SpecOps Decision Intelligence

Governed Analytics for the Enterprise PMO

Ron Griffin

September 2026

Executive Summary

Enterprise PMOs are not short of data. They are short of decisions made on time, with the evidence, the options, the owner and the approval in one place, and a record that survives the meeting. Every system of record in a large delivery portfolio captures its own slice of operating reality. None of them tells a VP which four decisions this week protect the most value, what each option costs, and who has to sign.

SpecOps Decision Intelligence closes that gap. It is the analytics and decision layer of the SpecOps.AI platform, an AI-powered project management and governance platform built around three agents, Sherpa, Overwatch and Orchestrator, designed to collect, stream and govern the operating data a portfolio produces. Sherpa is live today; Overwatch and Orchestrator are in beta and limited release, detailed in section 11. Decision Intelligence sits on top of the tools a PMO already runs. It does not replace the PPM tool, the ERP, the resource plan or the GRC system. It reads from them through their sanctioned paths, modernizes and processes what they hold into one decision dataset, and turns it into decision insight, real-time reporting, governed analytics, and a durable record of who decided what on which evidence, the design this paper lays out and section 11 checks against what is actually running.

Five things make it different from a dashboard product:

1. **It is not standalone.** The data behind every card is collected and streamed by the platform the PMO is already operating on, not re-keyed into a reporting tool.
2. **Every figure carries its tier of truth.** Fact, calculation or estimate, with freshness and an interval, visible on the card.
3. **Analytics are governed.** Every card carries a trust score that stays quiet while it is healthy and speaks plainly when it is not, in the order that finds the mundane cause before the mysterious one.
4. **Humans stay accountable.** Nothing changes a committed milestone, a budget, a staffing assignment or a vendor relationship without a named approver, and every decision is recorded with the options that were on the table.
5. **It learns from the people using it.** Corrections and outcomes feed calibration, under approval, and the system never claims a human's insight as its own.

The product is measured by decisions improved, not dashboards viewed.

1. The Problem as a PMO Leader Lives It

The operating review is the most expensive meeting in the portfolio. A deck that took a week to assemble is walked in ninety minutes. Forty programs are reviewed at equal weight. The one decision that mattered is made afterward in the hallway, on a number nobody can trace, and the record of it is an email.

Four findings shaped what SpecOps built.

Status is not a decision. A red project, an overloaded team, a twelve percent schedule variance: these are symptoms. Leaders act on consequence. A two-week validation delay puts a committed date at risk and exposes a specific amount of expected value. Every surface in the product translates a symptom into a consequence, an owner and a due date.

Dashboards dilute attention. A wall of equal-weight metric cards asks the executive to do the triage the product should have done. The first thing a leader sees must carry one operating story: the decisions that need action and the two or three conditions driving them.

Trust comes before autonomy. In a regulated enterprise, a recommendation is only useful if it can be inspected. The product separates observed fact from calculation from forecast from assumption from recommendation, and shows freshness, confidence and limitations next to the number, not in a tooltip.

The record is the product. A recommendation that is approved, overridden or ignored without a durable record has no governance value and no learning value. The decision record is the asset that compounds over time.

2. The Foundation: a Platform That Already Holds the Data

Most analytics products start with an integration project. Decision Intelligence starts with a platform that is already connected, because it is built into SpecOps.AI, which a PMO adopts to run delivery and governance, not only to report on it.

Sherpa is the platform's AI command authority. It reads the portfolio, classifies what a person asks for, assembles the evidence, drafts options, and acts only under approval. In the analytics layer, Sherpa is the assistant that explains a decision from evidence and routes a plain-English request to the right analytic method.

Overwatch is continuous anomaly detection. It compares three views of the same work, what was planned, what is actually happening, and what the governance baseline requires, and flags divergence as it happens. In the analytics layer, Overwatch is the streaming signal: a card does not wait for a monthly cycle to learn that a plan and its reality have separated.

Orchestrator is tiered governance. It enforces which actions require which approvals, brokers every AI action through a governed proxy, and writes the provenance record. In the analytics layer, Orchestrator is why a recommendation can be filed as an action with a named approver, and why every AI-assisted step is on the ledger.

Sherpa is live today. Overwatch is in private beta with federal partners. Orchestrator is in limited release. The description above is the design each is built toward; section 11 states plainly where each one stands.

Underneath all three sits the data pipeline the platform is building toward: ingest from the systems of record through their sanctioned paths, for example a PPM data warehouse's read interface, the engineering tools where work actually happens, and the finance, resource, vendor and compliance sources a portfolio already maintains, modernize that into one model, the seven-domain decision dataset, and keep it current. The systems of record remain authoritative throughout; the platform reads, reconciles and governs, it does not write back without confirmation.

This is the strength worth hearing first: the analytics are only as good as the data behind them, and here the data is collected and streamed by the platform doing the work, not assembled by hand for the report.

