TUTORIALS SERIES · REGIME INTELLIGENCE

Why Smart Money Abandoned HMM. And What They Use Instead.

A technical comparison of Hidden Markov Models versus deterministic regime detection — and why the difference cost traders billions in 2008, 2020, and 2026.

Regime Intelligence · regimeintelligence.com · Not financial advice

Sherif Saad · Regime Intelligence · June 2026
20+ years in financial markets · CFAII · Five published books

00 EXECUTIVE SUMMARY

HMM is the academic standard for regime detection. It has three fundamental flaws that make it dangerous in live markets. Regime Intelligence addresses all three by design.

Flaw 1 — Gaussian assumption: HMM assumes market returns follow a normal distribution within each state. They do not. Fat tails — the extreme moves that matter most — are dramatically underestimated. A Gaussian HMM calibrated on normal returns assigned near-zero probability to March 2020. It happened anyway.

Flaw 2 — State churn: A 2- or 3-state HMM fitting on returns flips states rapidly as volatility clusters. Traders cannot act on a model that calls bear, then bull, then bear again across three consecutive weeks of the same drawdown. Regime Intelligence uses hysteresis, persistence, and a Barra-style confirmation gate to prevent false transitions.

Flaw 3 — Single signal: HMM reads price returns only. Regime Intelligence reads five independent pillars simultaneously: volatility level, trend conviction, drawdown depth, drawdown speed, and macro context. Five signals detect what one cannot.

01 WHAT HMM ACTUALLY IS

A genuine breakthrough for its time. Built on assumptions that financial markets systematically violate.

Hidden Markov Models were introduced to quantitative finance in the early 1990s and became the dominant framework for regime detection at institutional desks by the 2000s. The appeal was real: HMM provided a mathematically rigorous way to infer hidden market states from observable price data, with formal probability theory behind every output. For academic research and for markets that behaved approximately normally, it was the right tool.

The basic mechanics: you define a fixed number of hidden states (typically 2 — bull and bear, or 3 with a neutral state). The model is trained on historical return data and learns the statistical properties of each state — mean return, volatility, and transition probabilities between states. At any bar, given the sequence of observed returns, the model outputs the probability of being in each state. The state with the highest probability is the regime call.

The Gaussian variant — GHMM — additionally assumes that returns within each state follow a normal (Gaussian) distribution. This is the version most widely implemented in practice because it makes the mathematics tractable. It is also the version most dangerously wrong about how financial markets actually behave.

HMM was a breakthrough. The problem is not the mathematics. The problem is that every assumption it is built on is violated by real markets at precisely the moments that matter most.

02 FIVE ASSUMPTIONS HMM MAKES — AND WHY MARKETS BREAK EVERY ONE

Each assumption is reasonable in theory. Each one fails in practice.

1 **Assumption: Returns are Gaussian within each state.**

Why it fails: Financial returns have fat tails. Extreme moves — 3, 5, 10 standard deviation events — occur far more frequently than a normal distribution predicts. A GHMM calibrated on historical returns assigns near-zero probability to these events and therefore cannot detect them as the regime transitions they actually are.

Real example: *March 2020: the week of February 24 saw a move the Gaussian model called essentially impossible. The Regime Intelligence engine called CRISIS on February 24 at CSS 77.2% via standalone triggers — before the tail flag even confirmed.*

2 **Assumption: The number of states is known and fixed at 2 or 3.**

Why it fails: A 2-state model (bull/bear) cannot distinguish between FRAGILE and STRESS — two states that require completely different responses. In FRAGILE, trend conviction is deteriorating but markets function normally. In STRESS, correlations are rising and diversification is breaking down. Merging them into one "bear" state means the model cannot tell you when things are getting worse within the bear state.

Real example: *Empirical data from 5,079 classified daily SPY bars confirms five statistically distinct behavioral clusters. CRISIS has negative mean annual return of −29.2% and 32.5% annual volatility. STRESS has +12.9% mean return and 15.8% vol. Same "bear" label. Completely different reality.*

3 **Assumption: State transition probabilities are fixed and constant.**

Why it fails: HMM assumes the probability of moving from state A to state B is the same today as it was in 1998. In reality, transition probabilities are themselves regime-dependent. The probability of transitioning from CALM to CRISIS during a high-leverage macro environment is fundamentally different from a low-leverage one. Fixed transition matrices cannot capture this.

