high/low VIX).

---

7. Decision & trade logic (human-executed final)

Decision anchor: 4H bar close → models produce `FinalProb`.

Thresholds (example, tune via backtest):

- BUY if `FinalProb >= 0.60` and TrendState = 1 and MomentumFactor ≥ 0
- SELL if `FinalProb <= 0.40` and TrendState = 0 and MomentumFactor ≤ 0
- HOLD otherwise

Entry timing assistance (Trader Agent) — You execute manually

- Bot returns:
- Suggested entry zone (limit range) based on 4H close + 15m/5m micro confirmation (e.g., pullback to EMA20 on 15m)
- Stop-loss level (distance defined by volatility rule)
- Take-profit level (R:R minimum 2:1; or dynamic TP based on historical move targets)
- Position size suggestion (see risk rules)
- Confidence & reasoning summary (top SHAP features + sentiment summary)
- You receive the structured alert (see Alert Template below), then confirm manual execution.

Exit assistance

- Suggestions: hard SL/TP, trail stop suggestions, time stop if no progress in N bars, early exit if `FinalProb` falls under safety threshold or if LLM reasoning flags high conflict.

---

8. Risk & portfolio agent — full metrics (Bot B, comprehensive)

Core (must implement first)

- Maximum Drawdown (MDD)
- Volatility (standard deviation)
- VaR (Historical / Parametric at 95% & 99%)

- CVaR (Expected Shortfall)

- Sharpe Ratio

- Sortino Ratio

- Beta vs relevant benchmarks (DXY, S&P;, 10Y yields)

- Correlation matrix (full covariance) between major drivers and portfolio returns

- Position sizing (risk per trade e.g., 0.5–1% adaptive)

- Exposure caps (max net exposure to gold factor, max leverage)

Additional advanced

- Treynor ratio, Calmar ratio, Information ratio

- Ulcer Index, Omega Ratio

- Tail risk measures, stress test scenarios, liquidity risk (depth/slippage)

- Portfolio-level volatility targeting and risk parity adjustments

---

9. Risk rules & defaults (practical rules you'll run live)

- Max risk per trade: **0.75%** of account (start); reduce to 0.4% when VIX > 25 or during major events.

- Minimum reward:risk for entry: **2:1**.

- Hard stop: set at `max(1.5 × recent volatility distance, prior swing low/high)`.

- Trail the stop once position reaches 1.5R (move to breakeven + small buffer).

- Max concurrent exposure: cap net gold delta to a percent of account (e.g., 3–5%).

- Circuit breaker: pause new trades if daily loss > 2% or running drawdown > 8%.

- Retraining cadence: sentiment encoder continuous; Macro & meta LightGBM monthly; LSTM quarterly (or after regime break).

---

10. LLM reasoning layer (role & guardrails)

- Purpose: structured sanity check & clear text rationale. NOT autonomous order placement.

- Input: proposed trade card + top SHAP features + current regime flags + recent historical analogs.

- Output: `approve/deny/adjust` + short rationale bullets + suggested adjustments to size/stop.

- Guardrails:

- LLM cannot directly place trades.

- All LLM decisions logged and auditable.

- If LLM repeatedly overrides model, flag to Research Agent.

---

11. Alert & signal template (what you'll see)

ALERT: BUY XAU/USD

Time: 2025-09-13 20:00 (4H close)

FinalProb: 0.72 (meta)

MacroScore: 0.65 (DXY down, real rates down)

SentScore: 0.55 (mild dovish tone)

SeqContext: bull-regime embedding id=42

Entry zone: 1950.5 - 1952.0 (limit recommended)

Stop-loss: 1935.0 (risk = 1.1% of account if full size)

Take-profit: 1979.0 (R:R ≈ 2.2)

Suggested position size: 0.9% risk (qty = X oz)

Reason: Top drivers — falling real rates, ETF inflows, low VIX relative

Micro-check: 15m pullback to EMA20 possible; wait 15m confirmation? Y/N

LLM note: approve with caution — near Fed speech in 12h; reduce size 30% if speech is hawkish.

---

12. Monitoring, logging & deployment

- Logs: every signal, model input snapshot, model outputs, SHAP values, LLM reasoning, risk approval / rejection.

- Dashboards: current positions, equity curve, drawdowns, rolling metrics (30/90/365 days), signal history.

- Alerts: missing feeds, model degradation (PSI > 0.2), abnormal slippage, exceeded exposure.

- Deployment: containerized services (microservices per agent) or one orchestrated serverless pipeline; versioned model artifacts and feature store.

---

13. Implementation checklist (priority order)

1. Build historical dataset & alignment scripts (4H resample, stamping, forward-fill rules).

2. Create feature engineering pipeline (macro PCA, sentiment embedding pipeline, technical factors).

3. Lag discovery process (CCF, Granger, mutual info) and lag map.

4. Train Macro LightGBM baseline; compute SHAP.

5. Train Transformer sentiment model (produce SentScore).

6. Train LSTM sequence encoder on curated features.

7. Train Meta LightGBM and validate using purged walk-forward CV.

8. Backtest with realistic execution costs; tune thresholds and risk params.

9. Build Research Agent (vector DB for features & feature cards).

10. Build Trader Agent (alert generation, micro-confirmation rules).

11. Build Risk & Portfolio Agent (metrics, sizing rules, circuit breakers).

12. Add LLM reasoning as approval layer (audit logs).

13. Paper trade; monitor performance & retrain as needed.

14. Small live deployment, strict limits; increase gradually.

---

14. Next steps I can do for you (pick any)

- Produce working **code skeletons** for: data fetch + 4H resampler, feature engine, lag discovery, LightGBM training, LSTM encoder, meta-model training.

- Provide **exact tickers & API endpoints** for each data item (FRED, Yahoo/AlphaVantage/IB, WGC, CFTC, etc.).

- Create the **alert JSON schema** and the **agent message contract** (producer/consumer examples).

- Design the **visual diagram** (PNG or SVG) for the architecture and an interactive dashboard layout.

- Build a **sample backtest notebook** (strategy with a few features) that you can run locally.