

HYBRID COGNITION FORUM · WORKING PAPER BRIEFING

Beyond Information Gain

Converting Data Science Concepts into Business Value

From AI Models to Decisions, Actions, and Measurable Business Value

Presented by John Aaron · Milestone Planning and Research, Inc.

A C-suite briefing drawing on the Hybrid Cognition Forum working papers

ratio-weekly.com

Ubiquitous AI is table stakes. The advantage is in decisions.

Most organizations start their AI journey in the same place. That start is worth making — but it isn't where the durable advantage comes from.

WHERE EVERY ORGANIZATION STARTS

- LLM copilots, chatbots, and off-the-shelf assistants — fast to deploy, genuinely useful.
- Low switching cost, minimal customization, available to any competitor within weeks.
- Real value — faster drafts, quicker answers, lower support cost — the automation gains this briefing turns to next.
- But an edge everyone can copy stops being an edge. Short-term advantage, quickly competed away.

WHERE THE ADVANTAGE ACCUMULATES

- Embedding AI into the business-process decisions that actually move a KPI or PPI.
- Requires the evidence, governance, and judgment architecture the rest of this briefing describes.
- Much harder to build — and much harder for a competitor to replicate quickly.
- Compounds over time: better decisions generate better outcome data, which improves the next decision.

This briefing is about the second kind of advantage — not the first.

AI Can Transform Information Into Durable Business Value Through Probabilistic Decision Making

1 Value has three channels

A changed decision, an automated action, or an earlier catch. A model that does none of the three is economically inert, however accurate it is.

2 The missing bridge

Decision-value theory asks a sharper question than accuracy: does new information change what you'd do, and does that change improve the outcome?

3 Autonomy is a dial, not a switch

Speed and labor economics push toward more autonomous AI. Risk — not enthusiasm — should determine how much authority is delegated at each step.

4 Amplification must be earned

“Human + AI” is not automatically better than the best human or the best machine alone. It is a testable, and sometimes false, claim.

Automation value vs. decisional risk

Automation earns its keep in more than one currency. This framework is narrower: it is about the specific risk that appears when AI participates in a consequential business decision.

AUTOMATION VALUE — ALWAYS WORTH PURSUING

- Faster service and shorter cycle times — a quicker, easier answer for the customer.
- Lower processing and labor cost per transaction.
- Higher customer satisfaction from consistency and speed alone.
- Applies whether or not a consequential decision is involved — this value is real and worth capturing on its own.

