Hybrid Cognition in Business

Working Hypotheses for KPI/PPI Optimization
Reference Models, Participant Experiments, and Shared Learning

A working paper for participation and testing in the Hybrid Cognition Forum

John Aaron, PhD

www.ratio-weekly.com
August 30, 2026 · Version 7

Working proposition: Hybrid Cognition can be used to improve Key Performance Indicators (KPIs) and Process Performance Indicators (PPIs) when human judgment and AI are deliberately combined, evidence remains auditable, and risk is managed. The Forum will test this proposition through existing reference models, participant experience, competing designs, and measured outcomes.

Gaining an Organizational Start With AI Efficiently and Safely

The Hybrid Cognition Forum brings participants together to test, challenge, refine, and extend working hypotheses about AI implementation across business disciplines. Its practical aim is to give organizations a head start in adopting AI efficiently and safely, with measurable business results. Testing shared reference models helps participants identify promising approaches and recognize unproductive paths before committing substantial time and resources. On-Time Delivery, Customer Retention, business forecasting, and project-risk reduction applications provide concrete reference models around which that work can begin. Each makes a proposed connection between evidence, human-machine decisions, and measurable performance available for examination.

The aim is to develop a cumulative empirical discipline for applied artificial intelligence in business process contexts. A reference model is a working starting point that participants can inspect, reproduce, adapt, or compare with alternatives. Results from one setting establish evidence within that setting; broader principles require further tests and counterexamples. The ten hypotheses below are provisional, including the proposed benefits of governance and risk management. Recommendations about safe participation are operating requirements for testing, rather than claims that every hypothesis has already been proved.

In practice the same analytical problem appears repeatedly across projects, markets, customer behavior, order fulfillment, operational risk, and other business processes,. Organizations collect large quantities of structured data, build dashboards, and increasingly deploy predictive models. Yet decision-makers still struggle with the ability of AI to answer the questions that matter most: Is performance departing from what should be expected? Is the departure temporary or structural? What is driving it? What might happen next? Should we let AI take actions autonomously? Where does human involvement come in?

The recurring design pattern in the reference models is a machine-assisted perception and decision process. It maintains an expected state, watches heterogeneous evidence, detects meaningful deviation, represents uncertainty, and presents alternative explanations or actions.

Human contributions may include context, novelty, forward information, causal scrutiny, values, and accountability. Participants can ask where each contribution adds value and where the allocation should change in their own organizational contexts.

Working Hypothesis 1: Organize Around the KPI and Its Supporting PPIs

The most practical organizing principle is an accepted business measure. A Key Performance Indicator (KPI) measures an outcome the organization is trying to achieve, such as On-Time Delivery or Customer Retention. A Process Performance Indicator (PPI) measures how the process producing that outcome is operating, such as schedule adherence, queue age, material availability, issue recurrence, customer-engagement change, or intervention lead time. A hybrid-cognition system should begin with an accountable KPI and the PPIs that help explain, predict, or influence it.

Business question	KPI / supporting PPIs	Expected trajectory	Decision enabled
Will we deliver what we promised?	OTD / commitment and flow PPIs	Feasible commitment / expected OTD	Accept, negotiate, expedite, reallocate
Is the project entering structural trouble?	Delivery KPIs / schedule, cost and risk PPIs	Plan and expected project trajectory	Investigate, recover, escalate
Is a customer relationship deteriorating?	Customer Retention / behavior and intervention PPIs	Expected customer behavior	Intervene, offer, route, retain
Is a market regime becoming abnormal?	Market outcome / regime and volatility PPIs	Contemporary fitted trajectory	Investigate, hedge, prepare contingencies

This KPI/PPI-first discipline matters for four reasons. First, it anchors AI to business value. Second, it defines what better means and therefore how the system can be validated. Third, PPIs expose the process mechanisms and leading signals through which the KPI may be improved. Fourth, the paired measures reveal whether a local intervention improves the target KPI by shifting cost, delay, quality loss, or risk elsewhere. A useful hybrid system should ultimately be judged by prospective improvement in the KPI and by credible evidence that the supporting process also became more reliable, timely, economical, or controllable.

Forum test: define the baseline, measure, owner, observation period, desired improvement, and acceptable cost and risk before changing a reference model. Compare KPI-focused and paired KPI/PPI evaluations. Check whether an apparent improvement survives accounting for false alarms, workload, quality, and effects on other processes.

Working Hypothesis 2: Departures from an Expected Trajectory Provide Useful Warning

A recurring architecture is to represent the process as an expected trajectory and then model departures from it. The level alone is often less informative than the deviation from what should be happening.

