Four Engines, One Regime Space

An Order-Flow-Aware, Path-Geometry-Gated Architecture for Digital-Asset Perpetual Calibration — Q7 AEGIS AI S1

Haneef R. Haqq, MBA[1]

[1]Quant7 Alpha, LLC

Technical Whitepaper · v1.0 · Strategy S1 (Digital-Asset Perpetuals)

June 30, 2026

Abstract

We describe S1, the reference architecture of the Q7 AEGIS AI family, which trades USD-quoted perpetual swaps on four high-liquidity digital assets (BTC, ETH, SOL, XRP) across four intraday horizons (1, 3, 5, 15 minutes). The design premise is that crypto perpetuals exhibit several distinct intraday regimes—trend persistence, volatility-expansion breakouts, mean-reverting drift inside ranges, and dealer liquidity sweeps—at frequencies high enough that a single-engine specialist forfeits alpha. S1 runs four concurrent engines that each target one regime, arbitrated by a seven-component confluence score and a five-state volatility-regime router. Position leverage is a function of order-flow conviction rather than a fixed multiplier, grounded in the empirical result that order-flow imbalance, not trade imbalance, is the dominant near-linear predictor of short-horizon mid-price drift. A path-geometry operator (velocity / jerk / curvature) supplies a confluence boost, and a per-configuration, train-equals-serve hybrid-AI gate filters entries. We detail the engine semantics, the confluence and gating logic, the order-flow leverage map, the adaptive four-leg exit, the risk construction, and the calibration acceptance criteria, and we state the architecture's limitations explicitly. No realized-performance claims are made.

1. Edge Thesis

S1 trades directional perpetual prices, where the four canonical Q7 engines each correspond to a real microstructural regime. **Trend (TRND)** fires when the moving-average stack and trend-strength index align—persistent directional drift is the edge condition, not the failure mode. **Breakout (BRK)** fires on volatility expansion beyond the Bollinger envelope, which on perpetuals captures genuine expansions driven by liquidations and funding-flow reversals. **Mean-Reversion (MR)** fires on intraday extremes inside ranges, where market-makers fade momentum bursts—a behavior with a direct microstructural basis (Section 5). **Sweep (SWP)** targets prior-high/low liquidity sweeps with rejection wicks, i.e. explicit dealer stop-runs. The four engines partition the regime space; the confluence and volatility-regime layers arbitrate which fires. This is the structural contrast with single-engine specialists deployed on stationary spreads, where trend/breakout/sweep semantics break down.

2. Architecture

S1 is organized in three independently testable layers.

Signal. Per (asset, timeframe), the four engines compute entry candidates each bar and write a candidate direction plus an engine tag. Arbitration selects the first eligible engine whose direction-aligned confluence clears its per-engine floor, subject to a per-engine cooldown.

Intelligence. Five subsystems run in parallel: the confluence aggregator (Section 3); the volatility-regime router (Section 6); the path-geometry operator (Section 7); the four-leg adaptive exit planner (Section 8); and the hybrid-AI overlay (Section 9).

Execution. Engine arbitration → confluence gate → cooldown → hybrid-AI gate → order-flow-scaled leverage sizing → entry. Exits run the four-leg cascade with a break-even arm, an MFE ratchet, and a runner.

Asset	Horizons	Configs	Leverage cap
BTC	1·3·5·15m	4	100×
ETH	1·3·5·15m	4	100×
SOL	1·3·5·15m	4	75×
XRP	1·3·5·15m	4	50×

The 16 configurations (4 assets × 4 horizons) are each calibrated independently; the leverage cap is a per-asset ceiling, not an operating point.

3. Confluence Scoring and Gating

Per bar, seven 0–100 component scores are computed—trend agreement, trend-strength band, direction-specific oscillator band, volume ratio, multi-timeframe agreement, Bollinger position, and candle-pattern strength—and combined into direction-conditional confluence:

$$\mathrm{conf}_{dir} = \Sigma_k\, w_k \cdot \mathrm{score}_{k,dir}, \quad \Sigma_k\, w_k = 1.$$

Each engine carries a calibrated confluence floor; the effective gate applies a per-direction asymmetry adjustment under a fixed feasibility cap. Engines can be disabled per configuration when structurally adverse for that (asset, horizon), which is how mean-reversion-dominant profiles emerge for tight-ranging assets—an outcome of calibration, not a hand-set rule.

4. The Four Engines and Regime Coverage

The engines are deliberately non-overlapping in their activation conditions, so that confluence arbitration is selecting among regimes rather than among redundant signals. TRND and BRK target directional persistence and expansion; MR and SWP target reversion and liquidity events. The coverage argument is empirical, not aesthetic: each engine's activation condition maps to a documented intraday behavior of perpetual order books, and the union of conditions spans the observed regime distribution.