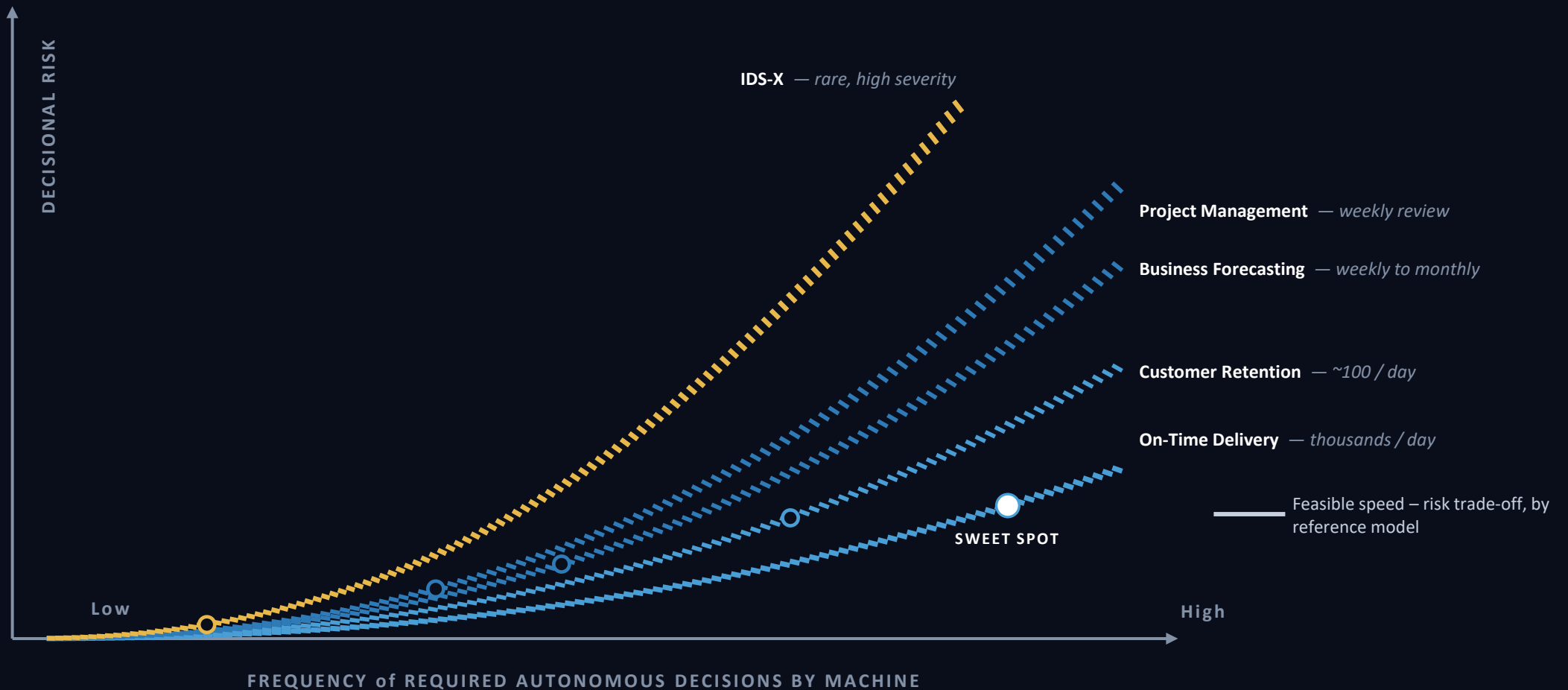
DECISIONAL RISK — WHERE THIS BRIEFING FOCUSES

- Arises specifically when AI participates in a consequential business decisions — some are routine, some are not.
- The same speed and labor savings that make automation attractive can be outweighed by the cost of a wrong decision.
- Governance, autonomy allocation, and audit trails earn their cost here — risk becomes the constraint.
- This is the risk the rest of this briefing is built to manage.

The practical test: does this task touch a decision whose outcome matters? If not, optimize for speed and cost. If so, read on.

The frontier depends on how often — and how severely — a decision fires

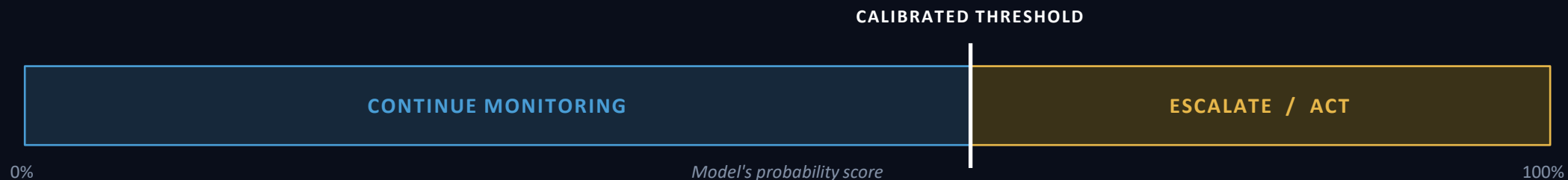
There is no single risk-speed curve. Volume and consequence per decision reshape it: high-frequency, low-severity processes can push much further before risk bites; rare, high-severity ones hit a rising risk price almost immediately.



The operating point moves with volume, severity, and governance strength. The dot marks OTD's sweet spot, where more speed stops paying for its added risk.

