PERCEPTUAL MACROECONOMICS

FOUR YEARS LATER

What Four Years of Markets Taught Us About Narrative, Regime Change, and Forecasting Under Uncertainty

An Anniversary Addendum to the IUVO™ Forecasting System

John M. Aaron
Milestone Planning and Research, Inc.
August 2026
www.ratio-weekly.com

Publication and Use Notice

This addendum accompanies Perceptual Macroeconomics: Narrative Co-Integration, Regime Dynamics, and the Limits of Federal Reserve Data in Asset Price Forecasting — An Introduction to the IUVO™ Forecasting System.

Authorship and Responsibility. The author assumes responsibility for the conceptual framework, empirical interpretation, and conclusions presented here. Analytical software and large language models were used as research, coding, interpretive, and drafting aids under human supervision.

Investment Disclaimer. This paper and the IUVO™ materials discussed in it are analytical and educational research. They do not constitute investment advice, a recommendation to buy or sell any security, or a representation that any forecast will be realized. Forecast cones describe modeled uncertainty, not guaranteed price ranges. Historical relationships may change.

Methodological Disclaimer. Statistical association, co-integration, error correction, regime conditioning, impulse-response estimates, and narrative classification should not be interpreted as proof of deterministic causality. The system is designed to support judgment under uncertainty, not replace it.

Abstract

Four years of continuous observation have changed the Perceptual Macroeconomics research program in an important way. The original proposition—that a measured public narrative field and the S&P 500 can share a persistent long-run equilibrium relationship—remains central. What changed is both the baseline and the forecasting philosophy built around it. After nearly four years, IUVO retired the pre-COVID growth-rate benchmark and replaced it with a fitted trajectory derived from the contemporary market record; the evidence increasingly suggested that the old reference was treating a changed process as though it were a temporary deviation.

The accumulated record also shows that narrative is informative without being mechanically causal; that the narratives associated with market stress change across eras; and that apparently similar shocks can generate materially different responses. A nonlinear challenger test using the 20-trading-day forward change in deviation from trajectory retained substantial out-of-sample explanatory power and repeatedly elevated narrative variables among its leading predictors. Taken together with the narrative-override backtest, these results support a common-cause/special-cause interpretation: persistent and recurring narratives appear to carry information about the ordinary state in which market variation is generated, while exceptional narrative configurations can sometimes identify conditions under which the baseline deserves to be challenged.

The evidence is asymmetric rather than universal: narrative intervention improved overall directional hit rate by 5 percentage points and added strongly in the later regime, but hurt in the earlier regime, while perfect-information experiments showed that knowing future driver values does not eliminate uncertainty or model-form risk.

The practical consequence is a conservative hybrid architecture. IUVO begins with an empirically fitted market trajectory and a 20-trading-day probabilistic forecast cone; narrative is used to interpret state and selectively challenge the baseline. A new analyst scenario laboratory extends that architecture by allowing forward-looking judgments about identified narrative drivers to be entered with explicit evidentiary weight, producing a conditional forecast without rewriting the production model. The result is a more mature view of machine-assisted forecasting: use statistics to define the current process, narrative to distinguish ordinary state from potentially exceptional conditions, and human judgment to incorporate information that the historical model cannot yet observe.

Keywords: perceptual macroeconomics; narrative economics; co-integration; regime change; forecast cone; narrative override; hybrid cognition; S&P 500; forecasting under uncertainty.

1. Four Years as an Empirical Test

The Perceptual Macroeconomics project began with a practical question: can the language surrounding economic events reveal information about market behavior that conventional macroeconomic indicators either miss or report too late? The original research treated the S&P 500 as a demonstration vehicle for a broader organizational proposition. If text generated by a complex environment contains a measurable latent state, and that state maintains a persistent relationship with an outcome of interest, then the text stream can become part of a forecasting architecture.

Four years of daily observation have produced a more qualified—and more useful—answer than the simple claim that news moves markets. Narrative matters, but it does not operate as a fixed collection of headline-to-price response coefficients. Markets are adaptive systems. Price, expectations, fear, policy, institutional credibility, political regime, media selection, and historical context interact. A narrative that is highly informative in one period can become background noise in another. A shock that dominates one episode may have little relevance to the next. A statistical response estimated from several heterogeneous episodes can obscure the episode that actually matters.

The anniversary therefore marks more than the accumulation of observations. It marks a change in the theory of use. The system has moved from asking how many narrative effects can be estimated toward asking a harder question: which evidence deserves the right to change the forecast?

2. What Survived: The Long-Run Narrative–Price Relationship

The strongest proposition to survive the four-year experiment is the distinction between long-run equilibrium information and short-run return prediction. The original research found a strong persistent relationship between the dominant narrative composite and the S&P 500 price level, formalized through co-integration and error-correction methods. At the same time, forward one-day narrative-return correlations were extremely small. Those two results are not contradictory.

Co-integration is a theory of levels. It asks whether two non-stationary processes share a common stochastic trend and whether deviations from their long-run relationship tend to correct. It does not require today's change in narrative to predict tomorrow's raw market return. The appropriate interpretation is therefore associative and equilibrium-based: narrative and price can characterize the same evolving macro state without narrative mechanically forcing the next price change.

This clarification matters operationally. IUVO is not a headline trading engine. Its purpose is to estimate where the market sits relative to an evolving historical trajectory, quantify the uncertainty surrounding the forward path, and identify evidence that may indicate the current environment no longer behaves like the historical baseline.

2.1 The Baseline Changed Too: Retiring the Pre-COVID Growth Assumption

An important methodological change is easy to miss because it concerns the reference line rather than the narrative layer. For nearly four years, IUVO retained the pre-COVID S&P 500 growth rate as the long-run benchmark. That was a defensible starting assumption: it provided a fixed historical reference against which post-pandemic departures could be measured. But a benchmark is itself a model assumption, and eventually the accumulated post-2022 record made the old reference less useful. We therefore stopped asking the current market to revert toward a growth path inherited from a materially different pre-COVID environment.

The current system instead estimates the market's fitted growth trajectory from the observed contemporary record and uses that empirical rate as the structural reference. This is not a cosmetic rebasing. It changes the meaning of deviation from target. Under the old approach, a persistent difference between the market and the pre-COVID path could continue to be labeled a deviation even after the market had established a different sustained growth regime. Under the revised approach, the baseline is allowed to learn from the realized post-2022 market while remaining a long-run reference rather than a short-term forecast.

For a non-specialist, the distinction is simple: we stopped insisting that today's market should grow at yesterday's pre-COVID rate merely because that was the rate with which the experiment began. After nearly four years of evidence, the market itself had supplied enough information to justify a new benchmark. The 20-day forecast is still anchored to the last actual S&P 500 observation; the fitted contemporary trajectory tells us what constitutes ordinary movement around the current long-run path.

This change is also a model-governance lesson. Baselines should be stable enough to prevent constant model chasing, but they should not become articles of faith. When a reference process has changed persistently, refusing to update the baseline can make normal behavior look anomalous and can distort both narrative interpretation and forecast error.

3. What Changed: Narrative Is State-Dependent

One of the clearest findings from the expanded record is that the narrative associated with stress changes across eras. FEAR appears repeatedly as a common state variable, but the secondary narrative has not been stable. Earlier stress episodes were associated more strongly with inflation and Federal Reserve tightening; later periods brought financial-collapse, China, tariff, fiscal-disruption, and geopolitical narratives to the foreground.

This distinction suggests a useful taxonomy. Persistent state variables describe the general condition of the narrative field. Recurring narratives reappear often enough to support historical estimation, although their effects may still vary by regime. Episodic shocks are sparse and event-specific. Treating all three classes identically creates false precision.

Narrative class	Typical behavior	Forecast role	Primary caution
Persistent state	Repeated, broad, slow-moving	Conditions baseline and uncertainty	May reflect regime rather than discrete cause
Recurring narrative	Multiple comparable episodes	Can support conditional historical estimates	Effect may drift across eras
Episodic shock	Sparse, event-specific, often unequal magnitude	Scenario context; override only if evidence warrants	Pooling can dilute or reverse the relevant response

3.1 Rethinking Variation: Common Causes, Special Causes, and Narrative

The four-year record also suggests a useful shift in how the forecasting problem is framed. Rather than treating every narrative movement as a discrete shock, IUVO can be viewed through the classical distinction between common-cause and special-cause variation. Common causes are the recurring forces and interactions that continuously generate ordinary variation within the current system. Special causes are unusual events, discontinuities, or combinations of conditions that push behavior outside what the existing process would normally lead us to expect.

That distinction maps naturally onto the narrative field, but not in a one-variable-one-cause way. Common causes may be embedded in persistent and recurring narratives: fear, policy expectations, trade concerns, institutional confidence, geopolitical background risk, and other themes that repeatedly rise and fall as part of the normal information environment. Their value may lie less in announcing a singular event than in describing the evolving state in which ordinary market variation is being generated. The co-integrating narrative structure and the forward TreeNet results are consistent with the proposition that text contains information about this changing state.

Special causes may also be embedded in narrative, but differently. A war, banking failure, abrupt tariff announcement, institutional rupture, or other exceptional development can appear as a new topic, an extreme intensity in an existing topic, an unusual combination of narratives, or a break in the historical relationship between narrative and price. The Iran war narratives demonstrate the difficulty: an event can be plainly exceptional without supporting a stable reusable coefficient. Special-cause detection therefore requires context and judgment in addition to statistical extremeness.