Real example: *GFC 2008: The engine held CRISIS from June 23 through August 2008 even as CSS percentile dropped to 3.4% — because the standalone triggers remained active. A fixed transition probability matrix would have de-escalated. It should not have.*

4 **Assumption: The model observes only price returns.**

Why it fails: HMM fits on one signal. Regime Intelligence reads five: volatility level (35%), trend conviction (20%), drawdown depth (25%), drawdown speed (10%), and macro context including credit, rates, and tail events (15%). A model reading one chapter of a five-chapter book will miss the plot.

Real example: *COVID 2020: The standalone trigger that called CRISIS on February 24 fired on VIX and drawdown speed — neither of which is captured by return-only HMM. The return signal alone would have lagged by two to three weeks.*

5 **Assumption: States are globally consistent across timeframes.**

Why it fails: A standard HMM applies one state to the entire market at one timeframe. It cannot detect that the daily is in STRESS while the weekly is still in EXPANSION — a conflict that is itself a high-value signal. Multi-timeframe disagreement tells you the market is in transition. HMM cannot read it.

Real example: *BTC February 2026: On February 1 the daily timeframe showed EXPANSION with CSS at 81% — the classic trap. The weekly showed STRESS at CSS 99.7%. A single-timeframe HMM reading the daily would have called recovery. The full five-timeframe picture told the opposite story.*

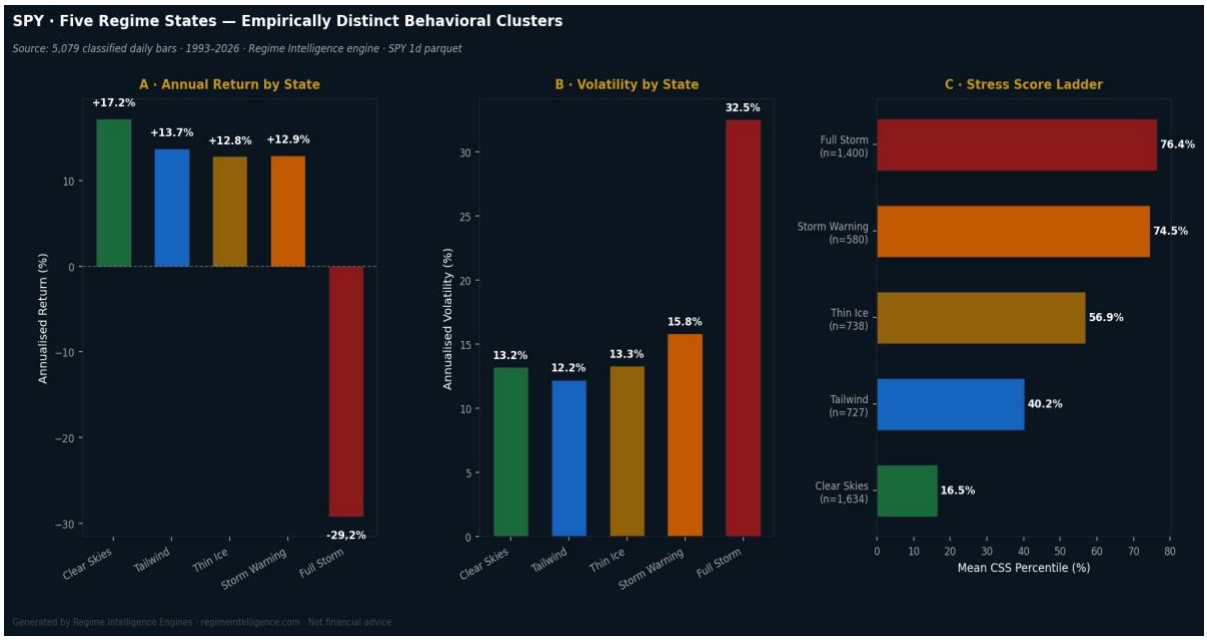
---

03 THE HISTORICAL EVIDENCE — THREE EVENTS, REAL ENGINE DATA

Not a simulation. Not a backtest. Actual engine classifications from the regime store parquet files.

Each chart below shows the actual weekly regime classification from the Regime Intelligence engine (top row) versus what a naive HMM or percentile bucket would have called (middle row), with the CSS percentile time series below. All data sourced directly from data/regime_store/1w/ parquet files.