The on-ramp: collapse to binary decisions first

A probabilistic model outputs a continuous score. The frontier from the last slide only becomes usable once that score is collapsed into a bounded, testable yes/no.



WHY BINARY FIRST

- A continuous 0–100% score is hard to govern, audit, or test directly.
- A threshold turns it into a bounded yes/no — false positives and false negatives become directly measurable.
- Bounded and reversible is exactly the risk profile that justifies automating early on the frontier.
- It's the same starting point every reference model in this briefing uses.

HOW IT EARNS MORE AUTONOMY

- Once the threshold's error rates are known, it can be tuned — or a gray-zone escalation band added.
- Each earned increment moves the operating point further right on the frontier: more speed at the same or lower risk.
- The binary call can later expand into a ranked or multi-way decision — but only after the simple version is proven.
- This is the earned-autonomy path: Bounded Decision, then Decide & Act.

Binary first isn't a limitation — it's what makes a probabilistic model increasingly robust, testable, governable, and safely automatable in the first place.

Two different levers, two different tests

Governance and model improvement both change the risk-speed frontier — for any given reference model's curve — but they change different things, and they're validated differently.

MOVE ALONG THE FRONTIER — GOVERNANCE

- Doesn't change the model's error rate — the model is exactly as accurate as before.
- Changes what a given error costs: audit trails, quality gates, escalation, and rollback re-price the risk of a mistake.
- Effectively slides the organization to a different point on the same curve — more speed becomes affordable at the same risk tolerance.
- This is the “govern the dial” lever from the autonomy table.

SHIFT THE FRONTIER — MODEL & EVIDENCE

- Better evidence fusion, calibration, and narrower scope reduce the error rate itself — not just its cost.
- This is the cognitive-amplification claim: bending the whole curve outward, not sliding along it.
- More speed becomes available at the same risk, or the same speed becomes available at less risk.
- Requires the same prospective, cost- and risk-adjusted test as any amplification claim.

Governance is validated by audit and control effectiveness. Model improvement is validated by prospective KPI/PPI results.

Traditional Data Science Thinking Is Necessary But Not Sufficient. Information gain is not the enterprise objective

Two levels are often conflated when organizations evaluate an AI proposal.

THE ANALYTICAL SYSTEM

Information Gain

- A reduction in uncertainty — a cleaner classification, a better prediction, a more informative posterior.
- Central to decision trees, feature selection, experimental design, Bayesian updating.
- Can improve steadily while the operating decision never changes.



