

MIXSHIFT INTELLIGENCE SERVICE

RAG-Enabled Hierarchical Causal Attribution Model

A Deterministic Structural-Attribution Oversight Layer for AI-Assisted
Decisions

*How structured performance attribution, governed retrieval, and evidence-constrained AI preserve
human control over complex business systems*

TECHNICAL WHITEPAPER

August 2026

Todd VanderStelt

MixShift

Patent pending · U.S. Patent Application No. 19/070,768

Contents

Abstract

Executive summary

1. The accountability gap in automated decision systems
2. Why AI increases the need for oversight
3. What HCAM is and what 'causal' means here
4. The HCAM methodology
5. From calculation to an AI-ready evidence system
6. Governing the AI that communicates the analysis
7. Novel contribution and internal validation
8. Integrated case: explaining a forecast miss with 'ACME'
9. Application patterns
10. Portability beyond Amazon commerce
11. Limitations and research boundaries
12. Conclusion: from explainable analytics to accountable automation

Appendix A. Worked H-Bridge and Shapley comparison

Appendix B. Evidence and claim matrix

References

Publication note

About the author

Abstract

As AI moves from drafting and search into analysis and action, organizations face an adoption paradox. They want lower workloads, faster decisions, and insights that would be impractical to produce manually. Yet those gains disappear when an accountable person must reconstruct every material analysis before trusting it. The barrier is not only model capability. It is the absence of an inspectable evidence path between source data, automated conclusion, and human decision.

Language models are powerful synthesizers, but generation alone **does not guarantee exact multi-step computation** or fidelity to retrieved evidence [7, 8]. In business analysis, the resulting risks include inconsistent aggregation, invalid ratio calculations caused by mismatched or incomplete bases, missing scope, unsupported causal language, and conclusions that do not reconcile to governed data. The cost is both risk and unrealized value: organizations either accept answers they cannot adequately inspect or confine AI to work whose errors matter less.

MixShift addresses this problem through its Intelligence Service: a governed decision-support architecture that keeps four functions distinct but composable. HCAM supplies deterministic attribution. Brand Context and mechanism evidence supply business meaning. Forecast models establish expected baselines; intervention and counterfactual models estimate actions and alternative outcomes under their own assumptions. Identity-resolved retrieval, skills, claim contracts, validators, and human review govern how AI uses that evidence. Remove any one of these functions and trust degrades: exact math without context can be irrelevant, context without attribution can be anecdotal, models without evidence boundaries can be overstated, and evidence without execution controls can be recombined into an unsupported conclusion.

This paper focuses on the RAG-enabled Hierarchical Causal Attribution Model because it is the analytical spine of that architecture. HCAM resolves the deterministic part of the problem before AI is asked to reason or write. Horizontal Bridges quantify through which declared metric relationships a result moved. Vertical Bridges locate where that movement manifested across business dimensions, groups, and entities. The outputs are packaged as scoped, replayable evidence artifacts with explicit coverage, caveats, provenance, and reconciliation checks. The Intelligence Service supplies those artifacts and companion evidence to task-specific skills. The language model synthesizes the evidence; it does not recreate the attribution.

HCAM provides deterministic structural attribution, not universal causal identification. It can show where an observed change occurred and through which measured levers it flowed. Root-cause claims require mechanism-specific evidence. Intervention estimates and causal counterfactuals require additional models and assumptions; a forecast comparison is an expectation comparison, not automatically a causal counterfactual. This separation is the foundation of accountable AI-assisted decision support.

Executive summary

AI use is broad, but trusted operational scale remains limited. McKinsey's 2025 global survey found that 88 percent of respondents reported regular AI use in at least one business function, while only about one-third said their organizations had begun scaling AI programs across the enterprise. Only 39 percent reported any enterprise-level EBIT impact from AI, and most respondents reported an impact below 5 percent [12]. A separate study of more than 48,000 people across 47 countries found that only 46 percent were willing to trust AI systems [13]. Different methods and populations sit behind those figures, but the tension is clear: access to AI has moved faster than confidence in using it for consequential work.

Business systems have also become better at producing recommendations than explaining them, but accountability does not move to the model. An operator still has to decide whether an unintuitive output is rational, defective, stale, incomplete, or misaligned with current goals. When a fluent analysis requires an expert to reconstruct the math, verify the population, check the denominator, and temper unsupported claims, the bottleneck has not disappeared. It has moved into review.

MixShift's response to that gap is not a single model. The Intelligence Service is a governed architecture for moving from measurement to interpretation to action while preserving the evidence boundary at each step. This paper examines HCAM in depth because it establishes the reconciled account of observed performance on which the rest of that architecture can build.

HCAM converts governed business data into an exact attribution structure before a language model enters the workflow. Its two analytical views answer different but composable questions:

- **Horizontal Bridge:** Through which declared metric levers did performance move?
- **Vertical Bridge:** Where across the business hierarchy did the movement occur?

For an identity such as $\text{Revenue} = \text{Traffic} \times \text{Conversion} \times \text{Price}$, the Horizontal Bridge allocates the observed revenue change across traffic, conversion, and price under a declared factor path. The contributions reconcile to the measured change. A Vertical Bridge locates the movement across dimensions and entities while preserving the appropriate bases and rollups. Together, they provide a navigable account of pathway and location.

That account is deliberately narrower than a claim of root cause. HCAM can quantify how much traffic contributed to revenue movement; it cannot, from the revenue identity alone, say why traffic changed. Traffic or impressions may be end nodes until another governed evidence source supplies an upstream identity. Cross-domain observations, such as advertising clicks rising more slowly than sessions, can provide direction and proportionality, but they remain signals rather than exact additive attribution.

HCAM artifacts are retrieved by identity and scope rather than approximated through semantic similarity. Brand Context supplies goals, events, product structure, operating posture, and reviewed history. “What We Know” observations package relevant measured facts and candidate explanations with their claim limits. Forecast, intervention, and identified counterfactual models retain their own assumptions and uncertainty. Skills govern required evidence, permitted claims, behavior when evidence is missing or incomplete, and validation. This evidence-contract approach supports the documentation, provenance, and human-oversight objectives emphasized in established governance frameworks [1, 9, 10].

Internal validation spans Seller Central retail, Vendor Central wholesale, and Amazon Advertising. A governed monthly-report workflow composed a real report from published evidence artifacts without rebuilding the attribution externally, and its figures reconciled to the source evidence [11]. In a separate test, the agent assembled a one-off combined June-and-July year-over-year analysis in approximately 15 minutes, with operator effort concentrated on review; the operator estimated that the equivalent manual work would require roughly a day. This is an internal existence proof, not a controlled productivity benchmark. Release-specific versions, tolerances, coverage, and deployment status remain in a dated validation note so this paper can stay stable as product surfaces evolve.

The result is an Oversight Layer, not an autonomous oracle. AI receives governed evidence, humans receive a traceable explanation, and the boundary between what is known, inferred, modeled, and unresolved remains explicit. This addresses a practical blocker to broader AI adoption: using automated analysis at speed without surrendering auditability or asking humans to rebuild the answer from scratch.

1. The accountability gap in automated decision systems

1.1 The gap between AI use and trusted scale

Organizations adopt consequential AI when the surrounding system makes the output reliable enough to use, inspectable enough to challenge, and efficient enough to avoid a second verification workflow. If every analysis requires a knowledgeable operator to reconstruct the math, resolve scope, and separate evidence from inference, much of the promised leverage disappears. Organizations then accept more risk than they should or restrict AI to lower-consequence work.

The trust deficit does not stop people from using AI; it can produce poorly supervised use. In the 2025 global trust study, 66 percent of AI users reported relying on output without evaluating its accuracy, and 56 percent reported making mistakes in their work because of AI [13]. The precise percentages will change as tools and practices mature. The durable point is that adoption requires trust at the level of each decision, not only strong average model performance.

HCAM addresses a foundational barrier for analytical decisions: can an accountable person trace a conclusion to governed evidence, see how the result was constructed, and know exactly where interpretation begins? That question predates generative AI; fluent synthesis has simply made it more urgent.

1.2 The problem that led to HCAM

The problem that led to HCAM became concrete through the author's work shaping and scaling Amazon's Hands Off the Wheel program. As the operations leader for an Amazon business unit, he partnered with the Supply Chain Optimization Technologies team to pilot an economic ordering model and expose the human corrections masking forecast bias, capacity constraints, and disconnected planning signals. He documented the changes required, helped test them through successive peak seasons, and was subsequently tapped as the business lead for the broader rollout, serving as a check between optimization teams and the category owners accountable for real-world outcomes.

Amazon has described Hands Off the Wheel as a large-scale supply-chain automation system using mathematical optimization and machine learning to decide what to buy, how much to buy, where to place it, and related actions [14]. The ambition was broader than automating an inventory calculation. It was to connect forecasting, purchasing, placement, and constraints across an end-to-end system while moving people from recurring decision makers toward overseers of automation.

That shift created a healthy tension between speed and accountability. Operators remained accountable for the business result even when they no longer made every underlying decision. Interdependence mattered: in-stock managers had been compensating manually for forecast bias and inbound-capacity constraints by adjusting reorder timing, planning horizons, and critical ratios while balancing in-stock rate, working capital, and turns. Targets were typically directional, framed as doing better than the prior year, rather than derived from a single integrated definition of optimal performance.