3. The Decision Dataset: Seven Domains, Three Tiers of Truth

Every analysis in the product reads from one dataset organized the way a PMO actually operates:

Domain	What it holds
Projects	Identity and classification, scope and delivery, status and health, dependencies, benefits
People and Teams	Reporting lines, allocation and demand, skills and readiness, performance signals
Financials	Budget, actuals, forecasts, capitalization and depreciation, benefits realization, the corporate finance layer (present value, net present value, internal rate of return, payback, cost of delay)
Supply Chain and Logistics	Vendors and contracts, orders and lead times, third-party risk, delivery performance, external capacity
Time	Periods, baselines, trends, seasonality, leading indicators
Risk	Registers, exposure, correlation, anomalies, trend breaks
Technology	Platform health, integrations, AI and automation inventory, governed actions

Every figure in the dataset is one of three things, and the tier is visible wherever the figure appears:

- **A fact**, read from a system of record, with its source and as-of time.
- **A calculation**, derived from facts by a stated method, inheriting the freshness of its oldest input.
- **An estimate**, a forecast with an interval, a model version, and a calibration record.

Deterministic and probabilistic are therefore properties of the data, not marketing labels. A card over facts is a calculation. A card with an estimate in it is a hybrid and carries the interval. A financial exposure card and a delivery risk card agree because they read the same rows.

4. Decision Insight: From a Worry to a Card

The customer-facing workspace is Sherpa Analytics. Its first page leads with Decisions Needed: the ranked list of decisions requiring action, each with severity, consequence, owner, due date and the recommended path. Around it sit the conditions driving portfolio exposure, the health of the portfolio, and roll-ups by team and by person.

Reports are role-layered and triage up. A first-line manager sees their team's projects, load and blockers. A program leader sees the roll-up of their managers' teams and the exceptions. The executive sees the portfolio, the risks, the decisions needed, and only what triaged up. Each level's view feeds the level above it; nothing is re-pulled or re-stitched.

The Decision Builder is where a PMO shapes its own analytics without a report designer. A person describes what they are worried about in plain English, for example "scope creep, logistics and financial exposure on Wave 9", and the builder turns that into a set of analysis methods drawn from a catalog of forty-nine, organized by the seven domains. The first concern named leads the card; the rest form its drill-down. They can add or remove methods from the domain tree, promote another method to the headline, split one set of methods into independent cards, and add the finished card to the dashboard for everyone in the organization. Nothing is a wizard. Every method declares how it reasons (deterministic, probabilistic or hybrid) and what data it needs, and where a method's source is not yet connected, the card says so in place of a number. A card never guesses.

5. Real-Time Reporting

A monthly operating review is a snapshot. Decision Intelligence is designed to be a stream: connectors keep the decision dataset current on the cadence of each source, Overwatch's divergence detection fires when plan and reality separate, and every card carries the as-of time of its oldest input so a leader knows how fresh the number is before deciding on it. The ingestion pipeline behind this is not yet live; section 11 states where it stands.

Real time does not mean noisy. Attention is earned by the score, not demanded up front. The workspace stays calm, the decisions that need action rise to the top, and the trust indicator described next is what tells a leader when a number has quietly stopped being reliable.

6. Analytics Governance: the Trust Score

This is the part of the product built specifically for a governance-first enterprise, and it applies to every card the platform shows, in every domain.

Every card carries a trust score from 0 to 10, computed from what the system can see about the card's inputs and, for a forecast, how its recent predictions fared against outcomes. The display rule is quiet until low: a small passive indicator at 8 to 10, the same indicator in neutral at 5 to 7, and a visibly more prominent flag at 0 to 4 that invites a click. It never interrupts. It is never a popup or a blocking modal.

Clicking the indicator opens an explanation in plain English, never a data-science readout. The diagnosis runs in a strict order that encodes a product decision: most low scores come from mundane causes, so the mundane causes are checked first.

1. **Staleness.** Has a required input gone past its refresh threshold? If so, the explanation is stated with full confidence: "This is low because the engagement score has not been updated in 34 days."
2. **Completeness.** Are required fields missing or too brief to rely on? Stated with the same confidence.
3. **Drift.** For a forecast whose inputs are current and complete but whose predictions keep missing, the explanation is bounded and honest: it states the observable pattern, names a correlated input only when the system can identify one, and says plainly that the system cannot tell you why. It never fabricates a cause.
4. **Fallback.** A low score with no identified cause says so and asks for a manual review.

Every explanation ends in two choices, and neither is wrong: investigate and recalibrate, which lets the person correct the field or leave a note describing what the system could not see; or proceed knowingly, which accepts the uncertainty and moves forward, recorded as an acknowledged risk rather than a resolved one.

Thresholds, drift sensitivity and the score bands are configuration, tuned per domain and per deployment, because a federal contractor's compliance cadence is not a mid-market PMO's reporting cadence. Every score is written to an append-only scoring ledger with the inputs it was computed from and the layer that fired, so a stored score can be read back against the rules that produced it.

The governance posture follows the NIST AI Risk Management Framework pattern: govern (accountable owners, policies, approval thresholds), map (use case, stakeholders, data classification), measure (calibration, coverage, override rate, groundedness), manage (routing, human approval, rollback).