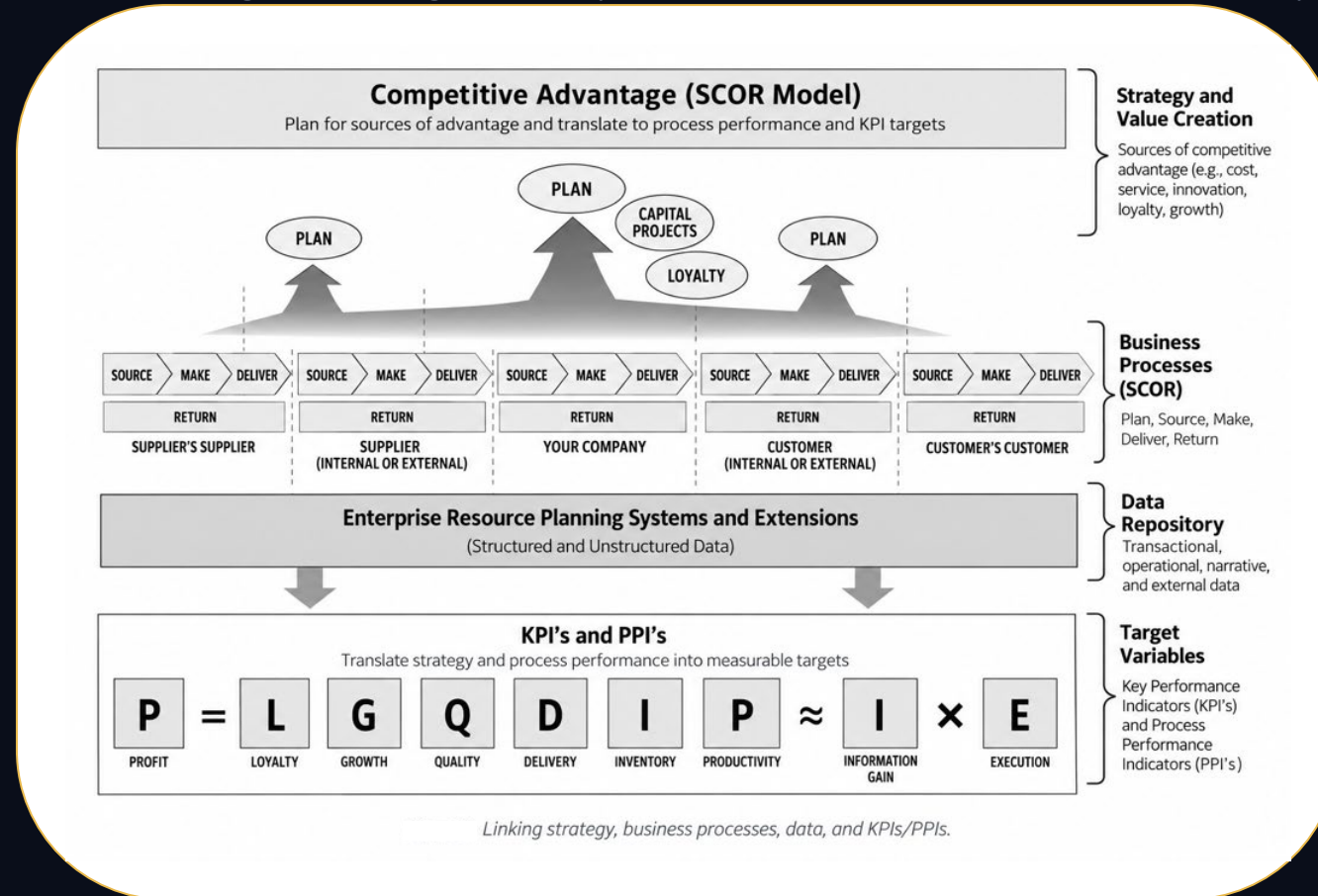
THE ENTERPRISE

Decision, Action, KPI/PPI

- Operates on accepted KPIs (On-Time Delivery, Retention, Delivery) and the PPIs that drive them.
- An insight is economically inert until someone has the authority, timing, and resources to act on it.
- The analytical objective must connect to the enterprise objective through a decision and an action.

Linking strategy, process, data, and KPIs/PPIs

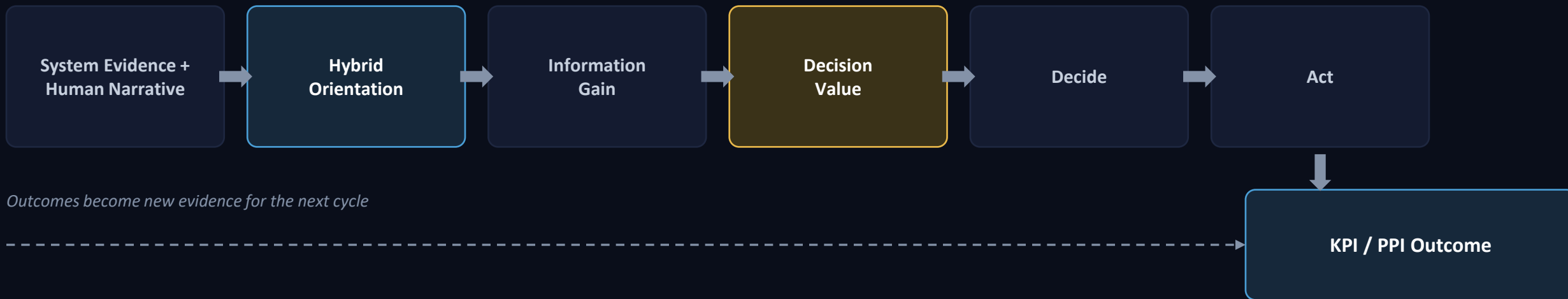
Adapted from the SCOR model: sources of advantage flow through business processes and the data environment into the accepted KPI/PPI targets an AI proposal must move.