Automating any one decision in isolation risked removing those compensating behaviors before the forecast, planning horizon, and capacity logic had been redesigned. The eventual end-to-end system became forecast- and capacity-aware, including an additional planning horizon for constrained inbound flow. That improved global coherence but made recommendations harder for an accountable operator to interrogate. A recommendation had to be understood as the output of the integrated system, not a number generated in isolation.

In one instance, the system recommended buying approximately \$10 million more inventory year over year for a business unit whose sales were approximately 10 percent lower year over year. The category leader pushed back, reasonably: what could justify expanding inventory while the business was contracting?

There were plausible answers. The mix might have shifted toward higher-cost products. Profitability might have improved enough to justify higher critical ratios and a higher target service level. Vendor lead-time variability or forecast uncertainty might have widened. Constraints, inbound position, or demand expectations might have changed. The recommendation itself could have been right.

The business owner saw an output that did not make intuitive sense, and the team could not show which explanations applied or how much each one mattered. Yet if declared mathematical relationships produced the recommendation, the \$10 million year-over-year increase should have been decomposable through those same relationships.

That experience produced the question at the center of HCAM:

Why did the recommendation increase relative to the prior year, and exactly how much did each input contribute?

1.3 Recalculation is not explanation

The available response was to rerun an expensive end-to-end simulation. That can help a technical team see what changed inside the model, but it answers a different question from the one the business owner asked. The business-facing explanation must first connect the recommendation to a meaningful decision anchor; in this case, the prior year.

Rerunning a model can incorporate fresher data, correct a defect, or test an assumption; it does not inherently explain why either output differs from the decision anchor. An accountable operator still needs to know:

- Which inputs, assumptions, or constraints changed?
- How much did each change contribute to the result?
- Where in the business hierarchy did the movement manifest?
- Which contributions are measured facts and which are model estimates?
- Did the new result reconcile to its stated bases and scope?
- Is the recommendation aligned with current goals and operating context?

Dashboards expose outputs and trends. Large spreadsheet models may expose equations to a sophisticated operator. Simulations can optimize a decision. None automatically converts an unexpected result into a reconciled account of which inputs moved, how their effects combined, and where those effects appeared. The missing capability is a reusable evidence path from the decision back through its governed inputs.

1.4 From an inventory decision to a general control problem

The same pattern appears wherever an automated system joins many inputs to a consequential output. A forecast may miss because an event moved on the calendar, demand weakened, price changed, or the operating conditions represented in its training history no longer match current conditions. An advertising recommendation may react to lower immediate return without accounting for funnel position or lagged effects. A report may calculate conversion using sessions from the full catalog but orders from only a filtered subset, producing a precise-looking rate whose numerator and denominator describe different populations.

More automation can reduce the effort required to generate an answer while increasing the effort required to validate it. The accountable human needs a control plane that preserves identity, scope, evidence, and the path through which the answer changed. That is the role of an Oversight Layer: it does not replace the underlying models, data systems, or operators; it makes their outputs inspectable and composable. HCAM is the deterministic attribution core of that layer.

1.5 The Intelligence Service as a trust architecture

HCAM is foundational, but trustworthy AI-assisted decision support requires more than attribution. The MixShift Intelligence Service joins four interdependent functions while preserving the evidentiary status of each:

- **Measurement and attribution:** HCAM determines what changed, through which declared levers, and where the movement appeared.
- **Context and mechanism evidence:** Brand Context, events, goals, operating history, and “What We Know” observations establish what matters and which explanations are supported or ruled out.
- **Expectations and actions:** forecast models establish expected baselines, while experiments, elasticity models, and other identified intervention or counterfactual designs estimate actions and alternative outcomes with their assumptions and uncertainty intact.
- **Governed AI execution:** identity-resolved retrieval, skills, claim contracts, validation, refusal, and human review control how the evidence is selected, combined, and communicated.

The functions reinforce one another. Without attribution, AI must reconstruct complex math from raw data. Without context, exact math lacks business meaning. Without expectation, intervention, and counterfactual models, the system can explain observed performance but cannot responsibly assess a miss, estimate an action, or compare with an identified alternative. Without execution controls, correct evidence can still become an unsupported conclusion. This paper concentrates on HCAM because the

reconciled analytical spine is the prerequisite for the rest of the system, not because it is the entire system.

2. Why AI increases the need for oversight

2.1 AI is a powerful synthesizer, not a substitute for governed business logic

Language models can compress complex evidence into an accessible narrative, connect context to measured observations, and generate useful investigative questions. That makes AI a compelling delivery mechanism for HCAM: it accelerates synthesis that previously required an expert to navigate a data cube, application, or spreadsheet.

Synthesis and governed computation remain different jobs. Given raw rows, a language model must infer definitions, choose populations, aggregate bases, calculate ratios, resolve identities, assess completeness, and distinguish structural relationships from causal mechanisms. Multi-step performance can degrade as task complexity grows, so fluency is not evidence that each operation was executed correctly [7]. Generation alone does not guarantee correct aggregation or identify causal effects from observational business data. Trust depends on the evidence, definitions, scope controls, and validation surrounding it.

2.2 The raw-data reasoning trap

Business data contains deterministic traps that can change a decision. Ratios generally require summed numerators and denominators rather than averages of row-level rates. A missing denominator is not zero. If advertising spend is positive while attributed sales are zero, advertising cost of sales (ACOS) has no finite value; treating it as 0 percent implies perfect efficiency, while passing infinity downstream contaminates comparisons.

HCAM handles the condition according to the declared metric family: the undefined state is typed and labeled rather than replaced with zero, infinity, or a fabricated finite rate, and the decomposition still reconciles to its account anchor. Section 6.3 details the governed treatment. The same discipline applies elsewhere: a top-N extract is not a complete population, an absent entity must not be replaced by an account total, and different metrics may require different weighting bases.

Asking a probabilistic model to rediscover these rules on every run consumes context and makes errors harder to audit. HCAM solves the deterministic part first. The language model receives a governed attribution artifact, together with prepackaged “What We Know” observations that stay within their claim limits, rather than a table and an invitation to improvise. It can then spend its capacity on interpretation, synthesis, and decision framing while the calculation remains exact and testable.

2.3 Grounding is necessary but not sufficient

Retrieval-augmented generation conditions generation on external memory rather than model parameters alone [2]. Authoritative evidence should live outside the language model and be retrieved for the task. Numerical evidence adds a stricter requirement: the correct artifact is defined by tenant, run, period, metric, grain, entity, schema, and capability, not by semantic similarity. For HCAM, retrieval is therefore structured and identity-driven; vector search is not required to fetch an authoritative bridge.

Grounding still does not guarantee a valid conclusion [8]. An AI can ignore a caveat, overgeneralize from partial coverage, or escalate a directional observation into a causal statement. The architecture therefore governs claims as well as retrieval.

2.4 The Oversight Layer recurses onto the AI

HCAM was conceived as a way for humans to oversee complex automated systems. In practice, the first system it had to oversee was the AI consuming HCAM itself. The same principles of identity, scope, evidence, typed absence, reconciliation, and human review must apply to the generated narrative. The agent should know not only what the data says, but what the evidence permits it to say. This operationalizes the knowledge-limits and human-oversight objectives in established AI risk-management guidance [1, 10] for analytical claims.

3. What HCAM is and what “causal” means here

3.1 Claim classes, not a ranking

The word *causal* carries several meanings in business conversation. HCAM separates five non-ranked claim classes:

- **Observation and association (C2).** What values were observed, and what moved with what?
- **Deterministic structural attribution (C1).** Through which declared metric relationships did the observed output move, and where did the movement manifest?
- **Evidence-backed mechanism (C3).** Is there attached evidence of an operational event or mechanism consistent with the movement, and have relevant alternatives been considered?
- **Intervention or identified counterfactual estimate (C4).** What outcome is expected if an action is taken, or under a defined alternative state, given a stated model, experiment, elasticity, or identification assumption?
- **Forecast or expectation comparison (C5).** How did actual performance compare with a formal expected baseline, given model fit and uncertainty?

Five governed claim classes

Different evidence contracts; no universal rank

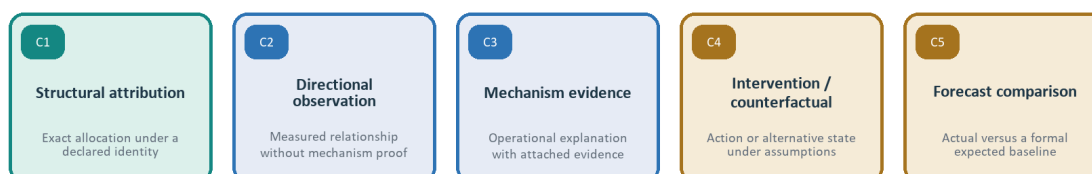


Figure 1. Claim classes preserve the evidence boundary among structural attribution, directional signals, mechanism evidence, intervention or identified counterfactual estimates, and forecast comparisons.