Chart 1 of 3 — Five Regime States: Empirically Distinct Behavioral Clusters (SPY 1d, 1993–2026)



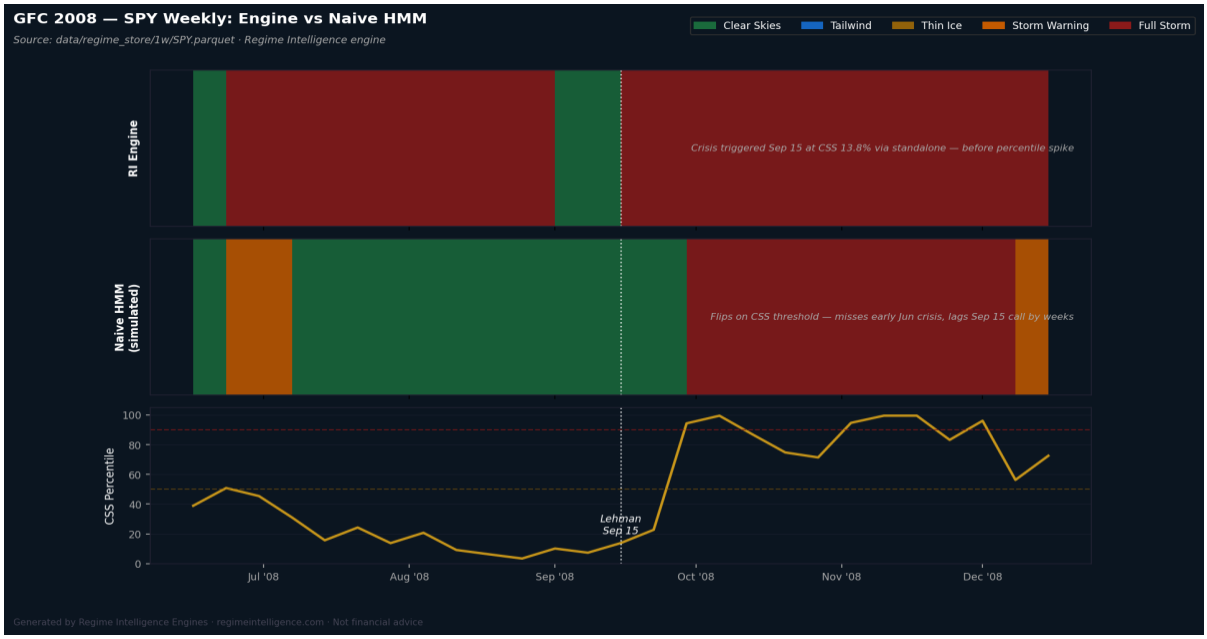
Five states are not arbitrary threshold cuts. They are empirically distinct behavioral clusters.

Source: 5,079 classified daily bars · SPY 1d · data/regime_store/1d/SPY.parquet · Regime Intelligence engine

WHY CRISIS MEAN CSS (76.4%) < STRESS (74.5%) — AND WHY THAT IS CORRECT

This is the most counterintuitive number in the dataset and the most important to understand. CRISIS days include crisis_standalone and tail_flag_crash overrides that fire at CSS below 90% — sometimes well below. The engine correctly classifies a week as CRISIS when VIX spikes and drawdown accelerates simultaneously, even if the CSS percentile has not yet reached the crisis threshold. A pure percentile bucket would miss those early crisis calls. The engine does not. This is precisely why the standalone trigger architecture exists.