Adapted from APICS Supply Chain Council (2017), Supply Chain Operations Reference (SCOR) Model, Version 12.0.

Information becomes value through a governed AI decision-action cycle

HOWARD'S BRIDGE



GOVERNANCE / RISK CONSTRAINTS

Authority · Permissions · Limits · Monitoring · Reversibility · Audit

The screening question for any AI proposal: which decision changes, who has authority to act, and which KPI or PPI should improve?

Three sources of value share this cycle — not one

Howard's test — does new information change the decision? — is necessary, but on its own it undersells this cycle. Two more value channels run through the same evidence.

DECISION VALUE

New information changes the response — Howard's test.

AUTOMATION VALUE

A stable, well-understood decision — like order routing — executed at machine speed and scale. Valuable even when the decision itself doesn't change.

COMPOUNDING QUALITY GAINS

Earlier intervention, fewer errors, and full automation reinforce each other. The highest-value zone is the overlap — automate the routine, decide well on the exceptions, and catch the anomaly before it becomes rework.

EARLY-DETECTION VALUE — catching the anomaly sooner avoids downstream rework, whether or not the final decision changes.

“Human + AI is better” is a goal to test — not a default to assume

Human-machine interaction is a configuration. Decisional accuracy is a performance effect — and it has to be demonstrated, not assumed.

COMPLEMENTARITY GAIN

Hybrid Performance – $\max (\text{Human-Only} , \text{Machine-Only})$

A positive, repeatable, cost- and risk-adjusted difference is evidence of amplification — within the conditions tested.

Errors must differ

Complementarity depends on humans and machines failing in different ways — not the same way, twice.

Reliability must be knowable

The system should recognize which contributor is more trustworthy in a given case.

Verification must be affordable

People must be able to check a recommendation without prohibitive time or cost.

Feedback must close the loop

Outcomes need to come back as evidence, or neither the human nor the machine learns.

The evidence supports optimism — and caution, in equal measure

WHERE AI ASSISTANCE HELPED

+40%

faster professional writing tasks with generative-AI assistance, and 18% higher rated quality (Noy & Zhang, 2023).

+15%

average productivity increase across 5,172 customer-support agents after an AI assistant was introduced — largest gains among less-experienced workers (Brynjolfsson, Li & Raymond, 2025).

WHERE THE EVIDENCE URGES CAUTION

A meta-analysis of 106 experiments and 370 effect sizes found human-AI teams beat humans alone on average — but underperformed the better of human-only or AI-only on decision tasks specifically, even as content-creation tasks gained (Vaccaro, Almaatouq & Malone, 2024).

Explanations attached to AI recommendations can help speed or accuracy in some settings — but do not reliably reduce automation bias, and sometimes increase it (Vered et al., 2023).

Takeaway: complementarity is configurable, not universal — which is exactly why it must be tested case by case.

The same speed that removes delay can increase loss exposure

THE TRADE-OFF

Agentic AI moves systems from producing advice to participating directly in execution. An error that once stayed inside a recommendation can now be executed at machine speed — propagated across systems before a person ever reviews it.

A 2026 study of 230 professionals found agentic configurations improved productivity — but a medium-autonomy design requiring confirmation best balanced productivity with users' need for control (Geninatti Cossatin et al., 2026).