Observed State = Expected Trajectory + Deviation

This is explicit in the IUVO forecasting reference example, where the S&P 500 forecast is constructed by forecasting DEV_NEW—the market's deviation from a contemporary fitted trajectory—and adding the forecast deviation back to the future trajectory. The same logic generalizes. A project can be compared with its planned trajectory; customer behavior with its learned trajectory; fulfillment performance with its feasible commitment trajectory.

The practical advantage is early warning. A system need not know the final outcome exactly to recognize that the process generating the outcome is changing. That distinction separates point prediction from detection of an emerging adverse state.

Reference Model: IUVO Forecasting

IUVO reference-model test: preserve the production forecast as the baseline and record prospective analyst inputs before outcomes are known. Compare the baseline with narrative-conditioned forecasts on held-out periods using forecast error, calibration, and warning lead time. These are analytical process measures; a business KPI must also reflect the decision the forecast supports. The Forum can test the marginal contribution of human forward information without assuming that a better forecast automatically produces a better business outcome.

Working Hypothesis 3: Narrative Can Add Decision-Relevant Evidence

A second recurring theme is that structured operational data alone rarely describes the entire state of the business. People continuously create potentially useful evidence in documents such as CRM notes, project status reports, customer complaints, service interactions, emails, management commentary, news, and other text. Modern language models make it practical to convert that unstructured material into structured analytical variables.

Narrative source	Possible structured measures	Business use
CRM / account notes	topics, sentiment, urgency, competitor mentions, objections, change rate	condition churn, opportunity and account-health models
Project reports	risk language, contradiction, uncertainty, causal claims, acceleration	detect structural deterioration before lagging metrics
Customer/service text	complaint themes, intensity, recurrence, novelty	explain KPI deterioration and identify emerging causes
External narrative	policy, financial, geopolitical, health, disaster narratives	condition forecasts when the external environment changes

The key point is not merely that narrative adds more predictors. Narrative can add explanatory power by describing mechanisms, context, intent, expectations, and environmental change that are poorly represented in transactional fields.

Working Hypothesis 4: Monitoring Changing Relationships Preserves Usefulness

Machine-learning performance in business should never be assumed constant. A model that explained or predicted an outcome well when trained can lose power because the environment

changes: customers change, competitors change, policy changes, incentives change, technology changes, management changes, and the process itself changes.

$$\textbf{Predictive Power}_t = f(\textbf{Model, Data, Current Environment}_t)$$

This creates a model-governance problem that conventional accuracy monitoring often detects only after performance has deteriorated. Narrative data offers a potentially earlier conditioning signal. People may begin describing a new competitor, changed buying behavior, policy shock, supply problem, management directive, or emerging failure mode before those changes are fully visible in the structured KPI history.

Narrative therefore has two major capabilities: it can improve the prediction directly, and it can provide evidence about whether the historical relationships on which the prediction depends are still valid.

Forum test: compare structured-data-only results with narrative-augmented results, including periods of environmental change. Measure whether narrative improves warning lead time or calibration and whether an apparent regime-change signal produces excessive false alarms. Preserve source observations so participants can examine whether extracted narrative variables actually represent the underlying evidence.

Three Narrative Mechanisms to Compare in the Reference Models

Narrative enters hybrid cognition in more than one way. The IUVO forecasting example makes the prospective human override especially visible, but operational applications such as an On Time Delivery (OTD) KPI Optimization and PRIMMS show an equally important use: narrative can operate continuously as a sensor, augmenting structured data before any explicit human override is required.

Role 1 — Continuous narrative evidence: augment machine perception

In operational settings, narrative from CRM notes, work-center risk reports, project status reports, customer complaints, and similar sources can be converted into structured variables and fused continuously with ERP, schedule, capacity, quality, and other quantitative evidence. The narrative is not an override; it is part of the current evidence state. This is the central narrative mechanism in the OTD KPI Optimization example and an important part of PRIMMS.

Role 2 — Narrative as a model-conditioning and regime-change sensor

Narrative can also provide evidence that the environment supporting a predictive relationship is changing. New customer objections, recurring work-center concerns, supplier commentary, policy changes, or other emerging themes may appear before their effects are fully visible in the KPI history. Narrative therefore can help answer not only “What is the KPI likely to do?” but also “Should we still trust the historical relationships on which this model depends?”