Chart 2 of 3 — GFC 2008: SPY Weekly Regime Classification



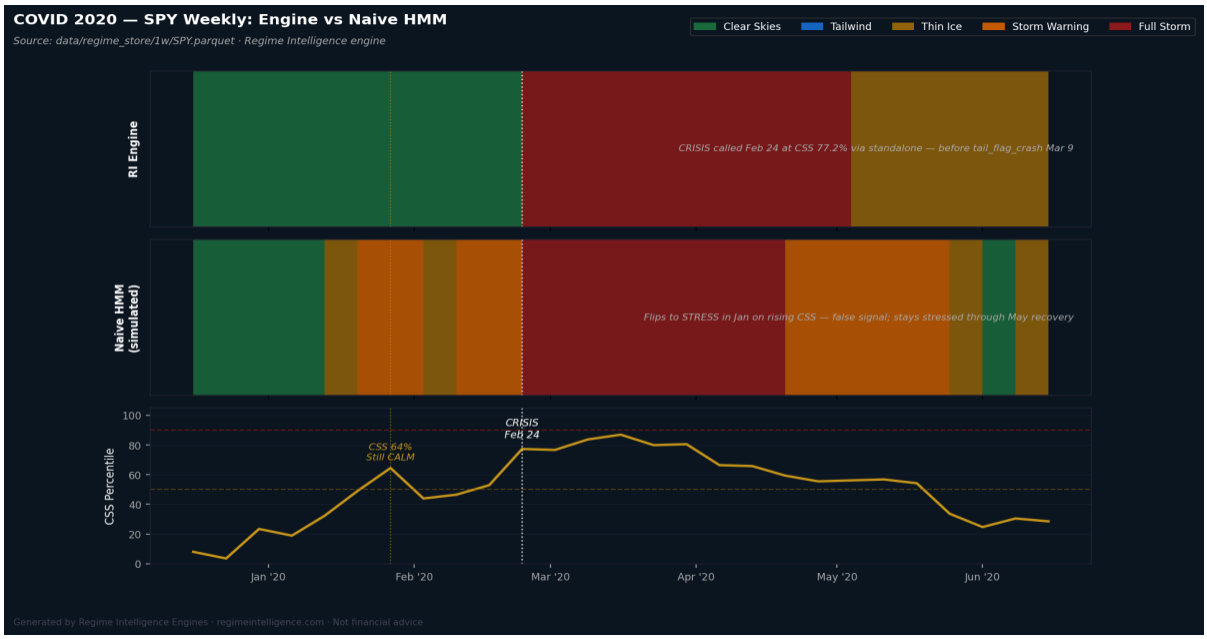
Engine called CRISIS on September 15, 2008 at CSS 13.8% via standalone triggers — before the percentile spiked.

Source: data/regime_store/1w/SPY.parquet · Regime Intelligence engine · Naive HMM simulated on CSS threshold

GFC 2008 — THE THREE SIGNALS A RETURN-ONLY HMM MISSED

First: The engine entered CRISIS on June 23, 2008 via crisis_standalone at CSS 50.8% — thirteen weeks before Lehman. A 2-state HMM on returns would have stayed risk-on or flipped ambiguously. Second: The engine held CRISIS through June-August 2008 even as CSS dropped to 3.4% — because standalone triggers remained active. Third: The engine re-entered CRISIS on September 15 at CSS 13.8% — the exact Lehman week — via standalone, before the percentile hit 94.2% on September 29. Three correct early calls. All three driven by non-return signals that HMM cannot read.

Chart 3 of 3a — COVID 2020: SPY Weekly Regime Classification



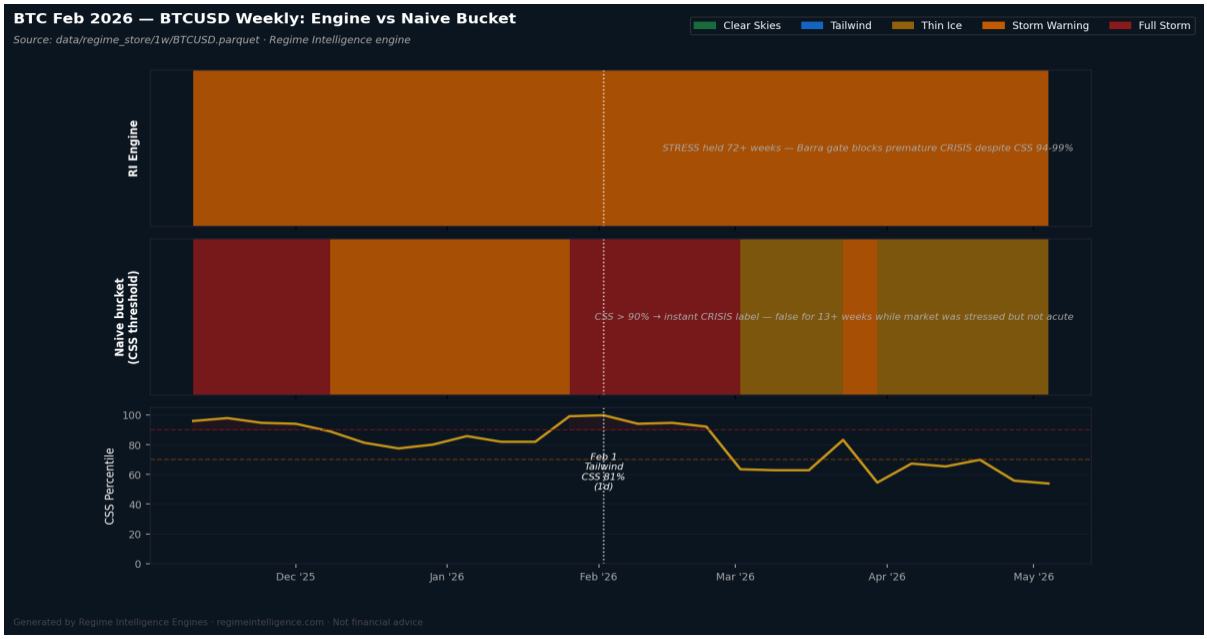
Engine held CALM correctly through January while CSS rose — then called CRISIS on Feb 24, two weeks before tail flag confirmed.