GOVERN THE DIAL, NOT JUST THE MODEL

- ✦ Repeatable, low-consequence, reversible actions can justify broad delegation
- ✦ Novel, weakly-observed, or irreversible actions require pre-approval and escalation
- ✦ Bounded tools/data access, transaction limits, and quality gates
- ✦ Continuous monitoring, stop conditions, and rollback paths
- ✦ One accountable owner for every delegated decision

The question isn't how much labor AI can remove — it's which allocation of human and machine authority produces the best risk-adjusted result.

Autonomy is an allocation across the cycle — not an on/off switch

Every reference model enters at Orient & Expand or Recommend, resolving a simple binary anomaly check. Bounded Decision and Decide & Act are earned later, as calibration improves.

Orient & Expand

MACHINE

Machine integrates evidence and develops causes and alternatives.

HUMAN

Human interprets, challenges, decides, and acts. PRIMMS is an example.

PRINCIPAL TEST

Does the synthesis improve search, judgment, timing, or outcomes?

Recommend

MACHINE

Machine ranks or proposes a response.

HUMAN

Human verifies, accepts or rejects, authorizes, and acts.

PRINCIPAL TEST

Can the recommendation be checked at reasonable cost, without automation bias?

Bounded Decision

MACHINE

Machine selects among permitted alternatives.

HUMAN

Human sets policy and limits; handles exceptions or approves execution.

PRINCIPAL TEST

Are thresholds, escalation, accountability, and residual risk explicit?

Decide & Act

MACHINE

Machine selects the response and changes process state.

HUMAN

Human designs constraints, monitors, intervenes, and stays accountable.

PRINCIPAL TEST

Are permissions, limits, detection, and rollback proportionate to consequence?

Ten working hypotheses for hybrid cognition

From the Hybrid Cognition Forum working papers “Beyond Human in the Loop” and “Hybrid Cognition in Business” — propositions to test, not settled conclusions.

- 1 Organize around the KPI and its supporting PPIs** — Start with an accountable business measure, not a technology.
- 2 Departures from an expected trajectory provide useful warning** — Deviation from the expected path warns earlier than the level of the metric itself.
- 3 Narrative can add decision-relevant evidence** — CRM notes, status reports, and commentary carry signal transactional fields miss.
- 4 Monitoring changing relationships preserves usefulness** — A model that worked yesterday can quietly lose power as the environment shifts.
- 5 Preserving alternatives improves decisions under uncertainty** — Keep competing explanations and paths visible instead of collapsing too soon.

Ten working hypotheses, continued

6

Explicit human contributions can add measurable value — Record what the human added, separately from the machine prior, so both become testable.

7

Human-machine interaction can produce cognitive amplification — Combined contribution can exceed either alone — but only as a tested effect, not an assumption.

8

Semantic expansion improves the search for causes and actions — LLMs broaden the hypothesis space beyond a single team's memory.

9

Auditable outputs make AI-assisted work more dependable — Every consequential recommendation needs a reproducible, challengeable trail.

10

Architecture and human-machine roles should adapt to evidence — No stack or authority split is universal; revisit it as calibration and conditions change.

Each hypothesis is tested against a reference model — On-Time Delivery, Customer Retention, IUVO, PRIMMS®, and IDS-X — not argued in the abstract.

One common logic, five different KPIs

Every organization starts the same way: a single binary anomaly check. Richer, multi-path judgment is a later, earned step — not the entry price.

IUVO™ Forecasting System

Short-Sale Identification

S & P Case Study: Market conditions and signal changes that may precede a short-sale opportunity

STARTING BINARY DECISION

Candidate present: yes or no?

On-Time Delivery

Commitment Reliability

Backlog, capacity, material readiness, and routing, read against customer urgency

STARTING BINARY DECISION

Requested date supportable: yes or no?

Customer Retention

Churn Risk

Usage, service activity, sentiment, and field narrative

STARTING BINARY DECISION

Intervention warranted: yes or no?