The important research question is no longer simply, “Do narratives predict the market?” It is more operational: can narrative help us distinguish variation that belongs to the current system from evidence that the system itself has changed? If recurring narrative states help characterize common-cause variation, they may improve the baseline forecast and its uncertainty. If exceptional narrative configurations help identify special causes, they may

warn that the baseline should be challenged, the cone widened, or an override considered. The two roles are complementary rather than competing.

This also clarifies why shock analysis was reduced rather than abandoned. Treating every statistically detectable shock as special creates noise and invites overreaction. Treating every observation as common-cause variation creates the opposite error: genuine discontinuities are absorbed into a historical average that no longer describes the current situation. The production architecture therefore uses the fitted trajectory and forecast cone to represent the ordinary process, while narrative provides both state information about common variation and a monitored channel for evidence of special causes.

The evidence now permits a stronger answer than merely stating a fifth-year objective. It does not prove a clean statistical partition between common and special causes, but it does show that narrative is operating in both roles. First, common-cause information is present in the narrative field: the long-run co-integrating structure, persistent FEAR state, and the 20-day TreeNet challenger all indicate that recurring text variables help describe the ordinary state around the current baseline. In the cleaner TreeNet specification, out-of-sample R-squared remained 37.87%, and narrative variables became more prominent after redundant chronology variables were removed. That is difficult to reconcile with a view of narrative as useful only when a rare shock occurs. Second, special-cause information is also present, but it is less stable. The three major downturns were accompanied by distinct clusters of unusually intense narratives; the conservative override layer intervened on only 17.86% of completed origins yet improved overall directional hit rate from 53.57% to 58.57%. That is consistent with a selective special-cause function. However, the effect is regime-sensitive: narrative intervention hurt the earlier/Biden-era backtest by 4.55 percentage points and improved the later/Trump-era backtest by 13.51 points. The Iran work likewise showed that an event can be exceptional without yielding a stable reusable response coefficient or a dominant structural-break dummy. The practical conclusion is therefore specific: recurring narrative state can help describe common-cause variation; exceptional narrative configurations can help flag some special-cause conditions; and forecasting improves when the two roles are separated rather than when every narrative movement is treated as a shock.

3.2 What the Results Say About Common and Special Causes

Changing the baseline itself is part of the answer. For almost four years the pre-COVID growth rate was treated as the expected process and post-COVID departures were measured against it. By 2026 that assumption had become increasingly difficult to defend. Re-estimating the fitted trajectory from the contemporary record amounts to recognizing that what originally looked like a special departure may have become the new common process. In statistical-process-control terms, the center line had to be reconsidered before individual departures from it could be interpreted intelligently.

This reframing also explains why the production system became simpler. The fitted contemporary trajectory and its forecast cone now represent the best estimate of common-cause behavior. Narrative does not continuously add or subtract points from that baseline.

Instead, it serves two narrower functions: describe the state in which common variation is occurring, and provide evidence that an observation or developing configuration may belong to a different process. The override gate is therefore analogous to a governed special-cause escalation rule rather than a second unconstrained forecast.

The empirical record supports this architecture, but with an important warning. A useful special-cause detector is not one that fires often; it is one that improves decisions when it does fire. IUVO's override layer improved the total 20-day directional hit rate while intervening on fewer than one in five completed origins. Yet the sharp regime asymmetry shows that the mapping from narrative to outcome itself can change. This means that the narrative measurement system must also be monitored as a process: a signal can look exceptional because the world changed, because media emphasis changed, or because the relationship between the two changed.

Accordingly, the common-cause/special-cause question is no longer unanswered. The present results say: yes, both appear to be embedded in narrative; yes, both can contribute to forecasting; but they contribute differently. Common-cause narrative is primarily state information. Special-cause narrative is primarily challenge information. The remaining research problem is not whether those roles exist, but how reliably and early the system can distinguish them.

4. The Same Narrative Does Not Necessarily Mean the Same Market

The geopolitical work provides a particularly useful example. The Iran-related narrative appeared in several episodes of very different magnitudes. A full-sample impulse-response estimate averaged those heterogeneous episodes and produced a response that was not representative of the dominant 2026 event. Restricting estimation to the dominant episode changed the inferred path materially.

The lesson extends beyond geopolitics. Historical occurrence establishes possibility; it does not establish recurrence. A banking crisis, tariff dispute, policy surprise, or geopolitical confrontation may provide a useful analogue, but an analogue is not an identity. Market structure, valuation, policy reaction functions, positioning, media amplification, and the surrounding narrative regime can all differ.

This is why the current system treats historical shock analysis as supporting evidence rather than an automatic forecast adjustment. A statistically interesting shock response is not, by itself, decision-relevant.

5. Political Regime, Media Regime, and the Measurement Problem

The four-year record also forces a more explicit treatment of political and media regimes. The source portfolio evolved over the observation period, while the political environment changed materially. The measured narrative field is therefore not generated by a fixed sensor. It is generated by a changing media ecosystem observing a changing political-economic system.

This is not merely a data-quality nuisance. It is theoretically important. Perceptual Macroeconomics is concerned with mediated economic reality: the representations that market participants actually encounter. If the media system changes its selection, emphasis, vocabulary, or amplification patterns across administrations, then the measurement environment itself has entered a new regime.

Table 2. 20-Day Narrative Performance by Policy/Administration Regime

Regime	N	Baseline hit	Narrative hit rate	Narrative gain	Override rate
Earlier/Biden-era regime	66	48.48%	43.94%	-4.55 pp	7.58%
Later/Trump-era regime	74	58.11%	71.62%	+13.51 pp	27.03%

This is a conditioning diagnostic, not a political-causality claim. The asymmetry is nevertheless large: narrative intervention hurt the earlier-regime backtest but added 13.51 percentage points in the later regime. That result motivated the separate media-regime investigation.

Table 3. Sensitivity to Dual-Scale Media-Regime Normalization

Regime	V34 hit rate	Dual-scale V35 hit rate	V35 vs V34	Decisions changed	Mean media-scale gap
Earlier/Biden-era regime	43.94%	45.45%	+1.52 pp	1.52%	0.000
Later/Trump-era regime	71.62%	68.92%	-2.70 pp	13.51%	0.470

Allowing narrative intensity to be evaluated on both global and regime-local scales changed relatively few decisions, but the changes were concentrated in the later regime. The result supports treating the media measurement process as a possible regime variable rather than assuming a fixed narrative sensor.

The regime results should be read as evidence of instability in the narrative-to-market mapping, not as evidence that one administration caused superior or inferior market performance. The later-regime gain from narrative intervention is large enough to matter, but the dual-scale test also shows that changing media intensity distributions can alter the apparent evidence set. This makes media-regime monitoring part of model governance rather than a side issue.

The resulting research problem is not adequately represented by a single administration dummy. At minimum, the system may contain interacting economic, political, and media regimes. A variable can have a different observed distribution because the underlying

economy changed, because the political agenda changed, because the media system changed how it describes the agenda, or because all three changed simultaneously.

For forecasting, the implication is conservative: regime-conditioned diagnostics are useful, but they should not be mistaken for political causality. The objective is to identify structural instability in the measurement-to-outcome relationship, not to assign partisan explanations to market performance.

6. What We Stopped Doing

- **We stopped equating model complexity with forecast quality.** Additional models, shock overlays, and diagnostics are useful only when they improve calibration, interpretation, or decisions.
- **We stopped treating every detected shock as forecast-relevant.** A shock can be statistically visible yet too small, too transient, or too historically ambiguous to justify changing the 20-day path.
- **We stopped assuming that one historical event defines a reusable response function.** Sparse events are scenarios and analogues before they are stable parameters.
- **We stopped allowing pooled history to erase exceptional context.** When episode magnitudes and mechanisms differ materially, full-sample averaging can be misleading.
- **We stopped presenting analytical detail merely because it exists.** The forecast interface should foreground information that changes the forecast or the user's interpretation of uncertainty.

This last change is particularly important. The current interface suppresses shock evidence unless it contributes to an explicit narrative override. Detection is not decision relevance.

7. The Current Forecasting Architecture

The present IUVO architecture is intentionally simpler than the earlier multi-model research environment. The research models remain valuable for validation and diagnostics, but the production question is narrower: what is the best disciplined 20-trading-day representation of the market's likely path, and is there enough current narrative evidence to alter it?