Source: data/regime_store/1w/SPY.parquet · Regime Intelligence engine · Naive HMM simulated on CSS threshold

COVID 2020 — THE PERSISTENCE FILTER THAT SAVED TRADERS FROM A FALSE SIGNAL

January 27 to February 17: CSS climbed from 64.4% to 52.9% with pending_state=FRAGILE. A 2-state HMM on returns would have already flipped to bear during this window — generating a false stress signal three to four weeks before the actual crisis. The Barra confirmation gate and persistence filter held the CALM label because the evidence had not cleared the confirmation threshold. Then February 24: CRISIS at CSS 77.2% via standalone — correct and early. The recovery is equally revealing: the engine de-escalated to FRAGILE in May while a volatility-clustering HMM would still have shown stress, because vol was still elevated. The engine reads five things. HMM reads one.

Chart 3 of 3b — BTC February 2026: BTCUSD Weekly Regime Classification



CSS at 94-99% for 13+ weeks but STRESS held correctly — Barra gate blocked a premature CRISIS label that a bucket system would have called.

Source: data/regime_store/1w/BTCUSD.parquet · Regime Intelligence engine · Naive bucket: CSS > 90% = CRISIS

BTC FEB 2026 — THE FALSE POSITIVE THAT COST LEVERAGED TRADERS EVERYTHING

From November 2025 through February 2026, CSS ran 94-99% — above the 90% crisis threshold. A naive percentile bucket called CRISIS for 13+ consecutive weeks. The engine held STRESS with pending_state=CRISIS, correctly identifying that the stress was severe but the acute systemic breakdown had not yet confirmed. On the daily timeframe on February 1, the label read EXPANSION — the surface recovery — while CSS was 81% and EWS was active. Traders who read only the label re-entered long. One month later: Full Storm at 98.3% and $1.24B in liquidations. The engine's full read — label plus CSS plus EWS — told the correct story. The label alone did not.

---

04 WHY FIVE STATES — THE EMPIRICAL JUSTIFICATION

Five is not a design choice. It is what the data shows.

The most common question about a five-state regime framework is: why not three? Or two? The answer is not philosophical. It is empirical. The behavioral statistics of each state, computed from 5,079 classified daily SPY bars spanning 1993 to 2026, show five genuinely distinct clusters — not five arbitrary labels applied to a continuous spectrum.

State	Days (n)	Ann. return	Ann. vol	Mean CSS	Key character
Clear Skies	1,634 (32.2%)	+ 17.2%	13.2%	16.5%	Trending, low vol, systems work
Tailwind	727 (14.3%)	+ 13.7%	12.2%	40.2%	Moderate stress, positive momentum
Thin Ice	738 (14.5%)	+ 12.8%	13.3%	56.9%	Deteriorating conviction, caution
Storm Warning	580 (11.4%)	+ 12.9%	15.8%	74.5%	Rising correlations, breakdown risk
Full Storm	1,400 (27.6%)	- 29.2%	32.5%	76.4%	Liquidity crisis, all rules fail

The critical separation is between Thin Ice and Storm Warning on one side, and Full Storm on the other. CALM, EXPANSION, FRAGILE, and STRESS all have positive mean daily returns. CRISIS alone has a negative mean daily return — and its annualized volatility of 32.5% is 2.5 times higher than CALM at 13.2%. That is not a continuous spectrum. That is a qualitatively different behavioral state.