PRIMMS-PM®

Project Risk

Schedule, cost, Risk Register, and Voice-of-the-Team narrative

STARTING BINARY DECISION

Intervention required now: yes or no?

IDS-X

Intrusion Detection

ML classification, signature alerts, and behavioral anomaly scores fused per network flow

STARTING BINARY DECISION

Escalate for analyst review: yes or no?

A governed path can include simulation—a lower risk approach



What migrates is the validated operating design — decision rights, thresholds, escalation policy, and audit requirements — not necessarily the reference model's original code.

PRIMMS® occupies a distinct position: a reference model for hybrid orientation and recommendation, and a mature specialized project-risk application that may simply remain the operational solution.

Simulation is an early verification step: use synthetic data to challenge assumptions before live testing. What ultimately migrates is the validated operating design — decision rights, thresholds, escalation policy, reward logic, and audit requirements — not necessarily the reference model's original code.

Five Reference Models Currently Available ---Local Calculation, Governed Communication

The current reference implementation gives MATLAB and Tableau Server distinct roles. It's a strategy-development reference, not a required production stack.

MATLAB — LOCAL CALCULATION ENGINE

- Ingests structured operational data, documents, sensor streams, and approved local narrative; applies statistical and temporal analysis.
- Compiled, and runs on MATLAB Runtime — no per-user MATLAB license needed to operate the application.
- Local execution creates a controlled boundary — sensitive source data can be filtered, minimized, or summarized before anything reaches an LLM or web-facing service.
- Preserves the machine prior, human addition, revised assessment, decision, and outcome for audit.

TABLEAU SERVER — COMMUNICATION LAYER

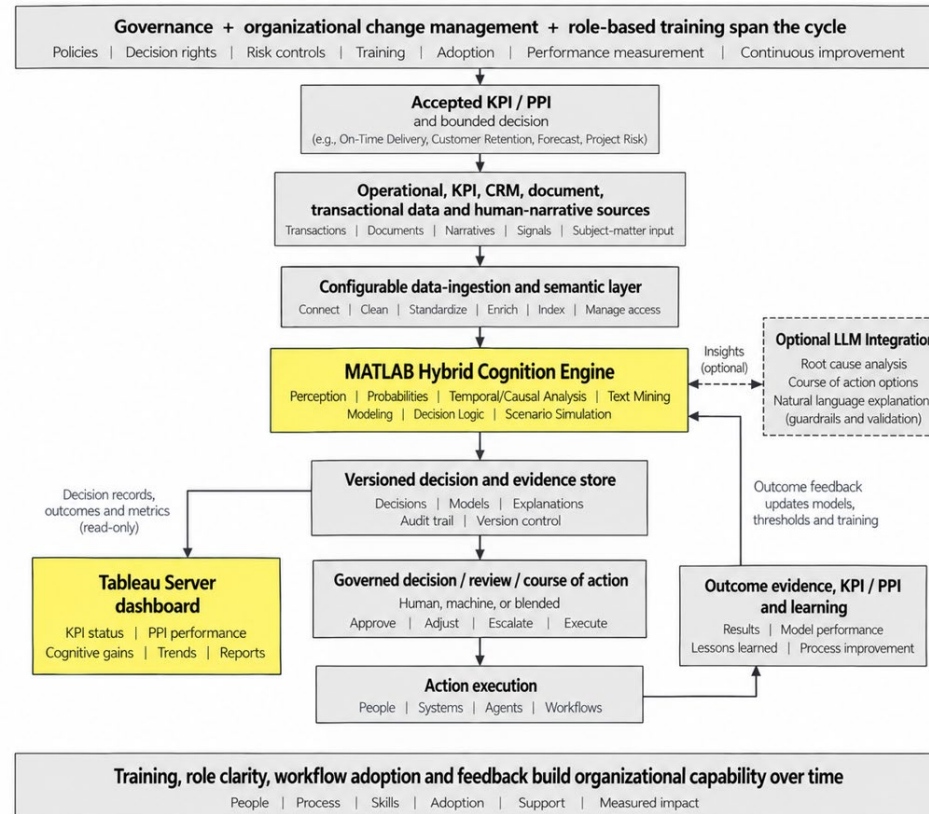
- Publishes only approved outputs — trends, classifications, summaries, aggregates — never raw source documents.
- Presents KPI/PPI trajectories, warning conditions, probability ranges, and action status through role-appropriate browser views.
- Gives managers, analysts, and sponsors shared visual context — and a retained history of how the decision evolved.
- A communication layer, not just a display — where interpretations get compared and reconciled.