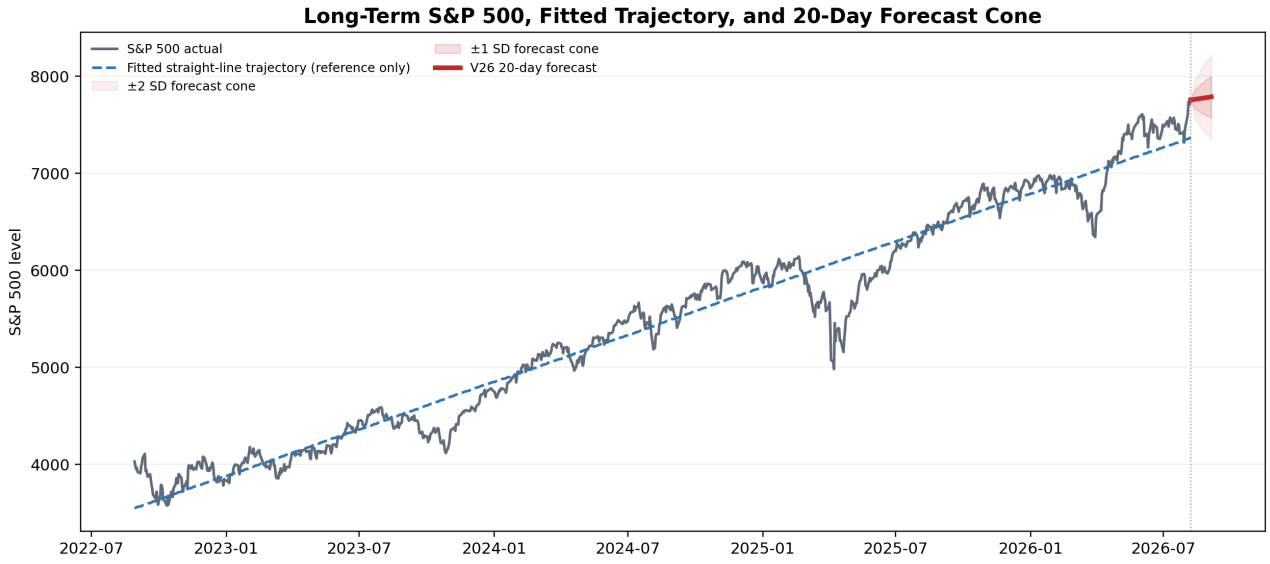
1. Observed market level: begin with the last actual S&P 500 observation.
2. Fitted long-term trajectory: estimate the historical reference path as a straight fitted trajectory rather than implying that the market must immediately assume the fitted value.
3. Anchored 20-day forecast: extend the forecast from the last observed market level, preserving continuity between history and forecast.
4. Forecast cone: express uncertainty around the central path rather than presenting a single deterministic endpoint.
5. Narrative review: evaluate current narrative conditions, regime evidence, historical analogues, and material emerging developments.

6. Narrative override: change the statistical trajectory only when the evidence is sufficiently strong and coherent to justify doing so. Otherwise the reported state is Narrative Override: NO.
7. Human interpretation: explain what would confirm, weaken, or reverse the forecast and identify the developments that deserve attention.

8. Why the Forecast Cone Is the Primary Product

A point forecast invites false precision. A forecast cone makes the model's epistemic status visible. The central trajectory is the system's best current estimate; the surrounding band represents modeled uncertainty. The cone is anchored to the last observed market value because the forecast is an extension of the current state, not a claim that the market will jump immediately to a fitted long-run level.

Figure 1. Long-Term S&P 500 Context with Current 20-Day Forecast Cone



The fitted straight line is a long-term reference only. The red V26 forecast and its uncertainty cone begin at the last observed S&P 500 close; the chart does not imply an immediate move to the fitted trajectory.

The long-term fitted trajectory and the short-horizon cone answer different questions. The fitted line provides structural context: where the market has been moving over the multi-year record. The cone addresses the operational question: given the current observation and current model state, what range of paths is plausible over the next 20 trading days?

Narrative evidence can affect this output in two ways. It can widen uncertainty without changing direction, or—when sufficiently strong—justify an explicit override of the central trajectory. These actions should be disclosed separately. Increased uncertainty is not the same as a changed forecast.

9. Why Narrative Stays in the Model

Removing narrative because individual shock responses are unstable would draw the wrong conclusion from the evidence. The instability is itself information. Narrative is valuable precisely because it can reveal that the environment in which a historical model was estimated is changing.

Table 1. Overall 20-Day Narrative-Override Test

Completed 20-day origins	Trajectory-only hit rate	Narrative-assisted hit rate	Gain	Override rate
140	53.57%	58.57%	+5.00 pp	17.86%

Across 140 completed 20-day origins, the conservative narrative layer improved directional hit rate by 5.0 percentage points while intervening on fewer than one in five origins. This is the principal empirical reason narrative remains in the production architecture.

9.1 Nonlinear Forward-Looking Challenger Test

A separate TreeNet regression was used as a challenger rather than as a replacement production model. The response variable was DEV_CHANGE_20: the change in DEV_NEW over the subsequent 20 trading days. This changes the question from whether narrative explains the market's contemporaneous deviation from trajectory to whether the current narrative field contains information associated with where that deviation moves next.

With the broad predictor set, the challenger produced 59.88% training R-squared and 42.47% test R-squared, with test RMSE of 153.05. After removing redundant raw chronology variables, the deliberately cleaner specification still produced 57.90% training R-squared and 37.87% test R-squared, with test RMSE of 159.08. In that cleaner run, narrative variables became more prominent rather than disappearing.

Table 1A. Nonlinear 20-Day Forward Challenger — Selected Results

Measure	Result
Broad TreeNet	Test R^2 42.47%; RMSE 153.05
Cleaner TreeNet	Test R^2 37.87%; RMSE 159.08
Leading narrative predictors	DOGE 81.9; USAID 63.0; HARVARD 59.8

Measure	Result
Additional state/narrative predictors	PC_2 58.6; Israel/U.S. 53.3; TARIFF 50.3; TRADE 45.2

The partial-dependence results also showed pronounced nonlinear and threshold behavior. DOGE, for example, became progressively associated with more negative subsequent DEV_CHANGE_20 as narrative intensity increased, then saturated at high levels; PC_2 also displayed a strongly nonlinear response. These findings explain why a simple additive shock coefficient can miss useful narrative information.

The experiment did not, however, justify replacing the production forecast with TreeNet. Nor did the added post-Iran regime indicators emerge as dominant predictors once tested directly. Iran-related narrative remained informative, but the evidence did not establish the most recent Iran episode as a single discrete structural break. The more defensible conclusion is that the narrative field contains forward-looking information while its mapping to market behavior remains nonlinear, regime-sensitive, and difficult to reduce to one event or coefficient.

Accordingly, V26 remains the production architecture. TreeNet is retained as a research and validation instrument: it provides independent evidence that narrative contributes information at the same 20-trading-day horizon used by the forecast, while the production model continues to require narrative to earn an override rather than mechanically importing a high-dimensional nonlinear model. Better understanding does not, by itself, require a different forecast.

The statistical layer provides reproducibility, auditability, and resistance to salience bias. It establishes the baseline before interpretation begins. The narrative layer contributes situational understanding: what is different about the current environment, which historical episodes are genuinely comparable, and whether an emerging development is likely to alter the variables on which the statistical relationship depends.

The governance principle is therefore asymmetric. Narrative may challenge the forecast, but it does not casually rewrite it. The default is NO override. The burden of evidence belongs to the intervention. This prevents the interpretive layer from becoming an after-the-fact explanation engine while preserving its value when the current situation is genuinely outside the model's comfortable historical range.

10. Narrative Override as a Formal Decision

A narrative override should be understood as a model-governance event, not an editorial opinion. It means that the current evidence has crossed a threshold at which retaining the unadjusted statistical trajectory would be less defensible than modifying it.

- Materiality: the narrative development is large enough to plausibly affect the 20-day horizon.
- Persistence: the evidence is not a single headline or one-day spike unless the event itself is structurally discontinuous.
- Coherence: multiple indicators or independent narrative channels point in the same direction.
- Historical relevance: analogous episodes, where available, support the interpretation rather than contradict it.
- Regime consistency: the effect is plausible under the current economic, political, and media regime.
- Forecast consequence: the evidence changes direction, slope, or uncertainty enough to matter to the user.

If these conditions are not met, the correct output is not a weakly qualified adjustment. It is Narrative Override: NO. The evidence can still be described in the analytical narrative, but it should not be allowed to masquerade as a forecast change.

11. A More Modest Role for Shock Analysis

Impulse-response functions remain useful research instruments. They help estimate the historical shape, timing, and decay of responses to recurring narrative disturbances. But the four-year experience suggests that they are better used as diagnostic evidence than as automatic production overlays.

Some shocks decay quickly. Others reverse. Some have asymmetric effects. Some are represented by too few comparable episodes to support a stable reusable coefficient. The result is that a screen filled with shock tabs can create an illusion of analytical richness while contributing little to the actual forecast.

The production rule is therefore straightforward: shock evidence should be surfaced prominently when it helps justify an override, materially changes the uncertainty assessment, or provides an unusually strong historical analogue. Otherwise it remains background research.

12. Forecasting as Hybrid Cognition

The four-year experiment reinforces the case for a division of cognitive labor. Quantitative models are strongest where consistency, historical discipline, reproducibility, and probability assignment matter. Human judgment is strongest where forward information, context, novelty, analogy, and causal interpretation matter. The practical question is therefore not merely whether to combine them, but exactly how the combination changes a forecast. The walkthrough below makes those mechanics explicit.

The statistical forecast should be locked before narrative interpretation begins. The interpretive layer should receive the forecast, uncertainty, regime diagnostics, and current

narrative evidence as constraints. Its role is to explain, challenge, and contextualize—not to manufacture a more interesting number.

This architecture is deliberately anti-deterministic. The system does not claim to know what the market will do. It organizes evidence so that users can distinguish the baseline expectation from the conditions that would make that expectation less credible.

12.1 Operationalizing Hybrid Cognition: The Turning-Point Prediction Lab

The hybrid-cognition concept has now moved from principle to an operational prototype in the IUVO HTML system. The production V26 level forecast remains locked and separately visible. Alongside it, the Turning-Point Prediction Lab uses analyst assumptions about forward-looking narrative drivers to test a different question: whether the system is moving toward a state in which the ordinary 20-day forecast deserves less confidence. The driver set includes FEAR, financial stress, geopolitical intensity, disaster, policy/macro conditions, health, and the more specific policy and institutional narratives that proved important in the empirical work, including tariff, trade, DOGE/administrative restructuring, USAID/federal activity, and related variables.

The analyst does not type an arbitrary S&P target. Instead, each slider represents a forward assumption about a narrative driver that may become observable during the forecast horizon. The machine-only precursor score and path probabilities are calculated from information available at time t; the hybrid laboratory then shows how the state assessment, conditional paths, and scenario likelihoods change when the analyst's forward information is admitted at an explicit evidentiary weight. These conditional laboratory paths do not silently replace the locked production forecast; a formal production Narrative Override remains a separately governed decision.