HCAM directly supplies structural attribution. It organizes observations, provides an accounting spine for mechanism investigation, and can be combined with separately governed intervention and forecast models. The classes are not strictly ordered: an intervention estimate may have stronger identification than a forecast comparison, and a forecast is not a causal counterfactual unless its design supports that interpretation.

3.2 Structural attribution

Business metrics are not arbitrary collections of correlated fields. Many have declared identities: revenue is the product of traffic, conversion, and price; advertising cost of sales can be expressed through cost per click, conversion, and average order value; and inventory coverage relates available supply to an expected demand rate. When an input in a declared identity changes, the computed output changes mechanically.

These are complementary analytical views. For ACOS, the Horizontal Bridge decomposes the account-level identity through cost per click, advertising conversion, and advertising AOV; the Vertical Bridge allocates the measured account-level rate change across entities on the governed attributed-sales basis.

HCAM uses *causal* in a bounded structural sense. Within a declared identity, the relationship is causally operative as mathematics: under the stated factor path, changing an input changes the computed output. HCAM can therefore state that reduced traffic contributed a specified amount to an observed revenue decline, or that a campaign contributed a specified number of basis points to an observed ACOS change. Those contributions are exact for the governed identity, scope, and allocation method. This differs from correlation because the terms are joined by a declared identity and reconcile to the measured output.

The boundary is upstream and prospective. The same bridge does not establish why traffic declined, why the campaign's economics changed, or what would have happened under an unobserved action. Those questions require mechanism evidence, an intervention design, or an identified counterfactual model. HCAM identifies the structural contribution of the measured input; it does not identify the external process that produced it. If traffic declined, HCAM can quantify the contribution to revenue movement, but the revenue identity cannot reveal whether the decline stemmed from seasonality, advertising, competition, search rank, distribution, or customer preference.

3.3 End nodes and directional signals

A Horizontal Bridge must stop when it reaches a measured node for which no governed upstream identity exists; such a node is an end node. In the Amazon commerce domain described here, traffic or sessions and advertising impressions are horizontal end nodes unless an additional governed model is supplied. This does not prevent a Vertical Bridge from locating where the measured traffic or impression change occurred across products, campaigns, ad groups, targets, or other governed entities.

An end node is not analytically useless. Cross-domain evidence can still supply direction and proportionality: if advertising clicks rise much faster than sessions, paid traffic may be contributing to growth; if sessions rise while clicks fall, other sources or measurement differences deserve attention.

Such cross-domain evidence should not be forced into an exact additive attribution unless the populations, identities, time windows, and measurement boundaries support the claim. HCAM keeps the distinction explicit.

3.4 Claim classes as typed contracts

The distinction is enforced through typed claim classes rather than left to writing style.

Class	Permitted claim	Example boundary
C1: Deterministic structural attribution	Exact values, changes, locations, and contributions that reconcile to an anchor	May quantify traffic's contribution to revenue; may not state why traffic changed
C2: Directional or contextual observation	Relationships among measured facts without an identified mechanism	May compare click and session direction; may not claim advertising caused the session change
C3: Evidence-backed operational explanation	A qualified mechanism supported by an attached event, changelog, experiment, or equivalent evidence	May say a recorded price change is consistent with measured ASP and conversion movement

Class	Permitted claim	Example boundary
C4: Intervention or identified counterfactual estimate	Expected outcome under an action or defined alternative state, with stated assumptions and uncertainty	Requires an experiment, elasticity model, or identification strategy
C5: Forecast or expectation comparison	Actual versus a formal baseline with model fit and uncertainty	A forecast miss is not automatically a causal effect or causal counterfactual

These classes are deliberately conservative and non-ranked. They do not prevent the system from generating a useful hypothesis. They prevent a hypothesis, forecast deviation, or structural allocation from silently becoming an identified causal fact.

4. The HCAM methodology

4.1 Two views of the same movement

HCAM combines two complementary decompositions.

The **Horizontal Bridge** answers: through which declared metric levers did the result move? It follows a metric identity such as revenue through traffic, conversion, and price.

The **Vertical Bridge** answers: where did the movement occur? It allocates the account-level change across dimensions, groups, and entities such as product groups, ASINs, campaigns, ad groups, targets, or search terms, depending on the domain and available evidence.

An executive needs both. A statement that conversion drove a revenue decline is incomplete if the deterioration was concentrated in one product family. A statement that one item group declined is incomplete if the movement was actually a deliberate price-volume tradeoff. Horizontal and vertical evidence are separate but composable.

Calculated metrics can expose **rate and mix** in both bridge directions. Horizontally, rate and mix may be declared levers inside the metric identity. Vertically, the bridge can separate change within entities from change caused by the portfolio shifting toward entities with different rates. That distinction often changes the decision. One pattern may call for an execution fix inside specific products or campaigns. The other may reflect a deliberate or unintended change in the composition of the business.

Consider a promotion after which revenue declines. A Horizontal Bridge might show that lower average selling price pulled revenue down while higher conversion partially offset the loss. The rate/mix detail can then distinguish prices falling on the same products from customers shifting toward lower-priced products. It can likewise distinguish the same products converting better from traffic shifting toward

products that already had higher conversion rates. A recorded promotion can support the explanation that the observed discount was consistent with the price-rate decline and conversion improvement. If the conversion gain did not offset the price loss, the observed bridge answers whether total revenue improved during the period. It does not, by itself, answer whether revenue would have been still lower without the promotion; that counterfactual requires an intervention design.

4.2 Calculation invariants

HCAM uses a small set of invariants across domains:

- Period 1 is the comparison period, such as prior month or prior year. Period 2 is the current or observed period.
- Change is always calculated as $\text{Period 2 value} - \text{Period 1 value}$.
- Growth is $(\text{Period 2 value} \div \text{Period 1 value}) - 1$.
- Positive contribution means the factor pushed the target metric upward under the declared identity.
- Aggregated ratios are calculated by summing numerators and denominators first, then dividing.
- An insufficient denominator is a typed state, not a zero.
- Every bridge declares its anchor, scope, and footing result.
- Path-dependent decompositions use a stable canonical factor order.
- Claims beyond the artifact's declared decomposition ceiling are typed as out of contract.

These rules sound elementary, but their value is consistency. An analytical system becomes difficult to trust when signs, bases, missingness, and populations change subtly between views.

4.3 Horizontal Bridge: an exact path decomposition

Consider the identity:

$$\text{Revenue} = \text{Traffic} \times \text{Conversion} \times \text{Price}$$

The representative equations illustrate how one declared factor path allocates and reconciles a measured change. They are not a comprehensive specification of HCAM's calculation methods or implementation system.

Let Period 1 be the comparison state and Period 2 the current state. One declared sequential-substitution path allocates the change as:

$$\text{Traffic contribution} = (T_2 - T_1) \times C_1 \times P_1$$

$$\text{Conversion contribution} = T_2 \times (C_2 - C_1) \times P_1$$

$$\text{Price contribution} = T_2 \times C_2 \times (P_2 - P_1)$$

The terms telescope exactly to:

$$R_2 - R_1$$

This is not an approximation. It is an exact allocation of the measured change under the selected path. The path determines where interaction effects land. If traffic is substituted first, its contribution is evaluated at the comparison conversion and price. Conversion is then evaluated at current traffic and comparison price. Price is evaluated last at current traffic and conversion.

The three-factor revenue identity is illustrative, not a ceiling. Other governed identities use two, three, four, or more factors, and calculated metrics can expand through nested relationships. TACOS, for example, joins advertising spend to total revenue and can require evidence that crosses advertising and commerce domains. The engine exposes only the factors and depths for which the full identity, bases, scope, and coverage are governed.

Different bridge families require different mathematics. Some are additive and order-invariant. Others use fixed-path sequential substitution. Vertical mix/rate analysis uses a deviation-form decomposition rather than the same three-factor product formula. Sequential and compositional decomposition have a long history in index-number, demographic, and variance analysis; HCAM adapts those established ideas to governed business identities and evidence contracts [6]. The shared kernel is therefore not one universal equation. It is a contract for selecting the correct family from the metric identity while retaining common sign, scope, insufficiency, and footing rules.

4.4 Why use a fixed path?

A fixed path makes the allocation inspectable and repeatable. It can also reflect an operational convention or mechanism. For inventory coverage, for example, an inventory-first order can be defensible because the question is explicitly framed around how inventory position interacts with demand rate. In other cases, the order may be a presentation convention rather than a causal claim.

Where an operational path is selected, it encodes domain judgment that an experienced human operator would otherwise apply manually: which input is controllable, how one metric is likely to affect another, and which cross-metric thread deserves attention first. Packaging that judgment with the identity makes it available to the AI as governed context rather than asking the model to rediscover it from prose or correlations. The path still governs allocation, not proof of an external mechanism.

The important requirement is honesty: not every selected order represents the unique physical sequence of the business. HCAM trades symmetry for a fixed path. That trade improves interpretability when the path is meaningful, and it keeps the calculation simple enough to audit. For path-dependent families, the allocation convention must remain stable and inspectable. Current artifacts preserve it through canonical component order and component-level basis descriptions.