Role 3 — Prospective human Narrative Override

In forecasting applications such as IUVO reference model, the analyst may possess forward information that is not yet in the observed data. The analyst expresses that information as a difference from the machine/default narrative expectation. If the analyst leaves the controls at their defaults, the override is zero and the machine forecast stands. If the analyst changes them, the system translates only the difference; a separate evidentiary-weight control expresses

confidence in the human information. This is a disciplined prospective intervention rather than an opaque manual forecast replacement.

$$\text{Narrative Override}_i = \text{Human Expected Narrative}_i - \text{Default Expected Narrative}_i$$

Working Hypothesis 5: Preserving Alternatives Improves Decisions Under Uncertainty

Another recurring design principle is to avoid collapsing uncertainty into one answer too early. A central forecast can remain visible while the system also represents alternative processes and their probabilities (p)—for example, normal continuation, material deterioration, and abrupt impulse.

$$\{\text{Path 1}, p_1\} \quad \{\text{Path 2}, p_2\} \quad \{\text{Path 3}, p_3\}$$

Human forward information can then condition both the shape of those alternatives and the probability assigned to them. This is particularly useful when the business question is not "What exact number will occur?" but "Has the probability of an abnormal state increased enough that we should investigate or act?"

A Common Decision Cycle for Hybrid Cognition Processes

1. Define the KPI and PPIs	Start with the accountable outcome, supporting process measures, and decision owner.
2. Establish expected trajectory	Define what normal or planned performance should look like.
3. Sense broadly	Combine structured operational measures with structured narrative evidence.
4. Detect deviation	Measure whether the system is moving away from expected behavior.
5. Accumulate evidence	Avoid reacting to one weak signal; combine corroborating evidence and uncertainty.
6. Generate alternatives	Maintain competing explanations, paths, causes, and courses of action.
7. Admit human forward information	Use explicit Narrative Overrides or comparable structured judgment mechanisms.
8. Recompute consistently	Let the machine translate the human contribution into revised forecasts, probabilities, or options.
9. Decide and act	Preserve human authority where consequences, values, novelty, and accountability require it.
10. Learn prospectively	Save machine prior, human inputs, decision, and outcome; measure whether either actually added value.

Consistent with John Boyd's emphasis on orientation, the working proposition is that better understanding of the situation should precede faster action. The common cycle is observations and percepts → model of the situation → causal interpretation → alternative courses of action →

judgment → action → new evidence. Deterministic calculation, statistical models, semantic methods, human review, and bounded automation can be assigned to different stages. The cycle makes failures in evidence, interpretation, execution, and feedback visible across different technology stacks.

Reference Model: PRIMMS and Project Performance

Project risk management follows the same paradigm as continuous processes especially clearly because the project already has explicit KPIs, a planned trajectory, repeated observations, heterogeneous evidence, uncertainty, and a responsible human decision-maker. PRIMMS® is an exemplar of the architecture: it surrounds conventional project controls with a machine-assisted perception layer designed to recognize emerging structural risk while recovery options still exist.

The project plan is the expected trajectory. Schedule, cost, milestone, risk, and delivery measures provide structured evidence. Weekly status reports, issue descriptions, management commentary, and other narrative sources add evidence that may reveal changing conditions before the lagging KPIs fully register the problem. PRIMMS accumulates this evidence rather than treating a single threshold breach as sufficient proof.

Project KPI → Planned trajectory → Deviation → Structured + narrative evidence → Evidence accumulation → Competing explanations → Courses of action → Human judgment → Decision → Outcome learning

This is also where the distinction between early warning and exact prediction becomes operationally important. A project manager does not need a perfect forecast of the final cost or completion date to benefit from knowing that the probability of a structurally adverse state has materially increased. The value lies in recognizing the condition early enough to investigate, preserve alternatives, and intervene before the KPI deterioration becomes irreversible.

Human judgment is not an after-the-fact approval step.

The human contributes information and reasoning that may not yet exist in the machine-readable record. A project manager may know that a supplier delay is recoverable, that a nominally green milestone depends on an assumption that has ingloriously failed, that a key technical resource is about to leave, or that sponsor support is weakening. Those facts can materially change the interpretation of otherwise identical quantitative evidence.

The machine contributes persistence, breadth, historical comparison, probabilistic evidence accumulation, pattern recognition, and consistent recomputation. The project manager contributes context, prospective knowledge, causal judgment, understanding of consequences, and accountability. PRIMMS therefore illustrates the intended division of cognitive labor: the system improves Observe and Orient; the human retains Decide and Act.

A mature implementation should preserve the machine's prior assessment, the human's added information or override, the action selected, and the realized project outcome. That makes both machine performance and human judgment prospectively testable. Over time, the organization can learn not only whether PRIMMS predicted risk, but when human intervention improved or degraded the decision.

Forum test: compare conventional project controls with the PRIMMS evidence-accumulation process using the same project information and observation dates. Examine warning lead time, false alarms, recovery options, and subsequent schedule or cost outcomes. Participants can test whether improved orientation creates actionable time before a milestone fails, while preserving the project manager's recorded judgment and decision.