This is the practical hybrid-cognition/Bayesian idea in its current form. The historical model supplies the prior state assessment; the analyst supplies prospective evidence about drivers and causal conditions that are not yet in the data; evidentiary weight governs how strongly that information can alter the turning-point assessment. The current implementation is therefore a Bayesian-governed evidence-integration prototype rather than a fully estimated posterior probability model. Its output is a research precursor index and state classification, not a calibrated probability of a crash.

The design directly addresses a limitation exposed by the perfect-information tests. An analyst may possess credible forward-looking information about a policy announcement, geopolitical escalation, institutional action, or a narrative likely to broaden or persist. IUVO records that information as an explicit conditional assumption rather than historical fact. The useful question becomes: if this information is approximately correct, does the combination of current market vulnerability and anticipated special-cause pressure materially strengthen the case that a turning-point state is developing?

This also clarifies the meaning of override. A model override remains an empirically governed event produced by the production narrative gate. A hybrid turning-point assessment is a separate warning or challenge signal. It can justify closer monitoring, wider uncertainty, or a

formal challenge to the baseline without silently rewriting the S&P forecast. Keeping the level forecast, model override, machine-only precursor score, and analyst-conditioned precursor score separate makes the division between model evidence and human judgment auditable. The next section shows the actual mathematical and graphical mechanics used by the current laboratory.

12.2 Forecast Mechanics: From Baseline to Hybrid Scenario

The mechanics are easiest to understand as three layers that must be kept distinct: (1) a trajectory/market baseline, (2) a machine narrative intervention based on narratives already observed in the data, and (3) a human Narrative Override based on the analyst's judgment about how future narrative conditions will differ from their default expectations. The first two are machine forecasts. The third is the hybrid-cognition layer. Keeping them separate makes the contribution of narrative and the contribution of human judgment observable and testable.

Step 1 — Machine-only level forecast

Let P_t be the last observed S&P 500 level, T_t the fitted long-run reference trajectory, and $D_t = P_t - T_t$ the current deviation from that reference. IUVO forecasts the future change in deviation rather than forcing price directly toward the fitted line:

$$\hat{D}_{t+h} = D_t + \Delta\hat{D}_h$$

The fitted reference is advanced over the same horizon and the two quantities are recombined:

$$\hat{P}_{t+h} = \hat{T}_{t+h} + \hat{D}_{t+h}$$

For the current forecast vintage, the last close is approximately 7,758 and DEV_NEW is −33.2. Because $D_t = P_t - T_t$, the fitted trajectory at the forecast origin is approximately 7,791.2. The 20-day model forecasts a further change in deviation of about −86.4 points, so DEV_NEW at T+20 becomes approximately −119.6. Over those same 20 trading days the fitted trajectory advances to approximately 7,908.6. Recombining the two components gives 7,908.6 + (−119.6) ≈ 7,789. Thus the market can rise modestly from 7,758 to about 7,789 while simultaneously falling farther below its advancing long-run trajectory. In compact form: forecast market change ≈ trajectory growth (+117.4) + change in deviation (−86.4) ≈ +31 points.

Step 2 — Derive the machine narrative intervention

Before any analyst touches a slider, IUVO can measure what observed narrative contributes to the machine forecast. At the same forecast origin, estimate a baseline forecast using the non-narrative market/trajectory predictors X_t and a narrative-assisted forecast using the same predictors plus the observed narrative vector N_t:

$\Delta\hat{D}_{20}^{B} = f(X_t)$ and $\Delta\hat{D}_{20}^{N} = f(X_t, N_t)$

The machine narrative intervention is the difference between those paired forecasts:

$I_{20}^{NARR} = \Delta\hat{D}_{20}^{N} - \Delta\hat{D}_{20}^{B}$

For example, suppose the trajectory/market-only model forecasts a 20-day change in DEV_NEW of −90 points, while the otherwise comparable model containing the narratives already observed at time t forecasts +35 points. The narrative intervention is +35 − (−90) = +125 points. That +125 is not a discretionary adjustment; it is the incremental forecast contribution associated with admitting the observed narrative variables. It is then incorporated into the forecast of DEV_NEW and recombined with the future fitted trajectory in the same way described in Step 1.

This is the meaning of the earlier "Why narrative stays in the model" chart. The trajectory-baseline series and the narrative-assisted series are paired machine forecasts across completed 20-day forecast origins. Their divergence shows how much the inclusion of observed narrative changed the sequence of forecasts. The chart demonstrates incremental forecast contribution; the separate out-of-sample hit-rate tests determine whether that contribution improved forecast performance.

Step 3 — Convert the machine forecast into alternative paths and prior probabilities

The turning-point laboratory asks a different question from the central level forecast. It represents three possible process states over the same 20-day horizon: Normal Continuation, Cascade/Material Deterioration, and Major Abrupt Impulse. Their fitted machine-only likelihoods form the prior probability vector $\pi^{M} = (\pi N, \pi C, \pi I)$. In the current vintage:

Path	Machine-only probability	T+20 level
Normal continuation	99.997%	7,789
Cascade / material	<0.001%	7,711
Major abrupt impulse	0.003%	7,379

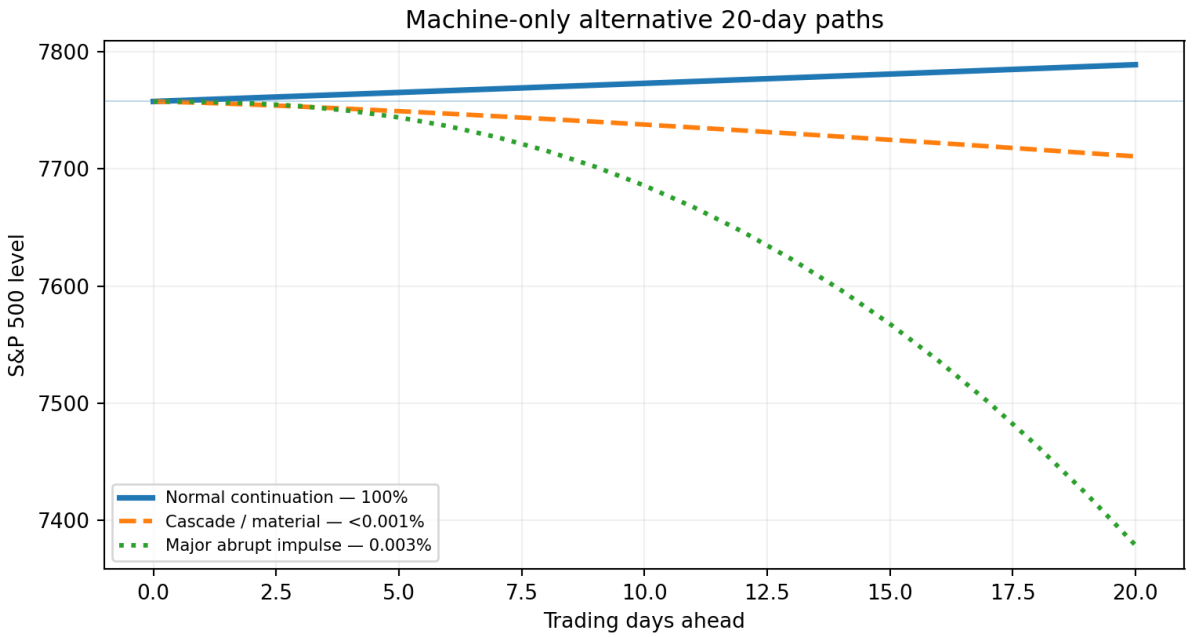


Figure 2A. Machine-only forecast: three alternative 20-day paths branch from the same current close. The probabilities are the fitted path-class likelihoods before analyst intervention.

At this forecast origin, the machine sees little evidence of an imminent nonlinear turning point. The normal path therefore dominates. Importantly, the cascade and impulse paths are still displayed. They make the low-probability alternatives visible rather than allowing a single central line to imply certainty.

Step 4 — Express human judgment as a Narrative Override

Each slider begins at a neutral default: the historically appropriate median or other model-derived expected value for that narrative driver over the forecast horizon. Leaving a slider at that default means the analyst has no forward information that warrants departing from the machine/default expectation. The Narrative Override for driver i is therefore $O_i = N_i^H - N_i^D$, where N_i^H is the analyst's expected future narrative value and N_i^D is the default expectation. If the slider is untouched, $O_i = 0$. The override is scaled against the driver's historical P5/P95 range to form a standardized score s_i: 0 is no override, +1 is a strong stress departure, and −1 is a strong relief departure. The prototype then combines those standardized departures into a signed narrative-pressure score:

$$S = 0.18 \cdot FEAR + 0.18 \cdot Financial + 0.24 \cdot Geopolitical + 0.10 \cdot Disaster + 0.20 \cdot Policy + 0.10 \cdot Health$$

FEAR itself combines the assumed 20-day mean and assumed end-of-horizon change. The coefficients are transparent research weights, not structural causal coefficients. The important point is that S measures departures from default expectations, not raw narrative levels. If every slider remains at its default, S = 0 and there is no human Narrative Override. The machine narrative intervention described in Step 2 remains part of the machine forecast; the analyst has simply chosen not to add prospective information.