4.5 Relationship to Shapley attribution and SHAP

HCAM was designed outward from a business-control problem, not derived from Shapley attribution, SHAP, or do-calculus. Technical readers may nevertheless recognize a family resemblance. The comparison is useful because it places HCAM alongside established allocation methods and makes its deliberate tradeoffs explicit.

The Shapley value comes from cooperative game theory. It distributes a joint payoff by averaging each participant's marginal contribution across every possible ordering [3]. SHAP, short for **SHapley Additive exPlanations**, adapts that idea to machine learning: relative to a baseline prediction, it estimates how much each input feature moves a particular model prediction up or down [4]. HCAM also allocates a total change through marginal contributions, but it uses a declared business identity and fixed path rather than averaging every order.

That design serves a different objective from Shapley symmetry. When a business-relevant path exists, HCAM preserves the operational sequence and assigns interaction effects to the step where an operator expects to inspect them. The same selected order can be applied across periods, levels, and delivery surfaces, producing a stable explanation that maps directly to controllable levers. Averaging every mathematically possible order can be useful when no ordering should be preferred, but it can also blend operationally implausible sequences and obscure the pathway the business uses to understand the metric. Where no operational path is preferred, HCAM uses a stable allocation convention; a symmetric all-order variant remains a possible policy choice (Section 11.3).

HCAM is not SHAP. It decomposes observed business performance through a declared metric identity and hierarchy, not a black-box model prediction. A fixed-order marginal-contribution allocation satisfies **closure**, meaning the allocated legs sum exactly to the observed change, as well as the null-player property and additivity. It does not satisfy symmetry because interaction terms are assigned to the selected order rather than averaged across all orderings. Shapley's efficiency axiom is conceptually related to closure, but the terms are kept separate here because HCAM is not presented as a cooperative game.

For the positive multiplicative factors admitted by the engine's denominator and positivity gates, changing the factor order cannot reverse the sign of a contribution; it can change only the magnitude

assigned to that factor. This direction-invariance result does not eliminate path dependence, and it is not asserted for identities outside those conditions [11].

Internal sensitivity testing across representative bridge families found that the fixed-path and all-order allocations were close under typical operating conditions, while path dependence became more material when large opposing movements nearly canceled into a small net change. Those are precisely the cases in which an account can look flat while its components are moving sharply in opposite directions. HCAM applies magnitude and masked-flat guardrails in those conditions [11]. Appendix A provides one transparent example; the broader study remains in the dated implementation record.

Neither fixed-path attribution nor Shapley attribution identifies an intervention effect merely by allocating a change. That boundary matters more than which allocation rule is chosen.

4.6 Relationship to the do-operator

Sequential substitution can look superficially like an intervention because one factor is changed while specified others are held at prior or current values. Pearl's do-operator has a more demanding meaning: it represents an external intervention in a structural causal model and supports causal identification only under the assumptions encoded in that model [5].

An H-Bridge substitution does not remove confounding, alter a causal distribution, or estimate $P(Y \mid \text{do}(X=x))$. It does not need to do so for the question HCAM is designed to answer. The comparison and current states were both observed, the governing metric identity is known, and the task is to allocate the observed difference exactly across its inputs. No unobserved intervention effect must be identified to complete that retrospective accounting.

Do-calculus becomes relevant when the question changes from *What contributed to the observed difference?* to *What would happen if we intervened?* or *What would have happened under a different action?* At that boundary HCAM can supply the measured structure and evidence, but an intervention or counterfactual model must carry the identification assumptions. HCAM therefore avoids causal machinery that is unnecessary for exact decomposition without claiming that such machinery is unnecessary for prospective causal decisions.

4.7 Vertical Bridge: locating the movement

Horizontal decomposition explains the metric pathway. Vertical decomposition locates the movement across the business.

A vertical bridge begins with the account-level change and allocates it across a declared dimension and hierarchy. Depending on the metric, it may separate changes in composition from changes in within-entity rate. It preserves the appropriate bases so that group and entity results reconcile to the

account anchor. The analysis can then move from account to item group to entity, or through advertising grains such as campaign, ad group, target, search term, and purchased product.

For a ratio metric, let entity i have period-specific rate r_i , denominator weight w_i , and share $s_i = w_i / \sum w$. The account rate is $R = \sum (s_i \times r_i)$. One shipped deviation-form allocation separates each entity's contribution into:

$$\text{Rate}_i = (r_{2i} - r_{1i}) \times s_{1i}$$

$$\text{Mix}_i = (r_{2i} - R_2) \times (s_{2i} - s_{1i})$$

The rate term captures change within the entity at its comparison-period share. The mix term captures the effect of shifting denominator weight toward or away from an entity whose current rate differs from the current account rate. Summed across the governed population, rate plus mix closes to $R_2 - R_1$. The method requires denominator-complete coverage and returns a typed insufficient state when the required bases are unavailable. The equations above illustrate one representative allocation pattern and its closure property. They do not disclose the full set of calculation families, decision rules, edge-case treatments, or governed implementation.

A two-entity example shows the closure. The account rate rises from 14 to 15 percent. Entity A's rate rises from 10 to 12 percent as its share falls from 60 to 50 percent; Entity B's rate falls from 20 to 18 percent as its share rises from 40 to 50 percent.

Entity	Rate change at prior share	Mix from share change	Total contribution
A	(+2.0 pts × 60%) = +120 bps	((12% - 15%) × -10%) = +30 bps	+150 bps
B	(-2.0 pts × 40%) = -80 bps	((18% - 15%) × +10%) = +30 bps	-50 bps
Account	+40 bps	+60 bps	+100 bps

The vertical contributions close exactly to the account movement: 15% - 14% = +100 bps. The example does not say why either entity's rate or share changed; it separates within-entity performance from portfolio composition.

This is not a niche statistical refinement. Aggregate metrics are where leaders allocate capital, judge teams, and train automated systems. When large gains and losses cancel, a flat total can certify inaction even while one part of the business is deteriorating rapidly and another is masking it. An improving aggregate rate can also coexist with broad within-entity deterioration if the portfolio shifts toward stronger entities. A top-N table can omit the entity responsible for a material residual. HCAM evaluates the governed population before making claims such as *largest*, *only*, or *all*.

Every artifact declares how deep its decomposition goes. A request beyond that ceiling is not treated as an empty result. It is marked out of contract, preserving the difference between “nothing happened” and “this artifact does not answer that question.”

4.8 Composing the bridges

The two bridges create a repeatable analytical path:

1. Establish the account-level movement and comparison basis.
2. Decompose the movement through the relevant Horizontal Bridge.
3. Identify the material levers and any large offsetting contributions.
4. Use the Vertical Bridge to locate those contributions across groups and entities.
5. Retrieve relevant cross-domain evidence, Brand Context, events, and model state.
6. Separate deterministic findings from directional signals and mechanism hypotheses.
7. Validate the figures, coverage, caveats, and claim language before rendering the result.

The language model enters after the analytical spine exists. It can navigate and explain the structure without recalculating it.

5. From calculation to an AI-ready evidence system

5.1 Reference architecture

The Intelligence Service operationalizes the four trust functions introduced in Section 1.5. It does not flatten attribution, context, forecasts, intervention or counterfactual estimates, and generated prose into one undifferentiated answer. It gives each evidence source a governed interface and preserves its claim class as it moves through the workflow.

The method becomes an Oversight Layer through a sequence of governed interfaces:

Business data → HCAM engine → evidence artifacts → Intelligence Service → skill → AI synthesis → validation → human decision support

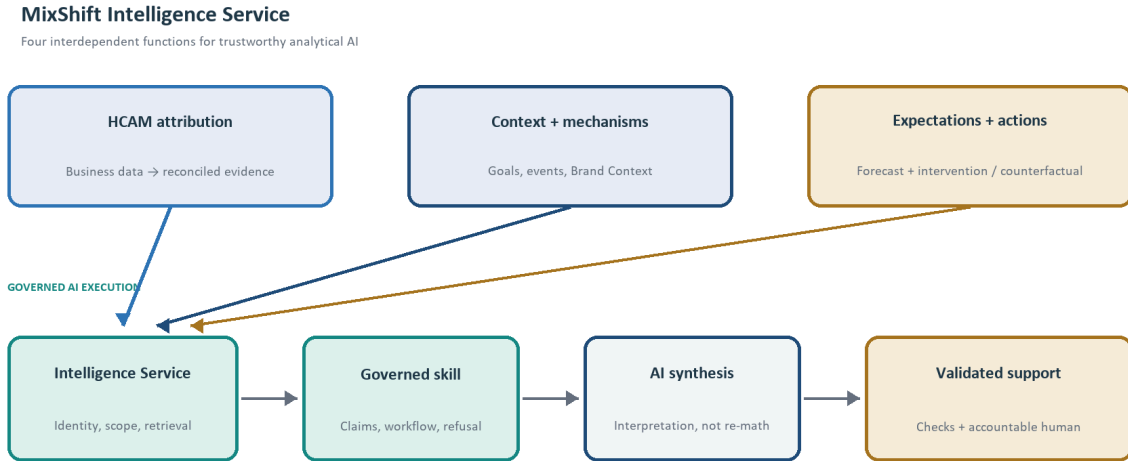


Figure 2. HCAM supplies the deterministic analytical spine; context, expectation and action models, and governed AI execution complete the MixShift Intelligence Service trust architecture.