Reference Model: On-Time Delivery Optimization

The On-Time Delivery (OTD) Key Performance Indicator optimization example on the Ratio Weekly website demonstrates a different but complementary form of hybrid cognition. The business objective is explicit: improve OTD and the quality of the customer commitment. Supporting Process Performance Indicators can include schedule adherence, queue age, material readiness, work-center reliability, first-pass quality, commitment confidence, and recovery lead time. The machine evaluates structured operational evidence such as capacity, queue position, material availability, routing, work-center condition, and the probability that a requested date is attainable. At the same time, narrative evidence from CRM and work-center risk information is continuously converted into structured analytical evidence and added to that assessment.

This matters because two orders with similar ERP fields may not represent the same decision. CRM narrative may indicate that a customer is strategically important, has already experienced repeated service failures, cannot tolerate slippage, or is considering an alternative supplier. Work-center narrative may reveal emerging congestion, quality instability, maintenance concerns, staffing limitations, or other conditions not yet fully reflected in historical cycle-time statistics. Those narratives alter the evidence state before the commitment decision is made.

OTD decision state = f(ERP / operational evidence, CRM narrative, work-center risk narrative)

The human role is also different from a forecast override. Sales, customer-service, operations, and work-center personnel generate much of the contextual narrative in the normal course of work. Their observations become machine-readable evidence. The decision-maker can then judge the resulting probability and alternatives in light of customer consequences and operational realities. Hybrid cognition is occurring continuously: human observation enriches machine perception; the machine integrates evidence consistently; and the responsible person retains the commitment decision.

This example also shows why pairing the KPI with PPIs is essential. OTD supplies the measurable business outcome; the PPIs reveal where the fulfillment and commitment process is strengthening or failing. The value of narrative can therefore be tested prospectively: does adding CRM and work-center narrative improve calibration, lead time, commitment quality, customer outcomes, OTD performance, or the process indicators that precede those outcomes beyond the structured-data model alone? The combined scorecard also discourages a superficial OTD improvement achieved through excessive inventory, unstable rescheduling, unrealistic customer promises, or cost transferred to another part of the operation.

Forum test: evaluate incoming orders using the same backlog, requested dates, customer priorities, and capacity assumptions. Compare structured operational evidence alone with the narrative-augmented assessment. Track OTD, promise-date reliability, queue stability, expedite

frequency, and cost. Challenge the reference model's policies and propose alternatives that improve commitments without destabilizing the schedule.

Reference Model: Customer Retention

Customer Retention is another widely accepted KPI around which hybrid cognition can be designed and tested. The KPI records whether valuable relationships are preserved. Supporting PPIs provide the earlier process view: changes in order or usage trajectory, engagement, service burden, complaint recurrence, relationship coverage, competitor signals, account-team follow-up, and the time between a credible warning and a retention action. The practical objective is to recognize a deteriorating relationship while management still has choices that can change the outcome.

The Ratio Weekly customer-retention proof of concept examines this problem with 175 synthetic business-to-business customers observed over 156 weeks, including 18 known departures. Multiple model architectures are compared against locked outcomes and an explicit economic measure that weighs Economic Detection Potential against False Positive Penalty. The study is a controlled research demonstration rather than a production-performance claim, but it illustrates a testable proposition: earlier, well-calibrated detection can preserve more recoverable value than equally accurate detection that arrives after the intervention window has narrowed.

The hybrid design combines machine surveillance with field evidence the behavioral history may not yet contain. A model can identify an elevated but not-yet-confirmable departure pattern; an account team may know that a customer champion has left, a competitor has entered the conversation, sentiment has changed, or a budget event is approaching. Bayesian Weight of Evidence expresses machine and human evidence in a common log-odds unit, preserves each as a separate named contribution, and recomputes one posterior probability. Neither source silently overrides the other.

The study's C011 example makes timing concrete. Every tested path ultimately identifies the same true departure, but the earlier surveillance result carries $713.1 thousand of modeled Economic Detection Potential compared with $54.0 thousand for the later confirmation path. The $659.1 thousand difference comes from timing rather than classification. The example supports a broader hypothesis for Forum testing: retention systems should be evaluated not only by eventual accuracy, but also by warning lead time, false-positive cost, actionability, recoverable value, and the quality of the retention process that follows the warning.

Customer Retention KPI → behavioral and narrative PPIs → machine evidence + logged human evidence → auditable posterior → proportionate retention action → realized outcome → recalibration

Forum test: compare machine-only and machine-plus-human assessments with inputs recorded prospectively. Examine retention outcomes, intervention lead time, false-positive cost, and realized value. Evidence weights must be calibrated or clearly labeled as judgment; dependence between signals must be checked to avoid double-counting. Synthetic detection economics motivate a test, while real retention improvement requires evidence about the interventions and outcomes themselves.