7. Human-in-the-Loop Accountability

The assistant explains, assembles evidence, drafts options and routes actions. It does not decide.

The design is this: any option on a decision can be filed as a tracked action with the chosen option and any override of the recommendation recorded. High-impact actions route to a named approver role and nothing is applied until a person executes it. The decision record is meant to capture the options shown, the option chosen, who decided, the approver, the evidence basis and the outcome, the thing an auditor asks for first. That record is not yet built; section 11 states where it stands.

The trust score's two choices are part of the same record. When a person proceeds knowingly on a low-trust number, that is logged with the score they saw, the explanation they read and their role. When they investigate and leave a note, that note is kept with the card. The system will never later present that person's insight as its own reasoning; a future explanation influenced by a correction attributes it to the person who supplied it.

8. Self-Learning, Under Governance

A governed system learns without becoming ungoverned.

Outcomes close the loop. The decision record is the label source. When a forecast's date arrives, the platform scores the claim automatically against what happened. That produces per-method calibration and coverage: when a card says 80 percent, is it right about 80 percent of the time, and do its intervals contain the outcome at the stated rate.

Corrections feed calibration. Notes left through the trust score become labeled data points for tuning drift sensitivity. Field corrections resolve on the next scoring cycle.

Improvement is approved, not silent. A new method version runs as a challenger beside the incumbent, scores against the same outcomes, and is promoted only when its calibration and lead time beat the incumbent over a stated number of outcomes. Promotion is a versioned, approved change with rollback. Learned adjustments carry a confidence that decays, so a stale lesson fades instead of persisting forever.

Degrade honestly. When drift crosses a threshold, a card downgrades its tier, widens its interval or shows "not available", routes to its owner and logs why. A card that keeps showing a confident number through drift is the failure the design prevents.

The benchmarks that matter to a PMO are the ones it already cares about, and they are what the platform reports on itself: calibration and coverage per method, lead time ahead of the standard reporting cycle, freshness and coverage of inputs, override and acceptance rates with reasons, decision lift against the baseline for that decision type, and the share of material claims in a brief linked to evidence.

9. What a PMO Gets in the First 90 Days

SpecOps does not propose a platform rollout. It proposes one decision.

Discovery. A working session with PMO leadership produces a decision inventory: the ten to twenty recurring decisions the organization makes, scored for frequency, materiality, pain, data readiness and pilot feasibility. Typical entries: reallocating scarce QA capacity, escalating a vendor integration delay, approving a project recovery plan, accepting a compliance exception, prioritizing new intake, adjusting capital allocation.

Selection. One decision with high materiality, a willing owner, accessible data, a recurring cadence, and a result measurable inside 60 to 90 days.

Pilot. The platform connects to the first sanctioned data source, the decision is instrumented end to end, and the operating cadence runs on it: weekly decision-queue triage, biweekly sponsor review, monthly value and governance reviews.

Success, agreed in writing before work starts. At least one executive decision is made or materially improved using the product, the decision brief took materially less preparation effort than today, and the outputs show their evidence, limitations and approval trail.

10. Architecture and Data Posture, at the Business Level

Three layers, each with a clear owner. The experience layer is the decision workspace, the builder and the governed actions, tenant-isolated and role-aware. The analytics layer computes the seven-domain methods with every output carrying method, confidence, assumptions, limitations and version, inside the customer-facing platform boundary rather than on a third-party AI host. The record layer is the provenance ledger holding the decision record, the scoring ledger and the AI analysis record, chained so they can be verified later.

A pilot demonstrates value on operational and project metadata; it does not require raw regulated data. Where regulated data later becomes necessary, the architecture states what data, why, who

can access it, where it is processed and retained, how it is logged, and how a user can challenge an output. The platform's core governance and detection capabilities are patent-pending.

11. Where the Product Stands

The decision-first analytics workspace, the role roll-ups, the Decision Builder with its plain-English parser and add-to-dashboard, and the trust score with its two-choice resolution are built and running. The append-only scoring ledger described in section 6 is real and hash-chained today.

Not yet built: the decision record and approver routing described in section 7, the seven-domain dataset and its live ingestion pipeline described in sections 3 and 5, live connection to a customer's PPM data warehouse, the embedded assistant explaining a decision from evidence, and the outcome-scored calibration loop. Overwatch is in private beta with federal partners and Orchestrator is in limited release, as stated in section 2; today's analytics run on Sherpa and the trust score layer. Collaboration features such as mentions and write-backs are also ahead. Each of these is gated on proof that it does what it says before it is shown to a customer. Where a demonstration uses seeded data, it is labeled seeded on the card. That discipline is a product rule, applied here to this document as much as to any card in the product.

I would rather show a smaller thing that is real than a larger thing that is a slide.

About the Author

Ron Griffin is the founder of SpecOps.AI. He has spent 26 years building and delivering enterprise technology, most recently as the founder of a global technology delivery company, and holds executive master's degrees in business and in technology. He built SpecOps because the tooling in the operating review has never been designed for the decision.