Brand Context, mechanism evidence, and sibling-model evidence enter the workflow as separate governed sources. The architecture preserves their different evidentiary status. HCAM evidence is deterministic attribution. Forecast evidence is a model-based expectation. Intervention or counterfactual evidence estimates an action or alternative outcome under stated assumptions and an identified design. An event may support an operational mechanism. A skill can combine them without pretending they are interchangeable.

The stable interface matters more than the current transport. A forecast artifact may enter through an operator upload today and a direct service integration later. The paper's claim is that the source is governed, labeled, and suppressed when absent, not that one delivery mechanism is architecturally essential.

5.2 The HCAM artifact as a formal contract

HCAM outputs are packaged independently from the interface that displays them. Formally, an HCAM evidence artifact can be represented as:

$$A = \langle I, S, B, M, \Delta, L, H, K, V, E \rangle$$

where:

- I is identity: run, tenant, account, marketplace, entity, and artifact key;
- S is scope: current and comparison periods, population, grain, and filters;
- B is the governed basis: numerator, denominator, weighting, currency, and sign convention;

- M is the governed metric identity and decomposition method; for path-dependent families, the current implementation carries the selected path through canonical component order and component-level basis descriptions;
- Δ is the measured anchor change;
- L is the set of horizontal and vertical contribution legs with a footing result;
- H is the hierarchy, parent-child mapping, and decomposition ceiling;
- K is capability and coverage state, including typed absence or insufficiency;
- V is schema, engine, and source provenance; and
- E is attached evidence metadata: caveats, claim permissions, and quotation constraints.

The contract is valid only if identity and scope are explicit, bases are preserved, contribution legs close to the anchor within the declared tolerance, and absence is represented as state rather than substituted data. A consumer may quote or compose the artifact; it may not silently change its scope, recompute its ratios under a different basis, or claim beyond its decomposition ceiling.

This packaging prevents every consumer from inventing its own bridge math. It also lets a report, plugin skill, standalone application, or later background process use the same computed evidence.

Versioned replayable artifacts are not the same as a complete enterprise audit log. They do, however, create inspectable analytical provenance: the consumer can identify which engine and evidence contract produced the result and can compare the artifact with its source run. The design is consistent with established provenance models that distinguish entities, activities, agents, and derivation relationships [9].

Replayability preserves and reproduces the evidence state used for the original result. It does not refresh that evidence or guarantee that the upstream source will not later restate it.

5.3 Structured RAG

HCAM is RAG-enabled in a specific sense. Generation is augmented with external, authoritative evidence, but numerical artifacts are resolved through structured identity rather than approximate text similarity.

Retrieval resolves the relevant tenant, run, period, metric, grain, entity, capability, and business context.

Augmentation supplies typed figures, decomposition legs, coverage, caveats, claim permissions, and contextual memory.

Generation composes a decision-ready explanation against those inputs.

This is structured RAG, or deterministic-artifact RAG. It preserves the core idea that generation should be conditioned on external memory while adapting retrieval to the requirements of numerical evidence. A vector database is not required to retrieve the correct bridge. Semantic search may still be useful for narrative corpora, policies, or unstructured context, but it should not decide which account total belongs to which tenant or which entity result belongs to which run.

5.4 Prepackaged “What We Know” evidence

HCAM does not stop at bridge legs. Deterministic evidence builders inspect the governed artifact and package relevant observations, exclusions, and candidate drivers for the AI and the human reviewer. These “What We Know” statements are bound to the same run, scope, metric, entity, and capability state as the underlying figures.

For example, suppose sales declined, the Horizontal Bridge shows that conversion mechanically accounted for most of the movement, and an event record shows that a promotion ended at the period boundary. The system can surface the promotion change as a likely driver consistent with the measured conversion decline. It cannot say the promotion ending caused the full sales decline unless stronger mechanism evidence supports that claim.

This middle layer is operationally important. It saves the user and the AI from manually searching every adjacent dataset for a plausible thread, while the claim contract preserves the distinction between exact attribution and a well-supported explanation. The system can also package exclusions. If availability improved, stockouts should not be offered as the explanation for a decline merely because they are common in the domain.

5.5 Brand Context

Performance data rarely explains what matters to a specific business. The same inventory level can be acceptable for a mature item and dangerous for a launch. The same advertising return can be weak for a bottom-of-funnel campaign and acceptable for an upper-funnel growth strategy. A calendar event can make a year-over-year comparison misleading even when the arithmetic is correct.

Brand Context supplies durable operating memory:

- business identity and account structure;
- goals and management posture;
- structural events and effective dates;
- active watches and reviewed discoveries;

- product structure, item-group mappings, and campaign structure;
- relevant operating history; and
- reporting and interpretation rules.

The context is retrieved, not assumed from the language model's training memory. Just as importantly, a detected pattern is not silently promoted into durable context. The system can propose a discovery; a human decides whether it becomes an accepted fact, watch, or operating rule.

5.6 Sibling analytical models

HCAM is the analytical spine of a broader decision-support system, not a substitute for every kind of model.

A forecast answers: **What was expected?** It carries fit, coefficients, uncertainty, seasonal structure, and a training window.

HCAM answers: **What changed in the realized result, through which measured levers, and where?**

An intervention model answers: **What might happen if we take a specified action?** It requires assumptions, experimental evidence, elasticity, or another identification strategy.

An identified counterfactual answers: **What would likely have happened under a defined alternative action or state?** It requires a design capable of supporting that comparison.

These outputs can be combined because they are labeled separately. The architecture should never turn a forecast coefficient into an HCAM contribution or an observed attribution ratio into a causal response coefficient simply because the values are numerically similar.

The combination is more valuable than a stack of separate outputs. A forecast establishes what was expected under its fitted assumptions. An intervention or identified counterfactual model estimates an action or alternative outcome under its causal assumptions. HCAM explains what actually moved and where the beat, miss, or estimated effect manifested. “What We Know” evidence then brings forward relevant events, exclusions, and candidate mechanisms. Together, they turn a residual into a prioritized investigation. When the companion model has a valid causal design, the combined system can support claims at that model's evidence level. HCAM does not create causal identification on its own.

5.7 Skills are analytical harnesses

A skill is more than a prompt. It defines a controlled analytical workflow:

- required inputs and scope;

- which evidence must be retrieved;
- which claims are permitted;
- what to do when evidence is missing or partial;
- how figures and caveats must be quoted;
- which validations must pass; and
- how the final artifact is rendered or persisted.

The model may perform one bounded composition step while deterministic extractors, validators, and renderers control the surrounding process. This division of labor makes the output more repeatable and turns observed failure classes into testable rules.

6. Governing the AI that communicates the analysis

6.1 Evidence and claim contracts

An evidence contract describes what the analytical system produced. A claim contract describes what a consumer is allowed to say about it.

The distinction prevents a common failure: a correct figure being used in an unsupported sentence. If a governed identity includes an advertising input, HCAM may state that the measured advertising change contributed a specified amount to the observed output change. Claiming that an advertising strategy caused the underlying traffic, conversion, or demand behavior is a mechanism claim and requires more. A forecast miss may be a C5 expectation comparison. “The business underperformed because demand weakened” requires evidence that separates demand from event timing, price, conversion, availability, and model error.

In a governed workflow:

- every material figure points to a defined evidence source;
- the figure's caveats follow it to each quotation site;
- coverage controls whether universal or superlative language is allowed;
- deterministic statements and hypotheses are formatted differently; and
- questions for investigation are not presented as conclusions.

6.2 Typed absence and refusal

Missing evidence has more than one meaning. A capability may be computed, partial, or skipped. A skipped result may be not requested, unavailable in the source, unsupported on the channel, or outside the contract. Collapsing these states into a blank cell or zero destroys information.

The same principle applies to entity retrieval. If a requested entity does not exist in the governed artifact, returning an account total would produce a fluent but false answer. HCAM's evidence service rejects the request instead. Refusal is not a degraded user experience in this context; it is a control that preserves scope.

Typed absence also helps the AI explain what it cannot conclude. “No decline was detected” is different from “the detector was not run,” “the source did not contain the required field,” and “this channel does not support the metric.”

6.3 Denominators, populations, and coverage

Many analytical errors begin with the denominator. HCAM follows a bases-first rule: sum the governed numerator and denominator over the intended population, then calculate the ratio. Row-level rates are not averaged unless that is explicitly the metric definition.

When an entity-level ratio is undefined in one period, the evidence may carry a labeled net-only decomposition produced under the declared zero-substitution rule. When it is undefined in both periods but the account-level rate remains defined, an advertising bridge may carry an explicitly labeled allocated remainder so the account decomposition still reconciles without inventing an entity rate. When the declared bases support neither a native decomposition nor an authorized residual allocation, the governed result is a typed insufficient state rather than zero, infinity, or a fabricated contribution. Coverage metadata bounds a claim to the population the artifact actually represents. A top-N extract is not a verified census. When completeness has not been established, the consumer must either qualify the statement to the observed population or withhold universal and superlative language.