Extending the Reference Models to CRM and Enterprise Analytics

CRM is a particularly natural environment for this architecture because it already sits near the intersection of KPIs, structured transaction data, human-entered narrative, workflows, forecasts, and decisions. The opportunity is not simply to summarize account notes, sentiment or add another score. It is to connect those elements into a governed perception-and-decision loop.

A customer application, for example, could maintain expected account trajectories for revenue, usage, service load, renewal probability, or margin; convert account narratives into structured evidence; detect when the relationship between historical predictors and outcomes appears to be changing; present alternative account futures; and let the account team contribute explicit forward information before the system recomputes risk and recommended actions.

The same pattern applies to sales forecasts, service performance, supplier management, project delivery, and operational commitments. The specific models change; the cognitive architecture remains remarkably stable.

Working Hypothesis 6: Explicit Human Contributions Can Add Measurable Value

Hybrid cognition changes the human role rather than eliminating it. As machines become better at continuous sensing, pattern recognition, probabilistic calculation, retrieval, and consistent recomputation, human value shifts toward the forms of judgment that remain difficult to infer reliably from historical data alone. The central workforce question therefore is not simply whether people can use AI tools. It is whether they can contribute information and judgment that materially improve a KPI/PPI-centered decision.

Where human judgment adds value

The recurring reference examples in this paper and found on the www.ratio-weekly.com website suggest several high-value human contributions. People can recognize context that is absent from the data; introduce prospective information before it becomes an observable historical signal; challenge a model when the environment has changed; distinguish plausible association from a credible causal mechanism; judge consequences and tradeoffs that cannot be reduced to prediction accuracy; generate or reject alternative courses of action; and accept accountability for decisions whose effects extend beyond the model objective.

This means that the human role should be designed into the analytical architecture. In IUVO, the analyst can contribute a prospective Narrative Override and state the evidentiary weight placed on that information. In OTD KPI Optimization, sales, CRM, operations, and work-center personnel continuously contribute observations that become structured evidence before a commitment decision. In PRIMMS, project managers interpret an accumulating risk state, assess competing explanations, develop courses of action, and retain Decide and Act authority. These are different mechanisms, but each makes human contribution explicit and testable.

Comparative capability is the design question. For each reference-model task, participants should identify what a person, deterministic rule, statistical model, or LLM contributes and how each can fail. Human decision authority in the present examples is a reference configuration.

Greater automation or greater human involvement can be tested where the consequences, controls, and evidence justify it.

Judgment must itself become more disciplined

Human judgment is not automatically superior to machine inference. People bring their own biases, incomplete information, organizational incentives, anchoring, overconfidence, and inconsistent treatment of evidence. Hybrid cognition therefore should not create an unrestricted manual override. It should structure judgment: identify what the human believes is different from the machine assumption, record the evidence supporting that belief, express uncertainty or confidence where possible, preserve the machine prior, and later compare both assessments with the realized outcome.

Machine prior + explicit human evidence + stated confidence → recomputation → decision → outcome → learning

This creates an important organizational learning opportunity. Firms can go beyond measuring model accuracy, but judgment accuracy: Under what conditions do particular roles, teams, or experts add predictive or decision value? When does human intervention improve the KPI? When does it degrade it? Which types of information consistently arrive through people before they appear in structured systems? Human judgment can therefore become an improvable organizational capability rather than an unmeasured exception to automation.

Future skills for a hybrid-cognition workforce

The resulting skills agenda is broader than prompt engineering or general AI literacy. Many employees will need enough statistical and analytical literacy to understand probabilities, uncertainty, model error, changing relationships, and the difference between prediction and causation. They will need to recognize when a model is operating outside the environment in which it learned, and to understand why narrative evidence may signal that a once-useful relationship is becoming stale. They will also need a new form of semantic and hypothesis literacy: the ability to frame an anomaly clearly, provide the relevant evidence state to an LLM, use the model to search beyond the team's local experience for analogous mechanisms and alternative courses of action, and then discriminate rigorously among the candidates it produces.

This changes the skill requirement for root-cause analysis. The scarce human capability is no longer only the ability to remember a personally familiar failure mode. It increasingly includes asking whether the current anomaly resembles mechanisms seen in other industries, functions, technologies, or operating environments; separating a useful analogy from a superficial resemblance; identifying what additional evidence would distinguish competing explanations; and knowing when an LLM-generated hypothesis is unsupported, infeasible, or inconsistent with local facts. The same discipline applies to courses of action: people must be able to evaluate a broader machine-generated option set for feasibility, unintended consequences, organizational fit, ethics, and risk rather than simply selecting the first plausible recommendation.