Step 5 — Translate the Narrative Override into conditional path displacement

The co-integrating relationship is then used as a scenario-response map. With w denoting the analyst information weight from 0 to 1 and $\sigma_r = 216$ points denoting the historical residual standard deviation used by the prototype, the endpoint displacement is:

$$\Delta P^{NAR} = -S \cdot w \cdot (1.25\sigma_r)$$

This is the point at which the distinction between scenario mechanics and causal inference is essential. IUVO is asking: if the assumed narrative state develops, what market-level displacement would be consistent with the observed long-run relationship? It is not asserting that the narrative variable mechanically causes that displacement.

The displacement is introduced progressively from the common current anchor. The prototype applies 55% of the endpoint displacement to the Normal path, 85% to the Cascade path, and 100% to the Major Impulse path:

$$\hat{P}_j(h \mid H) = \hat{P}_j(h \mid M) + m_j \cdot (h/20) \cdot \Delta P^{NAR}, \quad m_j \in \{0.55, 0.85, 1.00\}$$

Step 6 — Reallocate probability across the three paths

The analyst information also changes the likelihood weights. The laboratory first computes an observed machine challenge index from endogenous market vulnerability and current special-cause evidence. It then computes a hybrid challenge index after admitting the analyst's forward assumptions at weight w. Only the incremental challenge attributable to the analyst is allowed to reallocate probability away from Normal. In compact form:

$$C^H = C^M + w[0.62S^* + 0.18 \cdot \min(E, S^*) - 0.25C^M]$$

where E is the endogenous-vulnerability index and S* is the 0–100 special-cause pressure index. The incremental challenge $\Delta C = (C^H - C^M)/100$ is then mapped to the total turning-point probability, subject to a research cap:

$$\pi^H(\text{Cascade} + \text{Impulse}) = \text{clamp}[\pi^M(\text{Cascade} + \text{Impulse}) + 60\Delta C, 0, 65\%]$$

The turning-point weight is split between Cascade and Major Impulse according to the balance between endogenous vulnerability and special-cause pressure. This is intentionally a transparent research mapping. The machine probabilities remain separately visible; the hybrid probabilities are scenario quantities until prospective validation establishes calibration.

12.3 Worked Walkthrough: No Narrative Intervention versus Narrative Intervention

Case A — No human Narrative Override. Leave every narrative slider at its default value. Each analyst-minus-default difference is therefore zero, so S = 0. This remains true even if the evidentiary-weight control is set above zero: multiplying zero override pressure by any weight still produces zero incremental human adjustment. The machine forecast—including whatever contribution the observed narrative already made through the machine narrative

intervention—stands unchanged. The plotted paths and likelihoods therefore remain the machine-only values shown in Figure 2A. The decision implication is MAINTAIN.

Case B — Illustrative human Narrative Override. Now suppose the analyst possesses forward information that is not yet represented in the observed narrative data. To make the mechanics visible, move FEAR toward its historical stress extreme, move the remaining stress-oriented drivers toward their P95 stress values, and assign 70% evidentiary weight. This is an illustrative stress test, not a current forecast. Relative to the default slider values, the standardized departures combine to a signed pressure of approximately S = 1.00. With σ_r = 216 and w = 0.70, the prototype maps that judgment to an endpoint displacement of approximately $\Delta P^{NAR} = -1.00 \times 0.70 \times (1.25 \times 216) \approx -189$ points before the path-specific multipliers are applied. The machine challenge index is approximately 39; after admitting the analyst's prospective evidence, the hybrid challenge rises to approximately 78 and enters CHALLENGE territory.

Path	**Machine-only**	**Hybrid stress probability**	**Hybrid T+20 level**
Normal continuation	99.997%	76.6%	7,685
Cascade / material	<0.001%	5.5%	7,550
Major abrupt impulse	0.003%	17.9%	7,190
Any turning-point path	0.003%	23.4%	—

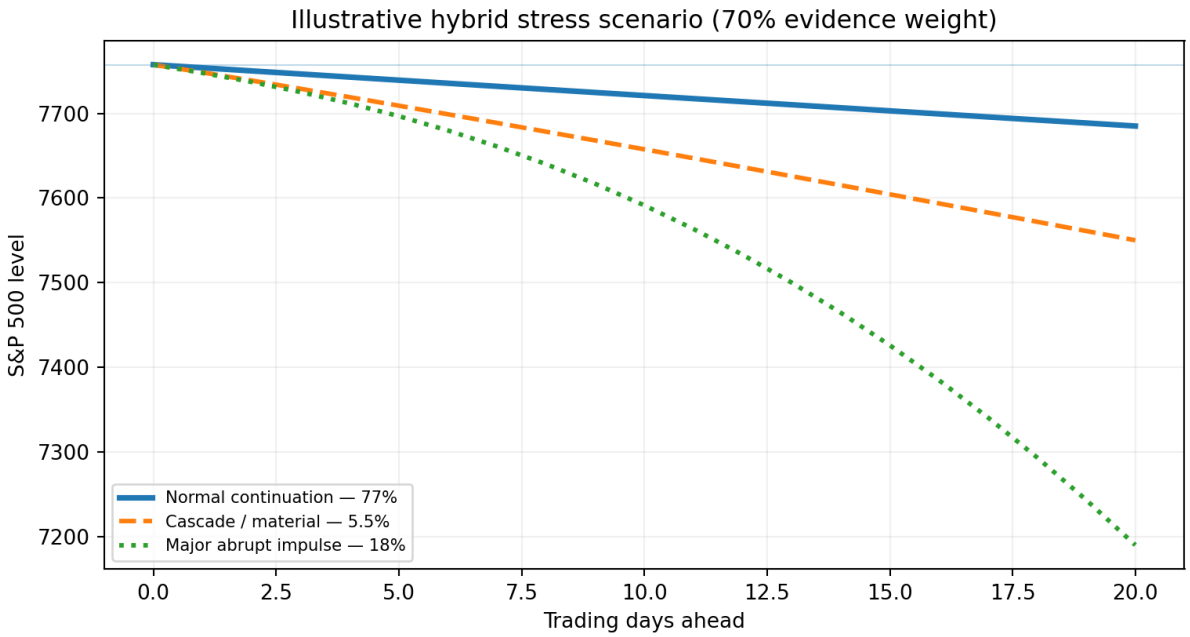


Figure 2B. Illustrative hybrid stress scenario. The same three paths are conditioned on explicit analyst assumptions at 70% evidentiary weight. Both the conditional path levels and their relative likelihoods change; the historical record and common starting point do not.

The figure shows why the laboratory is more useful than simply asking an analyst for a revised S&P target. The analyst never supplies the market answer. The analyst supplies forward information about conditions—policy, geopolitical risk, financial stress, fear, disaster, or health—that may not yet be present in the machine's data. The machine then performs the translation consistently and exposes exactly how much the judgment changed the result.

The complete sequence is therefore: trajectory/market baseline → observed machine narrative intervention → machine path probabilities → human Narrative Override → confidence weighting → conditional path displacement and probability reallocation. The term "intervention" refers to what observed narrative contributes inside the machine model. The term "override" refers specifically to the analyst's prospective departure from the default narrative assumptions. This terminology is used consistently throughout the remainder of the paper.

12.4 How the Sliders Create Useful Hybrid Cognition

The sliders are an interface for prospective evidence. Their default positions represent the model's neutral expectations for the future narrative drivers; they are not zero-information blanks. Leaving them alone means, "I have no reason to depart from the default." Moving a slider means, "I expect this narrative condition over the next 20 trading days to differ from the default by this amount." The evidentiary-weight slider answers a separate question: "How much confidence should IUVO place in my departure from the default?" This separation prevents a dramatic scenario assumption from automatically receiving high credibility.

Human contribution = Analyst expected narrative − Default expected narrative. If that difference is zero across all drivers, the hybrid result collapses to the machine result. If it is nonzero, IUVO—not the analyst—translates the stated narrative differences into conditional market paths and revised scenario likelihoods.

This produces a useful division of labor. The machine preserves the historical baseline, scales assumptions against prior experience, computes the same transformation every time, and records the prior and posterior-like scenario weights. The human contributes information that is inherently difficult to infer from historical data alone: a likely policy action, an anticipated geopolitical escalation, a judgment that several signals share a mechanism, or recognition that a current situation is novel enough to weaken an analogue. Neither contribution is hidden inside the other.

Operational use follows a simple discipline. First, lock and save the production forecast and machine-only path probabilities. Second, document the analyst's forward assumptions and their source. Third, move only the sliders for which genuinely prospective information exists and set the evidentiary weight to reflect its credibility. Fourth, compare the hybrid paths and probabilities with the machine prior. Fifth, use the size and architecture of the change to determine whether to maintain the baseline, watch more closely, widen scenarios, or formally challenge the production forecast. Finally, save the entire vintage before the outcome is known.

The last step is essential. If the analyst-conditioned forecast is not preserved prospectively, hybrid cognition can easily become retrospective storytelling. If every machine prior, slider assumption, evidence weight, hybrid probability, conditional path, and realized outcome is saved, the research can test whether human foresight actually improves lead time, turning-point discrimination, calibration, and false-alarm performance. Hybrid cognition then becomes a falsifiable forecasting hypothesis rather than a design philosophy.

13. Implications Beyond Financial Markets

The broader organizational proposition of Perceptual Macroeconomics remains intact. Any organization that produces substantial text before measurable outcomes occur has a potential narrative forecasting problem. Customer complaints precede churn. Sales-call language precedes win or loss. Project status narratives precede schedule and cost outcomes. Employee communications precede some forms of operational deterioration.

The four-year market experiment adds an important warning to those applications. Organizations should not assume that a text-outcome relationship is timeless merely because it was predictive during model development. Vocabulary changes. Management changes. Incentives change. Reporting practices change. The population generating the text changes. A successful system must therefore monitor not only prediction error but also changes in the narrative regime itself.