For example, a table containing the top five products may cover only 60 percent of account revenue. It can support a statement about the largest decline *among those five products*. It cannot support “the largest decline in the catalog,” because an omitted product may be larger. Likewise, conversion with zero sessions is undefined, not 0 percent. Treating it as zero invents performance that was never observed.

These constraints are mundane compared with a language model, but they are exactly the controls that prevent sophisticated prose from outrunning the data.

6.4 Mechanical validation

The final narrative is subject to mechanical checks.

Footings checks confirm that bridge contributions reconcile to their anchor. Figure contracts preserve source values and labels. In the governed monthly-report workflow, claim validators operate at the consuming skill's render seam. They reject unsupported comparison language, missing required caveats, and claims that exceed the coverage represented in the evidence. The validator rules are treated as reusable contracts rather than discretionary writing guidance. Negative fixtures reproduce prior failure classes so that a defect becomes an executable rule rather than an institutional memory.

Arithmetic reconciliation does not prove causal validity. It proves that the evidence was carried through the system without unexplained numerical drift. That is a necessary control even when additional causal evidence is required.

6.5 Compact failure and control model

The architecture treats common analytical failures as contract violations with explicit system behavior.

Failure mode	Preventive or detective control	Governed behavior	Residual risk
Wrong run, tenant, period, or entity	Identity- and scope-resolved retrieval	Reject the request; never substitute a nearby artifact or account total	Incorrect upstream identity mapping
Arithmetic drift or altered figures	Footings checks and figure contracts	Fail validation before publication	Error already present in the governed source
Invalid denominator or weighting	Bases-first aggregation and typed insufficiency	Return unavailable with a reason, not zero	Incorrect metric definition or source bases
Partial population used for a universal claim	Verified census or coverage state where available	Bound the statement to the observed population or suppress universal and superlative language	Completeness metadata may be unavailable or incomplete at some delivery surfaces
Unsupported mechanism or causal claim	Claim-class contract and evidence requirements	Reframe as structural attribution, signal, hypothesis, or refusal	Human overinterpretation after delivery
Grounded but unsupported generated prose	Evidence-bound quotation, claim linting, and source reconciliation	Reject or revise the sentence	Nuance and decision quality still require review
Stale or incorrect business context	Effective dates, source labels, and human promotion	Caveat, expire, or withhold the context	Delayed human governance

The controls do not make the source data true or eliminate judgment. They prevent the AI layer from silently repairing gaps, changing populations, or escalating a lower-evidence observation into a stronger claim.

6.6 Human control

The objective is not to remove the human from the loop. It is to improve the quality and speed of human judgment.

The AI can traverse the evidence, surface anomalies, draft explanations, and suggest next questions. Humans remain responsible for promoting context, approving consequential interpretations, selecting interventions, and deciding whether the evidence is sufficient for action.

The system therefore treats uncertainty and refusal as first-class outputs. For example, a well-governed agent should be able to produce a bounded statement such as: *The observed pattern is consistent with this hypothesis, but the current evidence does not identify the mechanism.* NIST's AI RMF similarly calls for documented knowledge limits and defined human-oversight processes [1, 10]. That bounded statement is often more valuable than a confident guess.

7. Novel contribution and internal validation

7.1 The novel contribution is the governed composition

HCAM does not claim that metric identities, variance decomposition, hierarchical drill-down, retrieval-augmented generation, or Shapley comparison are individually new. Each has established antecedents [2, 3, 4, 6]. The contribution is their governed composition into one decision-support system:

exact hierarchical attribution → scoped evidence artifact → identity-resolved retrieval → claim contract
→ mechanical validation or refusal → human decision support

The composition changes the role of AI. The model does not infer the accounting identity, choose a denominator, rebuild the bridge, or decide whether partial evidence represents a complete population. It receives a reconciled artifact with declared scope and evidentiary boundaries, then synthesizes that evidence for a decision. The same contract governs the deterministic calculation and the AI that communicates it.

MixShift's contribution lies in how the pieces are designed to operate together. Curated business identities, custom horizontal and vertical bridge mathematics, hierarchy, operator context, structured retrieval, claim controls, validation, and human review form a controlled system in which AI can perform more synthesis with less manual reconstruction. Patent-pending subject matter may cover aspects of the approach; this paper documents the system's governing principles and representative methods without disclosing private implementation recipes, test fixtures, or routing logic.

7.2 Internal validation across three measurement domains

The statements in this section summarize MixShift's internal implementation validation, not independent peer review. Release-specific evidence is recorded in the dated *HCAM Implementation Validation Note* (August 12, 2026; controlled access) [11].

HCAM has been implemented across three distinct Amazon measurement domains:

- Seller Central retail performance;
- Vendor Central wholesale performance; and
- Amazon Advertising.

The domains use different measures, grains, and insufficiency rules. Their common use of one attribution kernel demonstrates that the method is not a single dashboard formula. Metric identities, hierarchies, and source adapters are domain-specific; sign conventions, bases-first aggregation, footing, coverage, provenance, and claim controls are shared.

7.3 The evidence layer is a product contract

The evidence layer has been separated from the application interface and packaged as an independently versioned library. Artifacts carry an engine-version stamp and a capability manifest. Consumers can determine which declared capabilities were computed, partial, or skipped at the capability level, together with the stated reason. Entity- and label-level completeness require explicit census metadata; absent that metadata, the consumer must not infer completeness.

This separation is consequential. An interface can change without rewriting the method. A plugin skill, standalone application, report generator, or broader platform workflow can consume the same evidence contract. The deterministic engine remains the source of attribution rather than allowing each delivery surface to recreate the math.

7.4 A real consumer, not a laboratory demonstration

A governed monthly-report workflow has consumed published analysis artifacts from a live account, composed a complete report without residual analytical queries, and reconciled the report to the source evidence. The case matters because it tests the seam between calculation and AI consumption. It demonstrates that the evidence contract is sufficient for a real analytical output, not merely a theoretical API.

The architecture has also been used for a nonstandard combined June-and-July year-over-year analysis that was not a preconfigured report template. The governed agent assembled the analysis and report in approximately 15 minutes, with the operator's effort concentrated on review. The operator estimated

that equivalent manual work would ordinarily require roughly a day. This is an internal operating observation, not a controlled productivity benchmark. Its significance is narrower but important: ad hoc analytical requests can remain tractable when the agent begins with governed evidence and reusable analytical controls rather than raw data and a blank prompt [11].

Observed defects in prior reports have also been converted into must-fail negative fixtures. The bodies of those fixtures are not public because they encode recognizable operating incidents, but their existence establishes a repeatable governance pattern: when the system fails in a new way, the failure becomes part of the contract.

Release-specific package versions, test counts, tolerances, supported depths, and validation coverage are documented separately [11]. Keeping them in a dated implementation validation note allows this paper to remain accurate as the product evolves.

8. Integrated case: explaining a forecast miss with “ACME”

8.1 The apparent contradiction

“ACME” is an anonymized presentation of a real brand's monthly analysis. In the focal period, revenue was approximately +24 percent year over year yet finished approximately –10 percent versus the fitted forecast. Advertising spend was approximately +84 percent year over year, and TACOS changed by approximately +6 percentage points. A conventional report could choose any one of these facts as the headline and tell a very different story.

The apparent contradiction is exactly the kind of problem the combined architecture is meant to resolve. The brand grew versus the prior year, missed its expected baseline, and invested substantially more in advertising at lower immediate efficiency. Which comparison should govern the decision?

8.2 Forecasting establishes the expectation

The revenue forecast is an independent regression model. It estimates a baseline using organic growth, advertising response, and seasonal structure, and it reports model fit and forecast uncertainty. As new monthly observations enter the training window, its coefficients and diagnostics update.

The current advertising coefficient is therefore not an eternal property of the business. Its level, trend, uncertainty, and stability are part of the evidence. A degrading coefficient can prompt questions about saturation, funnel mix, competition, diminishing marginal return, or lagged brand effects. It does not, by itself, prove any one of those mechanisms.

The forecast's job in the “ACME” case was to establish what July performance was expected to be under the fitted historical structure.

8.3 The event moved; the model's history did not

Prime Day had occurred in July during the relevant training history. In the focal year, it moved to June. The model's July expectation therefore retained a historical event component that could not recur in the focal July.

This created two distortions. Month over month, July followed a June that now contained the major event. Year over year, July was compared with prior Julys that did contain the event. The fitted forecast was also exposed because the event timing had shifted outside the pattern represented in the training data.

The forecast miss was real relative to the model. But it was not sufficient evidence that the operating business had deteriorated by the same amount. Event timing changed the meaning of the baseline.

8.4 HCAM explains the realized movement

HCAM then decomposed the realized period across traffic, conversion, price, advertising, and availability and located material movements across the business hierarchy.

Availability had improved. That mattered because it ruled out one intuitive explanation: measured stockouts did not account for the miss. Traffic and conversion effects could be separated from price effects. Advertising performance could be examined without treating all sales movement as advertising-caused. The analysis retained advertising conversion and efficiency as legitimate operating concerns after the calendar distortion was isolated.