The practical workforce-development agenda therefore includes evidence literacy, narrative-to-data thinking, model and regime-change awareness, anomaly framing, semantic search and analogy evaluation, causal and systems reasoning, hypothesis testing, alternative-generation and option evaluation, KPI and decision literacy, prospective judgment discipline, and AI

governance. These are complementary skills: the machine can broaden the search space, but the human must know how to interrogate, test, narrow, and act on that expanded space.

Future skill	Why it matters in hybrid cognition
Evidence literacy	Interpret probabilities, confidence, uncertainty, corroboration, and competing evidence rather than treating model output as fact.
Narrative-to-data thinking	Recognize operationally meaningful information in CRM notes, project reports, customer language, work-center observations, and other unstructured sources.
Model skepticism without model rejection	Know when to trust a model, when to challenge it, and what evidence is required to justify a departure from the machine prior.
Causal and systems reasoning	Distinguish correlation from mechanism; understand dependencies, feedback, constraints, second-order effects, and changing environments.
Scenario and alternative generation	Develop plausible alternative futures and courses of action instead of prematurely collapsing uncertainty into one answer.
KPI/PPI and decision literacy	Connect analysis to the measurable outcome, supporting process, decision owner, intervention point, and economic or operational consequence.
Prospective judgment discipline	State assumptions, forward information, confidence, and rationale in a form that can be recorded and tested against outcomes.
AI collaboration and governance	Understand the division of authority between human and machine, escalation rules, auditability, and accountability for action.

Workforce development becomes part of implementation

If organizations deploy hybrid systems without developing these capabilities, the human side of the architecture can become the limiting factor. People may over-trust machine outputs, override them without evidence, fail to record valuable narrative, accept a semantically plausible root-cause story without testing it, or be unable to distinguish genuine regime change from noise. They may also underuse the technology by asking an LLM only to summarize what the local team already knows rather than using it to expand the hypothesis and course-of-action search space. Workforce development should therefore accompany model development. Training should use the organization's own KPIs, PPIs, anomalies, and decisions, allowing people to practice reading evidence states, structuring narrative, framing exceptions, asking for competing causal hypotheses, searching for analogues beyond local experience, generating non-obvious courses of action, challenging models, documenting judgment, and learning from realized outcomes.

A useful training exercise is to preserve the team's initial explanation and proposed actions before consulting the LLM, then compare them with the expanded candidate set produced by semantic search. The team can document which additional hypotheses or courses of action were genuinely novel, which survived evidentiary scrutiny, which were rejected and why, and

whether the resulting decision improved the KPI or its supporting PPIs. This turns the often-vague idea of "wisdom of the crowd" into a measurable learning process and helps people develop judgment about when expanded machine search actually adds value.

The long-term objective is a workforce that is neither displaced by machine intelligence nor asked merely to supervise it. It is a workforce capable of adding information, interpretation, causal scrutiny, and judgment precisely where those contributions improve measurable business outcomes and capable of using LLMs to reach beyond local experience without surrendering evidentiary discipline or decision responsibility. That is the human side of engineered complementarity.

Working Hypothesis 7: Semantic Expansion Improves the Search for Causes and Actions

A large share of business management is exception management and the expensive manual resolution of those exceptions. Dashboards, control systems, risk processes, customer-management systems, and operating reviews are designed to expose departures from expectation: an unusual KPI movement, a late milestone, a customer behavior change, a work-center constraint, a quality excursion, or some other condition that deserves attention. Once an anomaly is detected, however, the harder questions begin: Why is this happening, and what can we do about it?

Traditional analysis is constrained by the experience available locally. A project team, sales group, plant staff, or executive team naturally searches for causes and responses within the cases, practices, and mental models it already knows. That local expertise is indispensable, but it creates a bounded search space. Novel anomalies are especially difficult because the most useful analogue, causal hypothesis, or course of action may lie outside the team's direct experience.

LLMs expand the semantic search space

Large language models offer a new capability at this point in the decision process. Their value is they can search a much broader semantic space of patterns, mechanisms, analogous situations, failure modes, and response strategies for recovery than a local team can usually generate from memory. Given a well-described anomaly and its evidence state, an LLM can help construct a wider set of plausible causal hypotheses and potential courses of action for investigation.

Anomaly + evidence → expanded semantic search → candidate causes → candidate courses of action → human evaluation

This is particularly relevant to the project risk management tool PRIMMS, where an emerging project-risk condition can be examined against a broader semantic landscape of project failure mechanisms and recovery strategies; to OTD optimization, where an unusual commitment or work-center condition can trigger hypotheses that cross manufacturing, supply, customer, and policy domains; and to CRM, where an unexpected change in customer behavior can be considered against commercial, competitive, service, organizational, and external explanations.