5. Order-Flow-Driven Leverage

S1 does not size with a fixed multiplier; per-trade leverage is a function of confluence and *order-flow* conviction, capped per asset. This choice is grounded in the limit-order-book literature. Order-flow imbalance (OFI)—the net change in size at the best bid and ask, inclusive of cancellations—explains short-horizon mid-price drift far better than trade imbalance (reported $R^2 \approx 65\%$ vs. 32%) [1, 3]. Crucially, the price-impact slope is *inversely proportional to market depth* [1], which dictates depth-normalized order-flow: leverage is attenuated when our own flow is large relative to local depth. The Markovian queueing treatment further yields an order-flow-implied volatility, $\sigma_{OF} \propto \sqrt{(\lambda/D)}$, forecastable from flow statistics without observing price [2]—used to open the breakout gate and scale trend leverage when σ_{OF}/σ_{ATR} expands, and to favor mean-reversion when it contracts. The same theory supplies a closed-form queue-imbalance direction probability with empirical first-lag *negative* autocorrelation [2], the structural fact the MR engine exploits.

6. Volatility-Regime Router (Λ-Vol)

A deterministic five-state classifier—trend-calm, trend-active, mean-revert, stress-high, transition—maps the short/long volatility ratio and drift (in standard-deviation units) to regime-conditional stop and target multipliers. It is a rule-based state machine, not a learned model, chosen for interpretability and stability; it routes exit geometry and flags engine downsizing under stress. The motivation for an explicit regime layer is the documented tendency of high-volume regimes to shift return distributions from unimodal to multimodal [8], which a single static exit geometry serves poorly.

7. Path-Geometry Operators (ζ-Field)

The ζ-Field supplies a curvature signal to the direction-aligned confluence channel via three operators on the log-price path: velocity (first derivative), jerk (third derivative, via a Savitzky–Golay coefficient), and curvature (discrete circulation). The three combine into a single signed score that boosts the aligned side additively, leaving the gate structure unchanged. The rationale is that the *order* of moves and cross-channel geometry carry information that vanishes under bag-of-features representations—the same premise that motivates the signature method in machine learning [6]; the ζ-Field is a lightweight, deterministic instance of that geometric coverage.

8. Adaptive Four-Leg Exit with Runner

Each position scales out across four ATR-spaced target legs plus a runner, with engine-specific leg allocations and target multipliers (calibrated per configuration). A break-even stop arms after the first leg or once favorable excursion exceeds a threshold; an MFE ratchet advances the trail anchor after the second leg; a minimum-bars-in-position guard protects short-horizon mean-reversion signals from premature time-stops. Any hard-risk event (circuit breaker, rolling halt, liquidation guard, macro gate) closes all remaining legs immediately—the leg machinery operates only on the normal exit path.

9. The Hybrid-AI Gate (Train = Serve)

S1 filters rule-based entries through a hybrid-AI overlay that fuses three models—a supervised directional learner, a financial-language sentiment model, and a market-regime classifier—in the spirit of the multi-modal architecture of [9], but adapted to per-configuration, intraday, train-equals-serve operation. The motivation is the same one that paper documents empirically: trend and mean-reversion logic fail through regime transitions and are blind to news shocks, and a fusion of structured ML with sentiment and regime awareness is materially more robust than any isolated predictor.

9.1 Directional model — gradient-boosted trees (XGBoost)

The directional learner is an XGBoost classifier that predicts whether the forward move is favorable. The reference design [9] frames this as a binary next-period-direction problem (label 1 if the next-period return is positive) over a compact technical feature set, with ~200 boosting rounds, tree depth 6, and a 0.05 learning rate on a 70/30 split. S1 adapts that recipe to its short-horizon, per-(asset, timeframe) setting: labels are the forward 10-bar return thresholded at ±0.5% (favorable/unfavorable), and the feature set is *direction-conditional*—separate long and short framings (roughly a dozen technical and microstructure features plus an is-long indicator and the ζ-Field velocity/jerk terms) are computed so one model scores both sides. Mechanically, XGBoost is an additive ensemble of shallow regression trees fit by stagewise gradient boosting on a regularized logistic objective: each new tree fits the negative gradient of the loss given the current ensemble, scaled by a shrinkage rate, with row/column subsampling and L1/L2 leaf-weight penalties; the ensemble outputs a calibrated probability. Two disciplines distinguish S1's use from the reference: the model is trained *in-cycle* on the current cycle's own accumulated trades (no stale prior-cycle model), and it is admitted only if it clears a per-direction discriminability floor (AUC ≥ 0.55 long *and* short) and an adversarial leakage bound (adversarial AUC < 0.85). Its entry threshold is itself a calibratable parameter, and the gate is calibratable on/off—if the optimizer selects it off, the rule-based stack carries the decision.