An organization may later substitute other statistical environments, platforms, or visualization tools — the reference exposes the functions any architecture needs: sensing, calculation, decision authority, action, communication, audit, and feedback.

Every function, one governed cycle

Configurable reference-model architecture: a practical starting point

MATLAB executes the cognitive sequence; Tableau Server displays KPI status and cumulative performance



The architecture is a strategy-development reference: validated operating design can later migrate to the client's preferred enterprise stack.

Training and adoption are part of the decision system

A technically sound configuration can still fail if the organization doesn't change with it. OCM is a design variable — not a deployment activity appended after the model is built.

WHAT OCM DETERMINES

- Who sees which evidence, and when exceptions get escalated to a human decision-maker.
- How probabilities and thresholds get interpreted in practice, day to day, by the people using them.
- Where decision authority actually sits — versus where it's documented to sit.
- What work is genuinely automated or retained, role by role.

THE SECOND LEARNING LOOP

- Outcome evidence updates models, thresholds, and decision rules.
- Usage experience updates human skill, role design, workflow, and readiness.
- As organizational capability grows, configurations once too risky can become feasible.
- Low readiness can erase a superior model's value — through non-use, misuse, delay, or overreliance.

Adoption, proficiency, exception quality, and time-to-competence are PPIs in their own right — measured and learned from alongside the technical metrics.

Build AI strategy backward from the KPI, not forward from the tool

1 Select an accepted KPI and the PPIs behind it

2 Map the decision network — who senses, decides, authorizes, acts

3 Separate structured evidence, human narrative, machine inference, and human judgment

4 Define the pattern to detect and the initial binary decision it informs

5 Assign human/machine roles at the perception, decision, and action boundaries

6 Set acceptable false alarms, risk constraints, and escalation paths before deployment

7 Define prospective comparisons against human-only or machine-only baselines

8 Select the architecture that fits governance and confidentiality needs — last, not first

The governing equations describe the logic of the system. **Start safely: synthetic RL simulation tests the assumptions and reward structure first; prospective real-world validation comes before deployment.**

THE RL GOVERNING LOGIC — IN PLAIN ENGLISH

**OBSERVE CURRENT STATE → CHOOSE A PERMITTED ACTION → MEASURE BUSINESS
RESULT → LEARN**

Math shorthand: $s_t = \Omega_t \rightarrow a_t = (D_t, \Gamma_t, A_t) \rightarrow r_t = \Delta CA_t \mid \pi^* = \text{the best learned decision rule under the tested constraints}$

1 State s_t

What the system knows now: evidence, probabilities, discordance, KPI/PPI status, risk and context.

2 Action a_t

What it is allowed to do: decide, escalate, seek more evidence, involve a human, or act within governance limits.

3 Reward r_t

What happened after the choice: improvement or loss in KPI/PPI, cost, time, quality and risk-adjusted value.

4 Policy π^*

The decision rule learned from repeated outcomes: which permitted action tends to work best in each situation.

The enduring architecture is the decide-act-learn cycle tied to an accepted business measure.

That is the point at which information becomes valuable — and where the value of hybrid cognition can be demonstrated .

JOIN THE HYBRID COGNITION FORUM

One hour, every other Wednesday, 12:00 p.m. Central — beginning September 23, 2026

Bring a case, a reference model, a counterexample, or a question worth testing.

ratio-weekly.com