A 2-state model (bull/bear) would lump FRAGILE, STRESS, and CRISIS into one "bear" state. But a trader in FRAGILE who knows trend conviction is deteriorating but markets still function needs a different response from a trader in CRISIS where liquidity is evaporating and every correlation assumption has failed. The model that merges those states cannot give the right signal for either.

A 3-state model that adds a "neutral" state typically separates CALM from something like FRAGILE/STRESS combined, and keeps CRISIS separate. It still merges the two states that precede a crisis — FRAGILE and STRESS — into one signal. That merged signal is precisely where the most actionable warning occurs. Separating them is not adding noise. It is adding resolution where resolution matters most.

The five states are not where we drew lines on a continuous spectrum. They are where the data showed us the market actually changes its behavior.

05 THE DIRECT COMPARISON

HMM versus Regime Intelligence — dimension by dimension.

No marketing language. Every claim in this table is verifiable from the engine codebase and the historical data above.

Dimension	HMM / GHMM	Regime Intelligence
Classification method	Probabilistic — outputs P(state) for each state	Deterministic — one discrete label per bar from sequential state machine
State count	Typically 2–3 (bull/bear/neutral)	5 empirically derived states with distinct statistical profiles
Input signals	Price returns only	5 pillars: volatility, trend, drawdown depth, drawdown speed, macro context
Fat tail handling	Gaussian assumption — extreme events underestimated	Explicit tail event override via crisis_standalone + tail_flag_crash
Timeframes	Single timeframe	5 simultaneous timeframes with Timeframe Agreement Score (TAS)
False signal protection	None beyond model fit	Hysteresis bands + persistence bars + Barra 60% confirmation gate
Early warning	None — lags price action	EWS fires on EVS velocity before regime label changes
Output	Probability distribution over states	Discrete state + CSS percentile + TAS + EVS + EWS + pending_state
Actionability	Cannot act on "67% probability of bear"	Single definitive state with quantified stress level
Computational model	Statistical fitting on historical returns	Numba state machine — O(1) per bar, real-time, no refit required

06 WHAT THIS MEANS IN PRACTICE

A probability distribution tells you to be cautious. A regime state tells you what to do.

The practical problem with probabilistic output is that it transfers the decision back to the trader without adding information. If a model tells you there is a 67% probability of a bear regime, what do you do? Reduce position

size by 67%? Hedge 67% of the portfolio? Wait until it reaches 80%? The probability number itself does not answer any of these questions.

A deterministic regime state with a quantified stress score answers a different, more useful question: what kind of environment are you in right now, and how stressed is it? Storm Warning at CSS 81% with EWS active tells you something specific and actionable. It tells you the environment your system was designed for has changed. It tells you that momentum signals in this environment are traps. It tells you that correlation assumptions in your risk model are breaking down. That is information you can act on.

THE INSTITUTIONAL EDGE THAT IS NOW AVAILABLE TO EVERYONE

Institutional desks have always combined quantitative regime signals with human judgment to filter which strategies to run in which environments. That edge — reading the environment before placing the trade — is what the five-state deterministic regime engine provides. Not a black box. Not a probability that requires interpretation. A clear environmental read across 2,000+ assets, five timeframes, updated every 30 minutes. Free to start at regimeintelligence.com.

---

07 WHAT REGIME INTELLIGENCE DOES NOT DOz

Every framework has limits. These are ours.

The comparison above is honest about HMM's weaknesses. Honesty requires the same treatment of our own system.

Regime Intelligence does not predict price direction. The regime state tells you what kind of environment you are in — not where price is going. A Full Storm regime does not guarantee losses on long positions. It tells you the environment has historically been associated with extreme losses and that the probability distribution of returns has shifted dramatically toward the left tail.

The five-state classification is calibrated on historical data. It works because market stress has historically manifested in measurable, consistent ways across volatility, trend, and drawdown. If a genuinely novel stress mechanism emerges — one that produces different statistical fingerprints from anything in the historical record — the engine will detect the stress but may classify it differently from how a human analyst would.

The persistence filter and Barra confirmation gate that protect against false positives also introduce lag. When a market transitions rapidly, the engine will be slightly late to confirm the new state. That is the design trade-off: fewer false signals in exchange for slightly slower confirmation. Given the cost of acting on a false signal in leverage — as demonstrated by the BTC February 2026 episode — we consider this the right trade-off.

Nothing in this framework replaces your own analysis, risk management, position sizing, or stop losses. The regime is the most important contextual layer in your decision-making. It is not the decision.