9.2 Sentiment model — FinBERT

The sentiment channel is FinBERT [10], a BERT-family transformer fine-tuned on financial text. Each news item is tokenized and contextualized through stacked self-attention layers and mapped by a classification head to probabilities over {positive, neutral, negative}; domain fine-tuning is what lets it parse negation, hedging, and finance-specific jargon that generic sentiment models miss. Per asset per day the article-level scores are aggregated into a normalized polarity,

$$S_t = (1/N)\ \Sigma_i\ (\ P_{positive}^{(i)} - P_{negative}^{(i)}\ ),\quad S_t \in [-1, 1].$$

Following [9], S1 consumes S_t as a *risk filter and soft threshold skew, never as a standalone entry signal*: a sufficiently negative S_t (below a calibrated floor) suspends new entries and can flatten exposure, shielding core positions against adverse-news shocks—earnings surprises, regulatory actions, guidance cuts—that purely technical logic is blind to. *Honesty note:* in a historical backtest without a time-aligned, point-in-time news feed for the evaluated window, this channel is run neutral under a signed no-feed exemption (honest-empty) rather than fabricating sentiment; the exemption is recorded in the cycle dossier, and train=serve is preserved by deploying the same neutral configuration that was calibrated.

9.3 The tandem and the regime channel

The third model is the market-regime classifier—the Λ-Vol router (Section 6) at the strategy level, and a regime label at the overlay level—which conditions both the gate and the exit geometry. The three combine as a tandem: the directional model proposes a probability, the sentiment channel can veto or attenuate it, and the regime label selects the operating envelope. This is the multi-modal fusion that [9] shows outperforms any single isolated predictive model across changing regimes. The decisive engineering choice is that all three are computed in-cycle from the same data the system trades—the overlay calibrated is the overlay deployed, so the gate cannot be a flattering in-sample artifact of a model the live system never runs.

10. Calibration and Acceptance

Each configuration is accepted into a cycle only on a simultaneous five-criterion gate—win rate, gross profit factor, trade frequency, drawdown (a soft, logged criterion), and a commission-to-gross-profit bound (a hard block)—evaluated on the hybrid-AI-gated system, with the hybrid-AI discriminability bounds enforced separately. Cycle eligibility additionally requires full coverage with no blocked configuration, positive aggregate result, and in-cycle model freshness. Promotion toward live deployment requires consecutive consistent cycles, specification-to-implementation parity within tolerance, and an independent re-derivation—the governance described in the program's pipeline whitepaper.

11. Risk and Venue

Risk is enforced preventively at the sizing layer: positions are sized so that a worst-case adverse move, measured from the monotonically ratcheting equity peak, cannot carry the account past the hard drawdown ceiling, which is a structural constant identical across the specification, the calibrator, and the executor. Order-flow-aware leverage attenuation (Section 5) further reduces self-impact. The execution venue is a digital-asset perpetuals exchange with a depth feed; commissions use the venue's maker/taker schedule with a conservative taker-on-both-sides round-trip assumption, and perpetual funding is amortized into per-trade economics by hold duration.

12. Limitations

The order-flow results that motivate the leverage map are strongest with a live depth feed; under a bar-derived order-flow proxy the depth normalization is approximate, and the architecture treats the live-depth path as an enhancement, not a precondition. The network-momentum literature [4, 5] indicates that a learned cross-asset lead-lag graph supplies orthogonal alpha and is the principled replacement for any fixed leader heuristic; this is a roadmap item, not a present claim. The regime classifier is deterministic and can saturate; the hybrid-AI gate constrains but does not certify causal or durable structure; and execution, liquidity, leverage, and venue/regulatory risks are not eliminated by architectural integrity. As with every Q7 deliverable, no realized-performance representation is made, and the acceptance criteria above are calibration gates, not outcomes.

References

[1] R. Cont, A. Kukanov, S. Stoikov. *The Price Impact of Order Book Events.* 2011.

[2] R. Cont, A. de Larrard. *Price Dynamics in a Markovian Limit Order Market.* 2011.

[3] M. Gould et al. *Limit Order Books* (survey). Quantitative Finance, 2013.

[4] *Follow the Leader: Enhancing Systematic Trend-Following Using Network Momentum.*

[5] *Network Momentum across Asset Classes.*

[6] *A Primer on the Signature Method in Machine Learning.*

[7] R. Cont, J. da Fonseca. *Dynamics of Implied Volatility Surfaces.* Quantitative Finance, 2002.

[8] *Quantum-Probability-Theoretic Asset Return Modeling* (multi-modal return distributions at high-volume regimes). 2024–2025.

[9] V. N. K. Pillai, A. Ajith, S. K. J. *Generating Alpha: A Hybrid AI-Driven Trading System Integrating Technical Analysis, Machine Learning and Financial Sentiment for Regime-Adaptive Equity Strategies.* ComSIA 2026 (Springer LNNS).

[10] D. Araci. *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.* 2019.