The practical objective is not to automate judgment away. It is to provide a disciplined baseline, identify departures from that baseline earlier, and make the reasoning behind any intervention explicit.

14. Research Agenda for the Fifth Year

Table 4. 20-Day Known-Driver / Perfect-Information Diagnostic by Regime

Regime	N	Current hit	Known-driver oracle hit	Δ hit	Current MAE	Oracle MAE
Earlier/Biden-era regime	66	48.48%	71.21%	+22.73 pp	126.9	144.1
Later/Trump-era regime	74	58.11%	45.95%	-12.16 pp	194.2	360.4

The oracle experiment assumes future driver values are known, but still routes them through an estimated forecasting relationship. Its mixed results are therefore informative: perfect knowledge of inputs is not the same as perfect knowledge of the market response function. The later regime actually deteriorated under the oracle specification, reinforcing the importance of regime stability, model form, and judgment.

- Continue the weekly or monthly rolling forecast process using a fixed 20-trading-day horizon and preserve each forecast vintage for genuine out-of-sample evaluation.
- Measure directional hit rate, endpoint error, cone coverage, and override performance separately.
- Track whether narrative overrides improve results relative to the unadjusted statistical baseline; overrides must earn their complexity empirically.
- Use nonlinear models such as TreeNet periodically as challenger and feature-discovery models, not as automatic replacements for the production forecast. Promote additional complexity only when rolling out-of-sample evidence demonstrates incremental forecast value.
- Maintain regime-conditioned diagnostics while avoiding causal political interpretation not supported by the design.
- Monitor source composition and media-scale changes so that apparent narrative shifts can be distinguished from measurement-system shifts.
- Treat new shock topics as watchlist variables until enough comparable history exists to support formal response estimation.
- Re-estimate long-run relationships periodically, but avoid unnecessary model churn between scheduled review points.
- Preserve the narrative briefing because explanatory value is part of forecast usefulness even when the narrative does not alter the numerical path.

Extend the fifth-year research program beyond market forecasting by using the same disciplined workflow to study organizational datasets in which structured operating measures and text narratives coexist, with particular attention to probabilistic model behavior, temporal validity, maintenance burden, and whether narrative evidence improves practical business decisions.

- Treat the contemporary fitted trajectory as a governed baseline: re-estimate it at scheduled review points and test whether a persistent change in realized growth justifies rebasing, rather than retaining the pre-COVID rate by assumption.
- Operationalize the common-cause/special-cause hypothesis by separating persistent and recurring narrative state variables from episodic/extreme narrative configurations, then test whether the separation improves forecast calibration, cone coverage, and override performance.
- Test whether narrative contributes in both roles: improving estimates of ordinary variation around the current baseline and identifying special-cause conditions under which that baseline should be challenged.
- Evaluate the Turning-Point Prediction Lab prospectively by saving the locked production forecast, the machine-only precursor score, the analyst-conditioned precursor score, the assumptions entered, and the information weight at each forecast origin. After the horizon closes, score whether analyst-supplied forward information improves turning-point discrimination, lead time, calibration of the precursor index, false-alarm performance, and recognition of special-cause conditions.

15. Conclusion: Four Years Made the Model Less Certain—and Better

After four years, the strongest evidence for Perceptual Macroeconomics is not that narrative can replace conventional forecasting. It is that forecasting improves when statistical structure and narrative interpretation are allowed to do different jobs.

Statistics establishes the baseline, trajectory, uncertainty, and historical discipline. Narrative identifies changes in the environment that the historical model may not yet understand. Historical analogues establish what has happened before, not what must happen again. Analyst judgment evaluates the difference.

The 20-day nonlinear challenger adds an important empirical qualification: narrative is not only useful for contemporaneous explanation. In the present sample, narrative variables also retain material association with subsequent movement in deviation from trajectory. Yet the same experiment reinforces the case for restraint: discovering predictive structure is not equivalent to proving that a more complex production model will forecast better. The production architecture therefore remains intentionally simpler than the research environment.

IUVO therefore does not attempt to eliminate judgment from forecasting. It attempts to make judgment better informed, more explicit, and harder to exercise casually. The narrative

override makes that principle operational: the statistical forecast stands unless the current evidence earns the right to change it.

Four years of observation have made the system less certain, not more—and that is progress. The market is not a deterministic machine waiting for the correct equation. It is a probabilistic system populated by people responding to changing information, changing narratives, changing institutions, and one another.

The forecasting problem is not to make that uncertainty disappear. It is to recognize when the evidence has changed enough that our expectations should change with it.

The baseline itself is now part of that discipline. IUVO no longer treats the pre-COVID growth rate as a permanent truth. The contemporary fitted trajectory reflects the market process observed over the current record. That change embodies the first common-cause lesson: a persistent shift can cease to be an exception and become the process against which future exceptions must be judged. The narrative results then supply the second lesson. Recurring narrative variables contain information about the ordinary state around that process, while unusually intense or novel narrative configurations sometimes identify conditions under which the process forecast should be challenged. The 5-point overall gain from selective narrative intervention, the much larger later-regime gain, the nonlinear forward-looking TreeNet results, and the mixed perfect-information results all point in the same direction: narrative contains both state and exception information, but the mapping is probabilistic and regime-sensitive.

The Turning-Point Prediction Lab makes that philosophy operational. The statistical 20-day level forecast remains the governed baseline. The machine separately estimates whether the current configuration resembles a precursor state, and the analyst can then state forward-looking beliefs about important drivers and assign those beliefs evidentiary weight. IUVO therefore asks two different questions: 'What does the historical model forecast for the next 20 trading days?' and 'Given the current vulnerability plus what we credibly expect about developing drivers, how strongly should we suspect that the ordinary forecast is approaching a major turning point?' That is the practical hybrid-cognition objective: combine machine consistency with human foresight without concealing which contribution came from which source.

16. A Second Objective: What This Experiment Says About Machine Learning in Business

A second objective has emerged from the four-year experiment: to use the forecasting process as a practical laboratory for understanding how applied machine learning behaves when it is brought into an ordinary business setting. The financial-market application is useful precisely because it compresses many of the problems organizations encounter with their own data—time dependence, changing regimes, incomplete causal knowledge, text narratives, nonlinear relationships, model drift, and the need to make decisions under uncertainty.

The experience challenges a common simplification in discussions of enterprise AI: that the principal obstacle is poor data quality and that, once organizational data are cleaned, harmonized, and governed, useful AI follows naturally. Clean data are necessary, but they are not sufficient. A dataset can be internally consistent and still be weakly predictive, temporally unstable, structurally incomplete, or no longer representative of the environment in which the model must operate.

Applied machine-learning models are probabilistic systems. They do not discover a permanent deterministic rule hidden inside a clean dataset. They estimate relationships that are conditional on the observations, variables, regimes, measurement practices, and time period available during development. Those relationships can be nonlinear, interaction-heavy, threshold-dependent, and unstable across time. The TreeNet challenger in this study illustrates the point: substantial forward-looking information was detectable, but it was distributed across temporal structure, latent components, and multiple narrative variables, with partial-dependence functions that often departed sharply from simple linear assumptions.

This has direct implications for business analytics. Model development is analysis-heavy. It requires more than selecting an algorithm and supplying a prepared table. Analysts must define the outcome correctly, prevent leakage from future information, distinguish contemporaneous explanation from genuine forecasting, examine nonlinear response surfaces, identify redundant temporal variables, test alternative specifications, and determine whether apparent structural breaks are supported by the evidence. The model must then be evaluated out of sample and compared with a simpler baseline before its additional complexity is allowed into production.

Text narratives deserve particular attention in organizational applications. Much of the information that precedes business outcomes does not first appear in a structured numeric field. It appears in project status reports, meeting notes, customer complaints, service records, sales-call summaries, risk registers, executive commentary, employee communications, supplier discussions, and other language generated while events are still unfolding. These narratives can contain early evidence of changing expectations, emerging constraints, causal hypotheses, sentiment, and operational conditions that conventional transactional datasets may register only later.

The implication is not that organizations should indiscriminately feed all available text into a large model. Narrative data require the same discipline as quantitative data. Vocabulary changes, reporting incentives change, management teams change, and the meaning of a phrase can shift across organizational regimes. Text measures therefore require provenance, stable definitions where possible, temporal validation, and periodic reassessment. A narrative variable that was predictive last year may become commonplace, disappear, or acquire a different meaning this year.

The four-year experiment also argues for separating research complexity from production complexity. Machine learning can be extremely valuable as a challenger, feature-discovery mechanism, and diagnostic instrument even when the final production model remains

simpler. In this study, TreeNet helped establish that narrative variables contained material information about subsequent 20-trading-day movement and exposed nonlinear threshold behavior. That insight improved our understanding of the system without compelling us to replace the more governable V26 forecasting architecture.

For business users, this distinction is important. The value of AI is not measured by how much model complexity an organization can deploy. It is measured by whether the system improves understanding, anticipation, and decision quality at an acceptable level of operational and governance burden. A sophisticated model that cannot be explained, monitored, refreshed, or challenged may be less useful than a simpler model embedded in a disciplined analytical process.

Maintenance is therefore part of the model, not an activity performed after the model is finished. Organizations should expect to monitor forecast error, directional performance, calibration, data distributions, narrative composition, regime changes, and the continued usefulness of individual predictors. They should also preserve model vintages and forecast histories so that performance can be evaluated genuinely out of sample rather than reconstructed after outcomes are known.