A form of "wisdom of the crowd" — without surrendering judgment

The useful analogy is a computational form of "wisdom of the crowd." The LLM has absorbed patterns expressed across a very large body of human language and can surface possibilities for recovery that the immediate team may not have considered. This does not make those possibilities true, nor does it convert semantic association into causal evidence. It changes the search problem: instead of asking the local team to originate every plausible explanation or action from scratch, the machine can generate a broader candidate set that people can test against local facts, constraints, and consequences.

The distinction is critical. Root-cause analysis remains an evidentiary task. The LLM should generate hypotheses, analogues, questions, and candidate mechanisms; structured and narrative evidence should then be used to discriminate among them. Likewise, courses of action should be treated as alternatives to evaluate rather than recommendations to accept automatically. Human judgment remains responsible for determining plausibility, causal adequacy, feasibility, ethics, organizational fit, and acceptable risk.

From anomaly detection to actionable analytics

This extends the KPI/PPI-centered architecture beyond detection. A conventional dashboard may identify that the KPI is abnormal. A hybrid-cognition system can ask what evidence explains the abnormality, broaden the search for causes beyond local experience, generate courses of action that might otherwise be missed, and preserve uncertainty until the responsible person decides. In that sense, LLMs can help convert anomaly detection into actionable analytics.

The recurring pattern becomes: detect the exception; assemble structured and narrative evidence; use the LLM to expand the hypothesis and action space; test those possibilities against the evidence; and apply human judgment before action. The machine contributes breadth of semantic search. The human contributes local reality, causal scrutiny, consequences, and accountability.

Working Hypothesis 8: Auditable Outputs Make AI-Assisted Work More Dependable

Large language models contribute two especially important implementation capabilities. They can accelerate code writing, testing, documentation, and application construction; and they can expand the semantic search space for causal hypotheses, analogues, and courses of action. These capabilities can shorten development cycles and broaden the reasoning available to a team. They also make intermediate-output governance central to implementation quality.

A governed hybrid-cognition application should preserve enough of the analytical chain for an independent reviewer to reproduce and challenge the result. Depending on the application, that record can include source references, data transformations, model and code versions, prompts or specifications, intermediate calculations, thresholds, evidence contributions, uncertainty, human additions, selected actions, and realized outcomes. The interface should expose decision records, validation files, or other machine-readable outputs instead of asking the reviewer to trust a polished narrative or dashboard alone.

The customer-retention proof of concept illustrates this balance. AI systems wrote substantial portions of the code and sharply compressed elapsed development time. Human quality

management then challenged fabricated results, logic inversions, and double-counting; customer-level validation tabs traced the calculations to source CSV files; and cross-validation independently reconciled bottom-up results with locked dashboard values. The lesson is broader than the particular tools: AI-assisted speed moves more implementation risk toward evidence quality, audit design, and independent validation.

This yields another testable principle: every material recommendation should have an audit path proportional to its consequence. A reviewer should be able to identify what the machine produced, what a person added, how the two were combined, which intermediate outputs drove the decision, and whether the realized result later supported the original reasoning. Auditability turns rapid AI construction and semantic expansion into capabilities that organizations can validate, govern, and improve.

Working Hypothesis 9: Risk Management Enables Safer, More Reliable Improvement

The hypothesis is that a documented, proportionate AI risk-management plan reduces avoidable harm, rework, and disruption while supporting useful KPI/PPI improvement. The relevant risks include those the business is trying to reduce and those introduced by the AI application, its data flows, and its operating process. Risk management begins when the use case and reference-model experiment are defined and continues through development, validation, deployment, monitoring, and retirement.

Security and data confidentiality

Security assessment should cover unauthorized access, exposed credentials, insecure AI-generated code or software dependencies, malicious instructions embedded in documents or prompts, data extraction, and actions beyond the system's authority. Controls can include least-privilege access, separate development and production environments, secure credential handling, dependency and code review, input and output checks, and explicit approval for consequential actions. Desktop processing and cloud processing both require a defined security boundary.

Data confidentiality requires a map of what information is collected, where it is processed, what is sent to a model or cloud service, and who can access it. Classify customer records, CRM narratives, employee information, commercial terms, and intellectual property before use. Minimize and, where appropriate, redact or de-identify data; confirm permission to use it; apply encryption and access controls; and review provider retention, training-use, residency, deletion, and contractual terms. Prompts, logs, intermediate outputs, and validation workbooks can themselves contain confidential information and require protection.

A plan with owners, controls, and decision gates

For any test using nonpublic data or influencing live decisions, the Forum's proposed operating requirement is a documented risk-management plan approved by the responsible organization. It should define the intended use, affected parties, risk tolerance, decision rights, and system/data boundaries. A maintained risk register should record each material risk, its likelihood and impact before controls, the treatment, owner and due date, evidence that the control works, residual risk, and who may accept that residual risk. The plan also needs

validation and release criteria, escalation triggers, review dates, incident response, fallback or rollback procedures, and an accountable operational owner.