Traffic remained an end node. Advertising clicks relative to sessions could provide a directional signal about contribution and proportionality, but the system did not force the relationship into an exact causal allocation.

8.5 The decision-support narrative

The resulting explanation had five layers:

1. **Expectation:** July finished below the fitted forecast.
2. **Context:** the model's historical July event pattern did not match the focal year's calendar.
3. **Deterministic attribution:** HCAM quantified the realized movement through measured levers and across entities.
4. **Operational exclusion:** improved availability made stockouts an unsupported explanation for the miss.
5. **Investigation:** advertising conversion and efficiency remained areas for action, while the mechanism behind traffic required additional evidence.

This is more useful than declaring the month good because revenue grew year over year or bad because it missed forecast. It tells the operator which comparisons are distorted, which findings are exact, which explanations have evidence, and which questions remain open.

8.6 Rolling coefficients as operating signals

The same architecture can monitor model coefficients over time. If modeled advertising response degrades while spend rises, the agent can combine the trend with HCAM evidence on campaign, product, same-SKU, and cross-product flows. It can ask whether the change is concentrated in particular products, whether investment has moved up the funnel, whether immediate return is being exchanged for longer-term demand, or whether a marketplace is approaching saturation.

Cross-market coefficients can reveal a possible allocation opportunity, but they must be compared under like-for-like specifications, spend ranges, uncertainty, currency treatment, and local context. The coefficient generates a question. HCAM helps locate the measured evidence needed to answer it. Neither should become an automatic budget recommendation without an intervention model and human judgment.

9. Application patterns

9.1 Performance explanation on demand

The immediate use case is a governed monthly or weekly performance explanation. Instead of asking an analyst or AI to start with raw tables, the workflow starts with an evidence artifact that already reconciles. The operator can move from account movement to metric lever to group and entity, then retrieve context and relevant cross-domain evidence.

This can support:

- forecast beat and miss diagnosis;
- revenue, traffic, conversion, price, availability, and advertising bridges;
- account-to-product and account-to-campaign drill paths;
- item-group and portfolio review; and
- detection of large offsetting movements hidden by a flat total.

9.2 Proactive intelligence

Once the evidence system runs on a schedule, the same method can support background monitoring and “did you know?” discoveries. A governed agent can identify a material change, check its coverage and footing, retrieve current goals and events, and present the finding under its permitted claim class.

The discovery should not silently become durable truth. A human can confirm it, reject it, convert it into an active watch, or attach a new event. This creates a path from passive reporting to a closed oversight loop of goals, evidence, review, action, and outcome measurement.

The loop is an architectural direction, not a claim that every component is already generalized across every delivery surface.

9.3 Advertising and cross-sell flows

Advertising data creates a natural extension because the advertised product and purchased product may differ. A governed workflow can analyze advertised-to-purchased ASIN flow to identify possible cross-sell and halo relationships, then use HCAM to determine where the associated performance movement appears.

The system can distinguish same-SKU outcomes, cross-product flow, account-level return, and coverage. It can ask where cross-sell opportunities are concentrated and which products appear capable of absorbing incremental demand. It should not treat observed flow as identified incremental lift without an appropriate experiment or causal model.

9.4 Intervention analysis

Consider a proposed 20 percent promotion over Black Friday. HCAM can show what happened during prior promotions: how revenue changed, which portions were associated with traffic, conversion, and price, where the movement appeared, and how the rest of the catalog behaved during the same event.

That evidence improves an intervention model. Historical elasticity, experiments, controls, competitive conditions, event timing, and uncertainty are still required to estimate incremental revenue. HCAM does not replace those methods. It makes the prior outcomes interpretable and gives the agent a structured basis for testing whether a proposed analogy is valid.

9.5 Forecast oversight

Forecasting and HCAM create a useful feedback relationship. The forecast establishes an expected baseline. HCAM explains the realized beat or miss. Repeated misses can reveal structural breaks, event shifts, coefficient drift, or missing variables. The operator can distinguish a business problem from a model problem and decide whether to change the plan, retrain the model, or investigate the data.

10. Portability beyond Amazon commerce

10.1 The portable method

The current software is Amazon-typed. Its metrics, entity grains, narrative idiom, and adapters reflect Seller Central, Vendor Central, and Amazon Advertising. Claiming that it can accept any arbitrary dataset without configuration would be inaccurate.

The methodology and evidence contract are more portable than the product surface. The domain-independent elements include:

- explicit output/input identities;
- horizontal and vertical decomposition patterns;
- hierarchical rollup and reconciliation;
- evidence provenance and versioning;
- typed insufficiency and capability declarations;
- claim classes and refusal behavior; and
- human promotion of durable context.

10.2 Verticalization is engineering work

Applying HCAM to another domain requires a domain pack:

- metric ontology and identities;
- entity hierarchies and grains;
- source adapters;
- population and denominator rules;
- insufficiency semantics;
- event and context sources; and
- claim and intervention policy.

The right claim is not “works on any dataset.” It is:

Verticalization is an engineering problem, not an unsolved research question.

The Amazon implementation demonstrates the method in a complex domain with multiple measurement systems, ratios, hierarchies, partial populations, and cross-domain evidence.

10.3 Cross-channel commerce

A brand may eventually want performance across marketplaces, direct-to-consumer commerce, retail media, and other channels to pass through one Intelligence Service. The channels need not share identical metrics. They need explicit mappings, governed identities, and a common evidence contract.

HCAM can then preserve local truth while enabling portfolio-level questions: where growth is concentrated, which channels are gaining or losing efficiency, how mix shifts affect the total, and whether a cross-channel recommendation is supported by comparable evidence.

10.4 Financial and operational reporting

Internal financial reporting is another natural domain. Revenue, margin, operating expense, working capital, and cash flow have explicit relationships and business hierarchies. A finance organization can map outputs to inputs, allocate variance across units and accounts, attach budget and forecast evidence, and constrain the AI's explanation to what the ledger and operating context support.

Supply-chain planning, service operations, and other automated control systems present similar patterns. The requirement is not that every domain use the same formula. It is that the domain make its identities, hierarchies, evidence, and claim rules explicit.

11. Limitations and research boundaries

11.1 Structural attribution is not universal root cause

HCAM's strongest claim is deliberately bounded. Exact decomposition verifies how an observed change is allocated under a declared identity. It does not eliminate confounding or prove the upstream mechanism that changed a factor.

That boundary does not make HCAM blind to likely drivers. “What We Know” evidence can connect the exact bridge to relevant operational facts, events, and exclusions. As in the ended-promotion example in Section 5.4, the system can elevate a candidate driver as worthy of action or further testing without pretending that temporal alignment alone identifies the entire causal effect.

Root-cause language requires additional evidence: an event record, experiment, changelog, causal model, or other mechanism-specific source. When that evidence is absent, the system should state a hypothesis or question rather than a conclusion. The value lies in narrowing an otherwise open-ended investigation to the small set of drivers most consistent with the governed evidence.

11.2 End nodes remain bounded by evidence

Traffic, sessions, and impressions are not decomposed further without a governed upstream identity. Cross-domain signals can prioritize investigation but do not create an exact bridge by proximity alone.

Future domain packs or models may extend an end node. The methodological rule remains the same: decomposition stops at the artifact's declared ceiling, and the system states that boundary.

11.3 Fixed-path allocation is path-dependent

Sequential substitution assigns interaction effects according to order. The current method declares the order and uses guardrails when large opposing movements make path sensitivity material relative to the net.

A symmetric all-order variant is feasible for low-dimensional identities and may be useful where no business path is preferred. It is an alternative allocation policy, not a correction required to make fixed-path attribution valid.

11.4 Data and coverage limitations

Results inherit the freshness, completeness, restatements, and definitions of their source systems. Partial coverage constrains rankings and universal claims. Typed absence prevents false precision, but it cannot manufacture missing evidence. Detecting a later source restatement requires a subsequent pull and comparison. A cached artifact reflects the governed source data available at run time and is not automatically made current merely because it is replayable.

The system's refusal behavior should therefore be understood as a feature. A transparent “out of contract” result is more trustworthy than a complete-looking answer assembled from the wrong population.

11.5 Forecast, intervention, and counterfactual evidence is powerful but conditional

A formal forecast tells the operator what was expected under fitted assumptions. An intervention or identified counterfactual model estimates what might happen under an action or what would likely have happened under a defined alternative state. HCAM shows what actually moved, through which measured levers, and where the beat, miss, or estimated effect manifested. With “What We Know” context, the combination can turn a residual into a focused business explanation rather than a generic model error.

That is especially valuable when the companion model supports causal identification. The identified model can estimate the effect of an intervention; HCAM can then show where the realized result

appeared across traffic, conversion, price, products, channels, or other governed dimensions. The two methods answer different questions and become more useful when their evidence is kept distinct.

Forecasts carry model-specification risk, omitted variables, coefficient uncertainty, and exposure to structural change. A coefficient is conditional on the model and training window, not a permanent causal constant. Historical events can move, change form, or interact with new conditions.