The broader lesson is that enterprise AI should be treated as an ongoing probabilistic decision system rather than as a one-time data-cleaning or automation project. Clean data establish a foundation. Machine learning estimates conditional relationships. Text narratives add context that structured data may miss. Human judgment defines objectives, recognizes novelty, evaluates whether a model remains applicable, and decides when additional complexity has earned its place. The objective is not to remove uncertainty or judgment, but to make both visible, testable, and better informed.

A further objective follows from the common-cause/special-cause distinction. In business datasets, text may describe both the ordinary operating system and departures from it. Routine CRM notes, project reports, service records, and meeting summaries can encode recurring conditions that explain normal variation; unusual combinations, abrupt changes in language, or genuinely novel topics may signal a special cause. The practical machine-learning question is therefore not merely whether text adds predictive accuracy, but whether text helps the organization tell the difference between “the process is varying normally” and “something has changed that makes the old baseline less trustworthy.”

17. Three Market Anomalies: What Did the Narratives Know?

The long-term S&P 500 chart makes the practical question visible even to a reader with no background in econometrics. The market generally followed a rising fitted trajectory, but three conspicuous departures stand out in the period examined here: the July–October 2023 correction, the February–April 2025 selloff, and the January–March 2026 downturn. A business reader would recognize the same pattern in another setting as an unexpected fall in sales, a project suddenly slipping, customer attrition accelerating, or margins moving sharply away from plan. The useful questions are therefore ordinary ones: What happened? Was

there evidence in the narratives that helped describe what was happening? And could some of the deterioration have been anticipated?

The case studies below are descriptive, not causal claims. Narrative variables were compared with their own recent histories so that a large value means a topic became unusually prominent relative to its prior level. A narrative becoming prominent during a downturn does not prove that it caused the downturn. It does show what information environment surrounded the anomaly and gives us a disciplined basis for asking whether text contained information that conventional price and economic variables alone did not capture.

17.1 July–October 2023: Geopolitical Risk Arrives During an Existing Correction

The S&P 500 fell about 10% from its late-July peak to its late-October trough. The narrative record during the broader correction became dominated by Israel/Hamas/Iran-related discussion, with additional elevation in Yemen/attack, terrorist-attack, government-shutdown, election, wildfire/FEMA, and related risk narratives. The timing matters: the Israel–Hamas war began late in this market correction, so those narratives are much more useful for describing the changing information environment during the decline than for claiming that they predicted the correction from its July peak.

In plain language, the text stream shows a market environment that became increasingly crowded with geopolitical and institutional uncertainty while prices were already moving below their prior path. This is an example of why narrative analysis is valuable even when it does not provide a clean advance warning: it can tell the analyst that the context surrounding an anomaly has materially changed and that a model calibrated to quieter conditions may deserve additional scrutiny.

17.2 February–April 2025: Policy, Tariffs, DOGE, Trade and Institutional Change

The largest of the three episodes was the decline from the February 2025 peak to the April trough, roughly 19%. Here the narrative evidence is considerably more coherent. Tariff discussion was exceptionally elevated, while DOGE, federal-worker, executive-order, trade, Trump–Zelensky–Putin, mass-deportation and other new-administration narratives also rose sharply relative to their prior histories. Some highly episodic topics also registered large statistical spikes; those are useful as indicators of an unusually dense news environment but should not automatically be interpreted as economic drivers.

This episode is especially important because several of the same variables later appeared among the strongest predictors in the 20-trading-day TreeNet model. DOGE became one of the most important forward-looking narrative variables and displayed a pronounced nonlinear relationship with subsequent movement in deviation from trajectory. Tariff and trade narratives also remained prominent. The result does not establish that these topics mechanically caused the selloff. It does show that the text environment was registering policy and institutional change that was relevant to what happened next.

17.3 January–March 2026: Geopolitical and Policy Narratives Converge

The 2026 downturn was smaller than the 2025 episode but still substantial, approximately 9% from the January peak to the late-March trough. The most unusual narratives during the episode included attacks on Yemen, Israel/U.S.–Iran-related discussion, Greenland, oil-production, DHS, migration and other geopolitical or policy themes. The Iran-related structural-break variables we tested did not, by themselves, improve the forward TreeNet model enough to justify declaring a single Iran event the cause of a new statistical regime. That negative result is important.

A better interpretation is that the downturn occurred inside a multivariate narrative environment. Several geopolitical, energy and policy narratives were elevated simultaneously. Iran was part of that environment, but the data did not support reducing the entire episode to one event or one binary regime indicator. This is precisely where text analytics can improve judgment: it permits the analyst to see the configuration of concerns rather than forcing a complex period into a single explanatory label.

The three observed downturns illustrate why this matters. A predominantly endogenous episode can emerge from elevated market geometry and repetitive cycling without a broad narrative storm. A perfect-storm episode can combine endogenous vulnerability with broad policy and institutional disturbance, as in the 2025 configuration involving tariffs, DOGE, federal-worker/USAID activity, executive orders, trade and deportation-related narratives. An exogenous episode can arise from a more ordinary market state struck by concentrated geopolitical or policy pressure, as in the 2026 configuration involving Israel/U.S.–Iran, attacks on Yemen, oil production, DHS/migration, Greenland and related themes. The same narrative level therefore need not imply the same market consequence in every state.

17.4 Hybrid Cognition: Where Human Judgment Can Add Predictive Information

The analyst interface should consequently evolve from asking, ‘How many S&P points would this slider move the endpoint?’ to asking, ‘If these driver conditions develop over the next 20 trading days, how much does the evidence for a major-downturn state change?’ The driver inputs remain useful, but their purpose changes. FEAR, financial stress, geopolitical intensity, disaster, policy/macro conditions, health, tariffs, trade, DOGE/administrative restructuring, USAID/federal activity and related narratives become evidence about the probability or severity of a precursor architecture rather than direct point adjustments to the market forecast.

This suggests a two-forecast discipline. First calculate a machine-only turning-point assessment using information actually available at time t. Then calculate a hybrid assessment after adding the analyst's explicit forward assumptions. Both forecasts should be saved before the outcome is known. Over time the research can therefore test whether human forward-looking judgment improves turning-point discrimination, calibration, lead time, or false-alarm performance. Hybrid cognition itself becomes an empirical hypothesis rather than a philosophical assertion.

The machine contribution is measurement and consistency: deviation from trajectory, persistence above or below the reference path, cycling behavior, narrative breadth, extremity, concentration, acceleration, and historical analogues. The human contribution is foresight and causal abstraction: anticipated policy actions, geopolitical escalation or de-escalation, institutional decisions, expected persistence of an emerging narrative, recognition that apparently independent signals are causally linked, and judgment that a current situation is novel enough to weaken historical analogues.

The precursor architecture gives hybrid cognition a more precise role. Human judgment should not be used primarily to guess the future S&P 500 level. It should be used to estimate forward information about the drivers and causal conditions that the historical dataset cannot yet observe. An analyst may have credible information that tariff intensity will rise, that trade-policy uncertainty will broaden, that geopolitical escalation is becoming more likely, or that several apparently separate narratives are manifestations of one underlying mechanism. Those judgments can be encoded explicitly, assigned evidentiary weight, and evaluated against the machine's current state diagnosis.

17.5 Forecasting Implication: Predict the Turning-Point State, Not the Index Level

This is deliberately a conditional-state model rather than a crash-prediction claim. The present sample contains too few major downturns to support a stable fitted probability model. The fifth-year research task is therefore to score these precursor states prospectively, preserve every vintage, and test whether elevated vulnerability, exogenous special-cause exposure, and perfect-storm configurations improve discrimination of subsequent major turning points relative to the baseline forecast alone.

The practical forecasting implication is to separate the ordinary 20-day level forecast from a second question: is the system approaching a major turning-point state? The proposed architecture is Baseline State → Latent Vulnerability → Special-Cause Exposure → Forecast Consequence. The baseline and market geometry establish whether the system is already vulnerable; the narrative field indicates whether broad or concentrated special-cause pressure is emerging; and the combination determines whether the ordinary forecast deserves to be trusted, watched more closely, or challenged.

The phrase 'predict the turning-point state' is intentionally different from predicting the S&P 500 level. The ordinary V26 forecast still estimates the market's expected 20-day path and uncertainty from the current close. The turning-point model asks whether the conditions under which that level forecast was generated are becoming unstable. It therefore predicts a state of vulnerability rather than a price: whether latent market geometry, narrative breadth and acceleration, and concentrated special-cause exposure are combining in a way that historically deserves heightened attention.

Operationally, the assessment has four stages. Baseline State measures where the market is relative to its contemporary fitted trajectory. Latent Vulnerability measures endogenous warning features such as persistent deviation, repeated mini-cycles, loss of momentum, and other geometry that can make the system susceptible even before a dramatic narrative

appears. Special-Cause Exposure measures whether narrative intensity, breadth, acceleration, concentration, or novelty indicates an external or regime-changing pressure. Forecast Consequence combines those pieces into a governed classification—maintain, watch, or challenge—without pretending that the present five-event research sample can support a stable crash probability.

The distinction matters because the three downturns did not arise from the same configuration. The 2023 correction appears to have contained more endogenous vulnerability before geopolitical narratives became dominant; the 2025 selloff is the clearest perfect-storm case, with market vulnerability coinciding with unusually broad policy and institutional narratives; and the 2026 downturn is better interpreted as concentrated exogenous geopolitical and policy pressure interacting with the market's existing state. A useful warning system therefore must recognize configurations, not search for one universal trigger.

