

PROCUREMENT MISSION CONTROL

ONE EXCEPTION. SIX POINTS OF VIEW.



TECHNICAL REPORT / SYSTEM 006

Procurement Mission Control

A transparent multi-agent decision environment built around a deterministic procurement rules engine.

Fatima Aguilar · Decision Systems Lab · July 2026

Live prototype: fa366193.github.io/procurement-mission-control/

Source: github.com/fa366193/procurement-mission-control

Abstract

The research question is not whether agents can automate procurement.

The Procurement Heat Map Engine is a shipped operational system that transformed fragmented SAP exports, manual workbook preparation, and implicit purchasing logic into a daily, explainable exception queue. Its deterministic backbone reduced manual processing by more than 80 percent while preserving human authority.

Procurement Mission Control is an experimental extension of that system. Six bounded analytical roles inspect one synthetic exception from different perspectives: demand, supplier reliability, inventory policy, finance, adversarial risk, and decision synthesis. A visible trace shows what each role contributes, which rules fire, where evidence conflicts, what remains unknown, and how a recommendation changes under counterfactual inputs.

The prototype does not connect to SAP, call external AI services, learn from overrides, or place purchase orders. Its contribution is an interaction and evaluation architecture for studying whether specialized, contestable agent review improves procurement decisions without weakening auditability.

Research question

Can a hybrid rules-and-agents interface improve evidence coverage, calibrated judgment, and time-to-decision while preserving human accountability?

Artifact	Purpose	Status
Production heat map	Daily deterministic exception generation	Shipped
Mission Control	Interactive multi-agent decision review	Research prototype
ERP write-back	Execute approved purchasing action	Not implemented
Controlled evaluation	Compare hybrid system to baseline	Proposed

Development notebook

Five versions of the same underlying problem

V1 · Legacy workbook

Buyers exported ME2N, MB52, and ZMINVENTORY, updated quarterly inputs, refreshed a 24-sheet workbook, and manually interpreted thousands of materials.

V2 · Automated orchestration

PowerShell scheduled extraction, validation, cleansing, logging, workbook execution, publishing, and daily notification.

V3 · Deterministic heat map

Trusted Excel domain logic was restructured into explicit Pull In, Create PO, Push Out, Monitor, Research, and Obsolete classifications.

V4 · Decision surface

Power BI separated executive monitoring from the ranked buyer exception queue and rule-level drill-through.

V5 · Agentic review

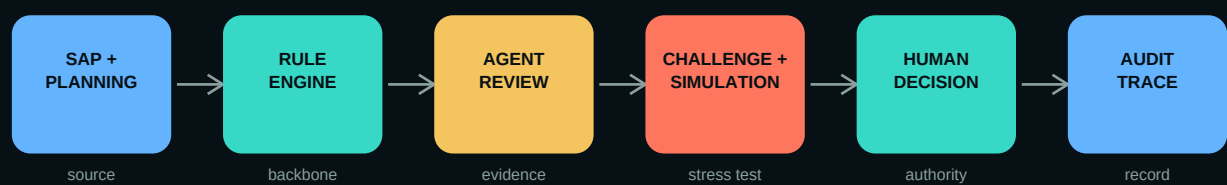
The research prototype surrounds the rules engine with bounded specialists, mandatory challenge, scenario simulation, and a human disposition record.

Design principle

The sequence matters. Agents were added after a transparent production backbone existed. The prototype does not ask users to trust a novel model instead of established rules; it asks whether additional perspectives can make exceptional cases easier to investigate.

Architecture

Rules remain the backbone. Agents make the review legible.



The browser prototype centralizes all calculations in typed deterministic functions. Each role reads the same scenario but is assigned a bounded analytical lens. The Risk Challenger is structurally required before synthesis, making disagreement part of the architecture rather than an optional prompt.

Six bounded roles

Role	Bounded responsibility
Demand Sentinel	Consumption, volatility, forecast bias, stockout horizon
Supplier Scout	Lead time, on-time delivery, open PO exposure, concentration
Inventory Strategist	Reorder point, safety stock, inspection inventory, excess and obsolescence
Finance Guardian	Working capital, expedite premium, carrying cost
Risk Challenger	Missing evidence, counterarguments, failure modes, confidence limits
Decision Orchestrator	Rules, findings, disagreement, and human-review memo

The interface physically arranges the six roles around a shared evidence board. Activity bubbles sit above each avatar; stage transitions change the active work, evidence, consensus state, rules, and trace. This makes orchestration observable without implying that animation is evidence.

Decision model

Deterministic formulas before narrative synthesis

The prototype uses synthetic but operationally plausible inputs: on-hand quantity, open purchase order quantity, weekly demand, demand volatility, lead time, supplier on-time delivery, unit cost, safety stock, inspection inventory, lifecycle, criticality, and supplier concentration.

Core formulas

Measure	Definition
Usable stock	on hand + 0.55 x inspection inventory
Weeks of supply	usable stock / weekly demand
Reorder point	weekly demand x lead time + safety stock
Reorder gap	max(0, reorder point - usable stock - open PO)
Supplier risk	weighted reliability, concentration, and lead-time factors; capped at 100
Carrying cost	excess units x unit cost x 22 percent annual factor
Expedite premium	reorder gap x unit cost x 18 percent scenario factor

Classification order

Obsolete lifecycle -> OBSOLETE. High volatility with supplier risk -> RESEARCH. Positive gap without open PO -> CREATE PO. Positive gap with an open PO -> PULL IN. Material excess -> PUSH OUT. Otherwise -> MONITOR.