Intervention estimates require stronger assumptions than performance attribution. A prior promotion that coincided with growth does not, by itself, identify incremental lift. The model must account for event effects, broader market movement, selection, timing, and uncertainty.

11.6 Context and language-model limitations

Brand Context can become stale or encode an incorrect human judgment. A language model can still make a prose error or draw an unhelpful conclusion. Mechanical controls reduce and detect classes of failure; they do not eliminate the need for review.

Human governance is therefore part of the architecture, not a temporary fallback until the model becomes more capable.

12. Conclusion: from explainable analytics to accountable automation

Organizations want AI to reduce analytical workload, improve performance understanding, and uncover opportunities that humans do not have time to search for. Yet those benefits are difficult to realize when every material answer must be manually rebuilt before it can be trusted. The verification burden limits adoption, and the cost extends beyond avoided errors. Insights that would take days of expert investigation are often delayed, simplified, or never pursued.

The question that led to HCAM was therefore not whether an automated system could produce a recommendation. It was whether the recommendation could be inspected well enough for a human to trust, reject, or modify it for the right reasons.

HCAM addresses the part of that question that can be made deterministic. Horizontal Bridges quantify through which declared metric levers performance moved. Vertical Bridges locate where the movement occurred. The contributions reconcile to governed anchors, preserve scope and coverage, and stop at explicit decomposition ceilings.

The surrounding Intelligence Service turns those results into AI-ready evidence and joins the four functions required for accountable decision support. HCAM supplies deterministic attribution. “What We Know” evidence, events, goals, and Brand Context supply business and mechanism context. Forecast models supply expectations; intervention and identified counterfactual models supply action and

alternative-outcome estimates while retaining their own uncertainty and status. Structured retrieval, skills, claim contracts, validators, refusal behavior, and human review govern how AI uses the evidence.

No one function is sufficient alone. Attribution explains observed movement but does not identify every mechanism or estimate every action. Context makes exact results relevant but does not make them causal. Forecast, intervention, and counterfactual models answer expectation or alternative-outcome questions but remain conditional on their assumptions. Execution controls keep a capable language model from blurring those boundaries. HCAM is the foundational analytical spine because it gives the system a reconciled account of what actually happened before the AI interprets what it means or what to do next.

This architecture does not promise that every business question has a causal answer. It promises to keep different kinds of answers from being confused. Deterministic attribution, directional evidence, mechanism evidence, intervention estimates, and forecast comparisons each have a place. Causal counterfactuals remain a separate claim requiring an identified design. The system's credibility depends on preserving those boundaries.

HCAM gives AI a map of the evidence that matters before asking it to explain the business. That can compress an investigation that once required a sophisticated operator to traverse multiple models and datasets, while leaving a traceable path for review. The method is being proven in Amazon commerce, but its core pattern applies wherever organizations can declare the relationships among outputs and inputs, preserve hierarchy, and govern the claims made from the result.

The purpose of an Oversight Layer is not to slow automation down. It is to remove a practical barrier to using AI for consequential analysis while keeping increasingly automated systems governable by the humans who remain accountable for their decisions.

Appendix A. Worked H-Bridge and Shapley comparison

Assume the comparison period has 100,000 sessions, 10 percent conversion, and a \$20 average price:

$$R_1 = 100,000 \times 0.10 \times \$20 = \$200,000$$

The current period has 110,000 sessions, 9 percent conversion, and a \$22 average price:

$$R_2 = 110,000 \times 0.09 \times \$22 = \$217,800$$

The observed revenue change is \$17,800.

Under the declared Traffic → Conversion → Price path:

- Traffic contribution: $(110,000 - 100,000) \times 0.10 \times \$20 = +\$20,000$

- Conversion contribution: $110,000 \times (0.09 - 0.10) \times \$20 = -\$22,000$
- Price contribution: $110,000 \times 0.09 \times (\$22 - \$20) = +\$19,800$

The contributions reconcile exactly:

$$\$20,000 - \$22,000 + \$19,800 = \$17,800$$

An all-order Shapley allocation averages the marginal contribution of each factor across the six possible factor orders. For the same example, the approximate allocation is:

Factor	Fixed path	All-order Shapley	Interpretation
Traffic	+\$20,000.00	+\$19,933.33	Growth in sessions contributed positively
Conversion	-\$22,000.00	-\$22,066.67	Lower conversion offset the traffic and price gains
Price	+\$19,800.00	+\$19,933.33	Higher price contributed positively
Display rounding	+\$0.00	+\$0.01	Reconciles rounded cents to the exact unrounded allocation
Total	+\$17,800.00	+\$17,800.00	Both displayed allocations foot exactly

The exact Shapley values are repeating decimals, so the table discloses the one-cent reconciliation created by displaying each factor to cents. The two methods agree on direction and produce similar magnitudes in this case. The fixed path assigns interaction terms to the declared business path. Shapley averages them symmetrically. Neither calculation explains why sessions, conversion, or price changed.

Appendix B. Evidence and claim matrix

Evidence available	Default claim class	Example	Evidence needed for a different claim
Governed metric values and exact bridge	C1 structural attribution	"Conversion contributed negatively to the observed revenue change."	Mechanism-specific evidence
Multiple measured domains with compatible scope	C2 directional observation	"Clicks rose more slowly than sessions, weakening the case that paid traffic explains the full increase."	Identified mapping or experiment
Event record plus aligned measured movement	C3 operational explanation	"The event-calendar shift is consistent with the inflated July baseline."	Stronger identification for causal effect size
Experiment, elasticity model, or identified response model	C4 intervention or identified counterfactual estimate	"Under the stated model, the proposed price change is expected to..."	Prospective validation and monitoring
Formal forecast with fit and uncertainty	C5 forecast/expectation comparison	"Actual revenue finished below the expected baseline."	Evidence identifying the miss mechanism or a causal counterfactual

References

1. Tabassi, E. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1, National Institute of Standards and Technology, 2023. <https://doi.org/10.6028/NIST.AI.100-1>. Accessed August 12, 2026.
2. Lewis, P., Perez, E., Piktus, A., et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Advances in Neural Information Processing Systems 33*, 2020. <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>
3. Shapley, L. S. "A Value for n-Person Games." In H. W. Kuhn and A. W. Tucker, eds., *Contributions to the Theory of Games II*, Princeton University Press, 1953, pp. 307–317.
4. Lundberg, S. M., and Lee, S.-I. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems 30*, 2017. https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
5. Pearl, J. "A Causal Calculus for Statistical Research." *Pre-proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics*, PMLR R0, 1995, pp. 430–449; reissued by PMLR in 2022. <https://proceedings.mlr.press/r0/pearl95a.html>
6. Das Gupta, P. *Standardization and Decomposition of Rates: A User's Manual*. Current Population Reports, Series P23-186, U.S. Bureau of the Census, 1993. <https://www.census.gov/library/publications/1993/demo/p23-186.html>
7. Dziri, N., Lu, X., Sclar, M., et al. "Faith and Fate: Limits of Transformers on Compositionality." *Advances in Neural Information Processing Systems 36*, 2023. https://proceedings.neurips.cc/paper_files/paper/2023/hash/deb3c28192f979302c157cb653c15e90-Abstract.html
8. Niu, C., Wu, Y., Zhu, J., et al. "RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models." *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024, pp. 10862–10878. <https://doi.org/10.18653/v1/2024.acl-long.585>
9. World Wide Web Consortium. *PROV-O: The PROV Ontology*. W3C Recommendation, April 30, 2013. <https://www.w3.org/TR/prov-o/>
10. Autio, C., Schwartz, R., Dunietz, J., et al. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1, National Institute of Standards and Technology, 2024. <https://doi.org/10.6028/NIST.AI.600-1>. Accessed August 12, 2026.
11. MixShift. *HCAM Implementation Validation Note*. Internal controlled-access validation record, August 12, 2026.
12. McKinsey & Company. "The State of AI in 2025: Agents, Innovation, and Transformation." November 5, 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai/>. Accessed August 12, 2026.
13. Gillespie, N., Lockey, S., Ward, T., Macdade, A., and Hassed, G. *Trust, Attitudes and Use of Artificial Intelligence: A Global Study 2025*. University of Melbourne and KPMG, 2025. <https://doi.org/10.26188/28822919>. Accessed August 12, 2026.
14. Amazon Science. "reMARS Revisited: Amazon's Supply Chain Optimization." 2022. <https://www.amazon.science/latest-news/remars-revisited-amazons-supply-chain-optimization>. Accessed August 12, 2026.

Publication note

This whitepaper documents the HCAM methodology and the stable architecture of the MixShift Oversight Layer. Release-specific implementation evidence, including package versions, test results, supported depths, closure tolerances, and deployment status, is maintained in a separately dated validation note. Current access and delivery surfaces are maintained on the MixShift product page.

About the author

Todd VanderStelt is the founder and Managing Partner of MixShift, where he focuses on forecasting, causal modeling, and marketplace analytics. A former Amazonian and Amazon Bar Raiser with more than 15 years of experience across ecommerce strategy, diagnostics, and client leadership, he has advised hundreds of brands across categories and growth stages. He developed the HCAM methodology through a decade-long effort to make automated business recommendations more inspectable, explainable, and actionable.