Human judgment enters one step earlier than the market forecast. The analyst is not asked to guess the index. The analyst is asked to estimate information the historical dataset cannot yet observe—for example, whether tariffs are likely to broaden, whether geopolitical escalation will persist, whether several apparently separate narratives share one causal mechanism, or whether a novel event makes the historical analogue unreliable. The hybrid score records how much those forward assumptions change the machine-only state assessment. Saving both versions before the outcome is known makes the value of human judgment testable.

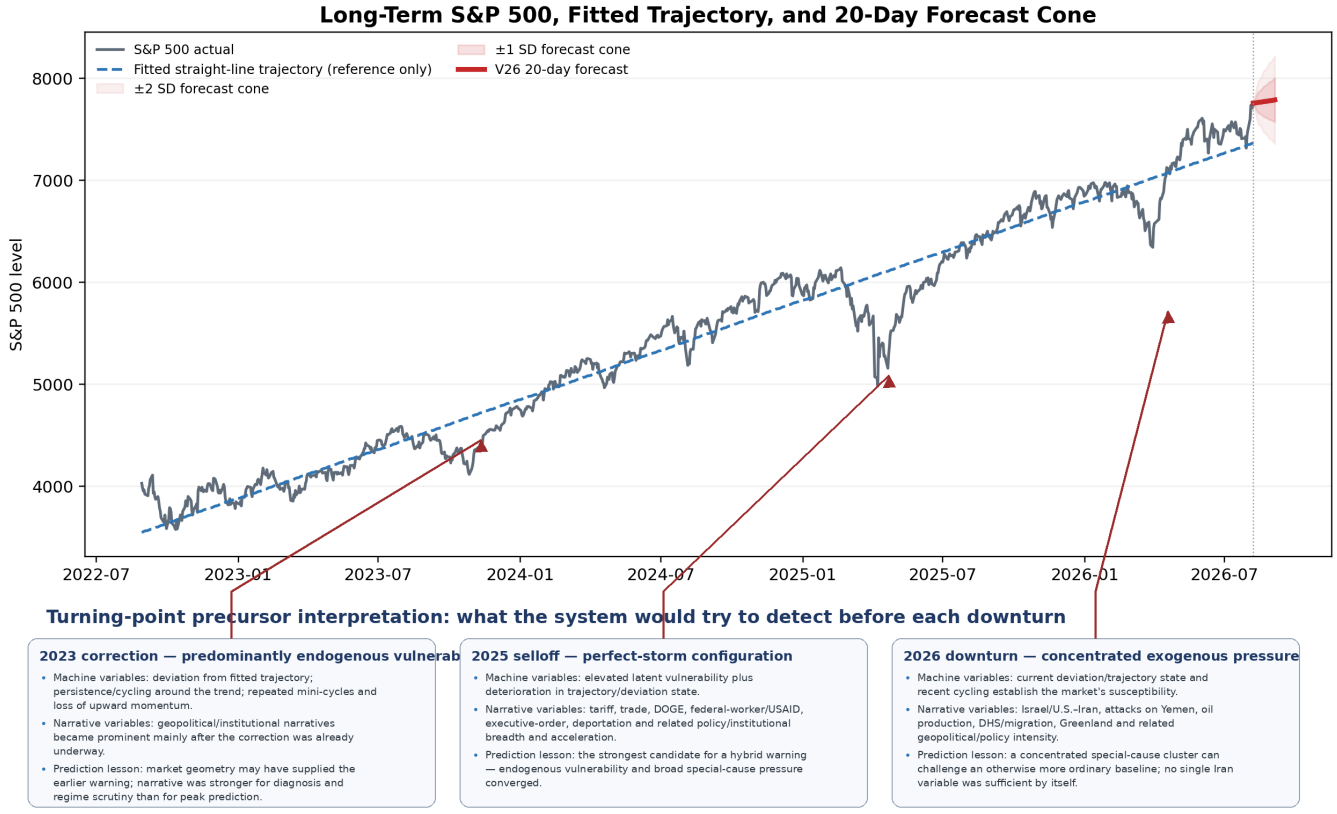


Figure 2. The same long-term S&P 500 plot, annotated with the precursor variables and hybrid-cognition interpretation associated with the three major downturns.

18. Explanation Is Not Prediction—and Perfect Information Clarifies the Difference

The three episodes show that narratives can help describe abnormal periods. The harder question is whether they could have helped us see those periods coming. Retrospective explanation and genuine forecasting are not the same task. A narrative that becomes prominent after a decline begins may be highly useful for diagnosis while offering little advance warning. That is why the research changed the response variable from contemporaneous deviation to the change in deviation 20 trading days ahead.

The forward-looking machine-learning tests found material out-of-sample predictive information. In the fuller TreeNet specification, test R-squared was about 42%, and even after redundant time variables were removed the test R-squared remained about 38%. Narrative variables such as DOGE, USAID, Harvard, Israel/U.S.–Iran, tariff, trade and other topics remained prominent. These results do not imply that 38–42% of future market movement is 'known.' They mean that, for the specific 20-day deviation-change target used in this experiment, the model explained a meaningful portion of out-of-sample variation while substantial uncertainty remained.

The perfect-information thought experiment asks a different question. Suppose the analyst did not know the future market value but somehow knew in advance what the important drivers would look like during the coming 20 trading days. How much forecast uncertainty would remain? This is an expected-value-of-perfect-information style question: it measures the potential value of better information about the drivers, not a claim that such information is actually available in advance.

For a business reader, the distinction is crucial—and the experiment produced a concrete answer. Perfect knowledge of the selected future drivers was not equivalent to perfect knowledge of the market. In the regime-conditioned test, the earlier/Biden-era known-driver specification improved directional hit rate from 48.48% to 71.21%, a gain of 22.73 percentage points, although mean absolute error increased from 126.9 to 144.1. In the later/Trump-era regime, the same idea moved in the wrong direction: hit rate fell from 58.11% to 45.95% and MAE rose from 194.2 to 360.4. The current scenario-lab validation tells the same general story: over the full 20-day sample the known-driver oracle improved directional hit rate only modestly, from 53.57% to 57.86%, while MAE worsened from 162.5 to 209.9; in the later holdout, hit rate fell from 73.08% to 57.69% and MAE rose from 178.3 to 291.6.

The implication is more informative than either success or failure alone. Better information about future drivers can be highly valuable when the estimated response function is reasonably stable, but it can make inference worse when the mapping between drivers and outcomes has changed. This is exactly why the Turning-Point Prediction Lab does not treat analyst slider settings as truth or mechanically convert them into S&P points. They are weighted assumptions used to challenge the current state diagnosis, while the production forecast remains separately governed.

Viewed through the common-cause/special-cause lens, the oracle results also suggest that knowing the magnitude of a driver is not enough if we do not know which process is currently generating the response. A future narrative level may be common under one regime and special under another. The value of analyst judgment is therefore not merely estimating a future driver level; it is judging whether that driver represents ordinary variation, an external special cause, or one component of a broader perfect-storm configuration.

19. What This Means for CRM and Other Business Text

The business analogy is direct. Organizations already possess narratives of their own: CRM notes, sales-call summaries, customer emails, service tickets, project status reports, risk registers, meeting minutes, supplier communications and executive commentary. Structured systems may accurately record revenue, contract dates, open opportunities, defects, schedule variance or service incidents while leaving much of the changing context surrounding those numbers trapped in text.

Consider a customer that eventually churns. The structured CRM may show an active account until very late in the process. The text may have been changing for months: repeated implementation complaints, references to a competitor, a new CFO reviewing vendors, declining enthusiasm in sales calls, requests for unusual concessions, or repeated unresolved service issues. No single sentence need predict churn. The potentially useful signal is the changing configuration of narratives before the structured outcome is recorded.

Modern language models make it practical to convert portions of that text stream into measurable variables—topics, concerns, sentiment, events, expectations and combinations of signals—and then use conventional statistical or machine-learning methods to test whether those variables explain or predict outcomes. The lesson from this market experiment is encouraging: text can contain material information that is absent from the conventional numeric record. But the lesson is equally cautionary: the resulting models are probabilistic, nonlinear, analysis-heavy and subject to drift.

This changes how an organization should think about 'AI readiness.' Cleaning and harmonizing data remain valuable, but data cleanliness is not the same thing as predictive usefulness. Before a company spends heavily cleaning every field in every system, it may be wiser to define a consequential decision, assemble a representative sample of the structured and narrative data relevant to that decision, and test whether a model can actually improve explanation or prediction. If it cannot, more cleaning alone may not solve the problem.

The market experiment also suggests that machine learning can create value without becoming the final production decision engine. A complex model can serve as a challenger, discover nonlinear relationships, identify candidate narrative variables, reveal thresholds, and test whether information exists. A simpler production model can then be retained when it is easier to govern, monitor and explain. Complexity should earn its way into production.

Finally, text models require maintenance just as quantitative models do. Organizational vocabulary changes. Products change. Managers and customers change. Reporting

incentives change. A phrase that once signaled serious risk may later become routine language. The organization therefore has to monitor not only model error but also the composition and meaning of the narrative stream itself.

The practical message for the non-specialist is straightforward: Clean data tell us accurately what has been recorded. Narrative data may help tell us what is developing. Machine learning can test whether either helps us understand what happens next. The opportunity is not to eliminate human judgment, but to give judgment earlier and better evidence.

Source Note

This anniversary addendum is based on the empirical and methodological record documented in Perceptual Macroeconomics: Narrative Co-Integration, Regime Dynamics, and the Limits of Federal Reserve Data in Asset Price Forecasting — An Introduction to the IUVO™ Forecasting System (2026), including its post-publication quality review, hybrid cognition architecture, narrative periodicity analysis, impulse-response analysis, and subsequent production-model refinements.