Confidence is displayed as a band rather than false precision. It begins with record completeness and is reduced by conflicting signals including high volatility, inspection stock uncertainty, phase-out status, and supplier risk.

Interaction model

A decision review, not a chatbot

Visitors begin in an executive portfolio containing twelve synthetic SKUs. Filters narrow action, site, and criticality. Selecting a record loads its planning inputs into the mission room.

Six-stage investigation

Stage	Visible work
01 · Intake	Validate completeness and units
02 · Parallel scan	Five specialists inspect their domains
03 · Challenge	Contest the leading action and expose missing evidence
04 · Simulation	Stress-test demand, delay, and cash
05 · Synthesis	Reconcile rules, evidence, and disagreement
06 · Human review	Approve, override, or request more evidence

The outcome contains the recommended action, calibrated confidence, reasons, strongest counterargument, unknowns, trigger evidence, quantity and cash implications, owner, and next action. The Full Trace drawer reconstructs every stage and fired rule.

Human disposition

Approve, Override, and Request More Evidence are functional local controls. Overrides require rationale and are described only as possible signals for future rule review. The downloadable JSON memo records the scenario, analysis, method version, timestamp, and human disposition. No transaction is written to an ERP.

Counterfactual lab

The useful question is: what would change the decision?

Four controls vary weekly demand, lead time, supplier reliability, and unit cost. The system immediately recomputes coverage, supplier and operational risk, cash exposure, and classification. Baseline and simulated results are shown together.

Why counterfactuals matter

A recommendation is more useful when a buyer understands its decision boundary. If a small change in lead time reverses Pull In to Monitor, the action is fragile. If a wide range of assumptions preserves the classification, the action is more robust. This is more actionable than a single confidence number.

Example: sterile process valve

Low usable stock, a positive reorder gap, critical material status, weak supplier on-time delivery, and high concentration support accelerating an existing order. The Challenger notes that normalizing demand could create avoidable inventory. The what-if lab lets the reviewer test that claim directly.

Failure-aware presentation

Unknown quality-release timing, alternate supplier capacity, and causes of demand variance are displayed explicitly. The system names the evidence that could change the action rather than hiding uncertainty inside a score.

Evaluation protocol

An impressive prototype becomes research only when it can fail.

The proposed study would compare Procurement Mission Control against the shipped rules-only interface using historical or controlled synthetic cases reviewed by experienced buyers. Cases should be held out from interface tuning and include adjudicated target actions plus acceptable alternatives.

Primary metrics

Metric	Operational definition
Time-to-decision	Elapsed time from exception opening to recorded disposition.
Decision consistency	Agreement across reviewers facing the same evidence.
Evidence coverage	Fraction of case-critical signals acknowledged before action.
Calibration	Relationship between confidence bands and adjudicated correctness.
Override quality	Whether rationale identifies valid context absent from the record.
Business outcomes	Stockout avoidance, excess reduction, expedite cost, and working-capital effects.
Comprehension and trust	Ability to explain why the action was recommended and when it should change.

Ablation tests

Rules-only baseline: isolates the value of visible specialization. **No Challenger:** tests whether mandatory dissent improves calibration. **No Finance Guardian:** tests whether cash tradeoffs alter dispositions. **Hidden vs. visible trace:** tests whether inspectability improves comprehension or merely increases time.

Governance and limitations

The prototype stops exactly where production authority should begin.

Synthetic data

The demonstration does not establish business effectiveness or integration feasibility.

No autonomous purchase orders

Every action remains a recommendation. ERP write-back would be gated behind authenticated human approval.

Security

Production use requires access controls, segregation of duties, data minimization, prompt and data-injection defenses, and supplier-sensitive information handling.

Traceability

Every future model, prompt, rule set, evidence source, tool call, and human disposition should be versioned and logged.

Agent failure

Specialized roles can still share correlated errors, hallucinate, omit evidence, or create persuasive but unsupported narratives.

Learning from overrides

Overrides are not automatically correct. They should trigger structured review, not silent retraining.

Research boundary

The objective is leverage, not autonomy. The human buyer owns the decision; the system owns the responsibility to make its reasoning inspectable.

Conclusion

From exception automation to accountable decision review

The production Procurement Heat Map Engine proved that explicit business logic, automated preparation, and attention-aware design could materially improve a purchasing workflow. Procurement Mission Control does not replace that achievement with an agent demo. It builds on it.

The hybrid design treats deterministic rules as trusted infrastructure and specialized agents as a contestable review layer. Evidence is shared, disagreement is visible, counterfactuals are interactive, and the final action remains human. This creates a credible research platform for asking whether agentic systems can improve operational judgment without obscuring responsibility.

Public artifacts

Live prototype

<https://fa366193.github.io/procurement-mission-control/>

Source code

<https://github.com/fa366193/procurement-mission-control>

Research notebook

<https://decisionsystemslab.org/systems/procurement-heat-map>