The assessment should also address erroneous or fabricated outputs, bias and unequal effects, weak calibration, model drift, double-counted evidence, unsupported causal claims, automation bias, legal and intellectual-property obligations, supplier changes, integration failures, outages, and operating cost. Independent testing should cover consequential outputs and security controls. Changes to models, prompts, data sources, access rights, or automation authority should trigger proportionate review before release.

Testing risk management through the reference models

In OTD, examine exposure of customer commitments and the effect of an erroneous scheduling action; in Customer Retention, examine confidential CRM evidence, false alarms, and inappropriate outreach. In IUVO and PRIMMS, examine sensitive forward information, misleading confidence, and unauthorized reliance on advice. Participants can compare simulated failures, detection and recovery time, rework, and control burden before and after safeguards. Rare incidents and small samples limit claims of risk reduction. LinkedIn discussion should use synthetic or approved, sanitized evidence; confidential customer data, credentials, and identifiable decision records must stay out of group posts.

Working Hypothesis 10: Architecture and Human-Machine Roles Should Adapt to Evidence

An architecture is a testable design choice. The hypothesis is that fitting the configuration to the decision process and revisiting it as evidence changes will produce better results than assuming one stack or one allocation of roles is universally appropriate. Reference-model components may be retained, modified, decomposed, or replaced. Forum participants can bring their own LLMs, analytics, enterprise platforms, visualization tools, and proposed allocations of human work.

One reference configuration uses desktop analytical applications to supply approved outputs to Tableau Cloud for visualization and distribution; the desktop computation remains outside Tableau Cloud. Other implementations may centralize processing or use an organization’s existing infrastructure. Compare alternatives on KPI/PPI performance, confidentiality, security, latency, cost, supportability, and auditability. Record the version and decision boundary being tested, and revisit automation authority when calibration, operating conditions, technology, or risk changes.

The Forum Working Method: Reference Models, Tests, and Shared Learning

The Forum will organize its work around the reference models described in this paper s well as additional ones under development. They provide shared examples that participants can inspect and discuss: an incoming order and its promised date, a deteriorating customer relationship, a project-risk signal, or a prospective forecast adjustment. The purpose is participation in testing

the hypotheses. Participants may use the existing implementation, modify it, reproduce its logic on another platform, or propose a competing approach.

For each working session (occurring one hour every other week), select a reference model, the hypotheses under examination, and a specific KPI/PPI question. Present an examinable evidence packet: the model version, synthetic or approved data, baseline output, assumptions, alternatives, and the next proposed test. Participant experience, research, counterexamples, unsuccessful experiments, and alternative architectures are all useful contributions. Agreement with the starting model is not a condition of participation.

Between meetings, continue the work throughout the week in the LinkedIn group, Hybrid Cognition Forum. A discussion thread for each model and test should identify the hypothesis, pose the question, and link to the approved reference material. Members can comment on assumptions, report attempted replications, post sanitized outputs, propose changes, and identify evidence that would distinguish competing explanations. The next meeting can review these contributions, resolve open questions, and agree on the next experiment.

Machine strengths	Human strengths
Continuous sensing and consistency	Context and novelty
Large-scale pattern recognition	Forward information outside the data
Probability and uncertainty calculation	Causal interpretation and mechanism
Historical comparison and memory	Values, consequences and tradeoffs
Reproducible translation of evidence	Accountability and decision authority

Maintain a simple hypothesis register with a stable identifier, current wording, reference-model version, proposed test, measures, contributor, evidence, limitations, and status such as proposed, under test, supported within stated conditions, revised, or rejected. Preserve earlier versions and unsuccessful results. This gives the Forum a cumulative record of what was tested and what changed, while keeping the list open to new principles and domains.

The working method: a shared reference model, a specific performance question, an examinable test, and learning carried forward.

Contribute a case, challenge an assumption, test an alternative, or help interpret the evidence.

The useful question is whether a proposed configuration improves the KPI and its supporting PPIs, under what conditions, at what cost, and with what residual risk. Preserve the contributions of machines and people, the intermediate evidence, the action taken, and the resulting outcome so others can examine the claim. The Forum develops practical knowledge by making those comparisons visible and repeatable.

Join the discussion: LinkedIn Hybrid Cognition Forum

Note on the Convening Organization

Milestone Planning and Research, Inc. is a project-management and risk-management organization that helps organizations beginning their AI journey follow a structured path designed to save time and reduce implementation risk. That path links the business objective to reference designs, evidence assessment, risk

planning, validation, integration, and measured outcomes. Existing applications are optional starting points; the approach can accommodate the organization's own ideas, systems, and preferred specialists. Examples are available at Ratio-Weekly.com.