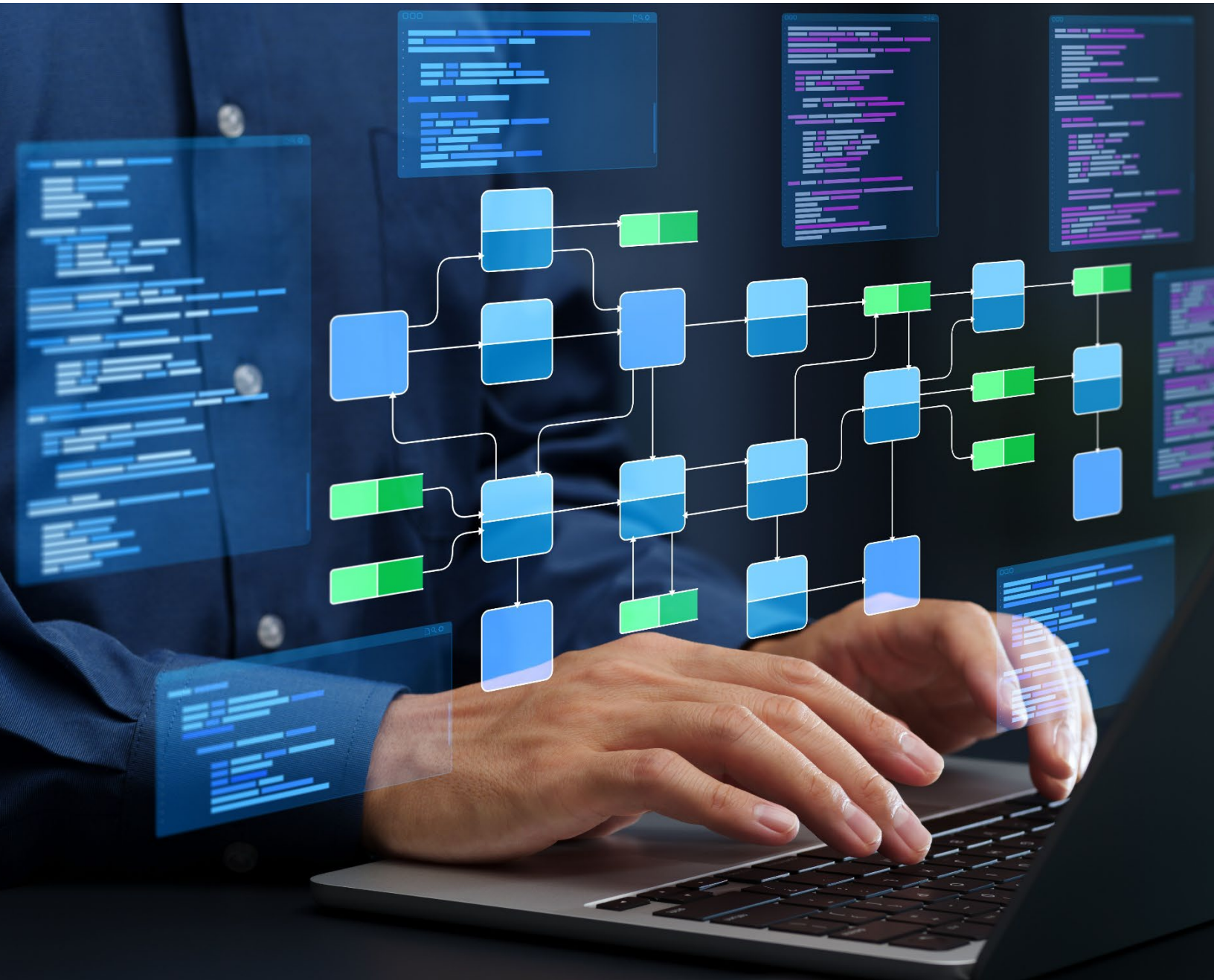
From Opportunity Identification to Opportunity Realization:

How AI Closes the Procurement Savings Gap

Authors: Vishal Suri, Pinky Singhal, and Rajat Singh

Evalueserve | 2026



Executive Summary

Organizations begin their procurement analytics journey with spend analysis. The step routinely identifies savings opportunities of 10 to 15 percent. Yet most organizations realize half or lees of that potential. The difference is the execution gap between identifying an opportunity and capturing it.

Today's macro-economic environment makes this gap starker. Commodity indices, labor costs, logistics rates, and supplier utilizations move faster than the quarterly or annual cycles at which most procurement teams refresh their models. By the time an opportunity is identified, quantified, and presented to a buyer, the underlying economics shift. Due to the static nature in which it is conducted, the analysis goes stale before action can be taken.

Agentic AI changes this equation. This paper describes four interconnected capabilities that together transform cost optimization from a periodic project into a continuous organizational discipline:

- AI-enabled spend intelligence that surfaces opportunities in real time
- Dynamic should-cost modeling, refreshed against live commodity and market data
- Automated supplier quote parsing and apples-to-apples benchmarking
- Embedded execution that delivers negotiation intelligence directly into the buyer's workflow

We illustrate the case through two Fortune 500 case studies where these capabilities were implemented and measurable savings realized. We also address what has not changed: the indispensable role of human judgment, domain expertise, and critical thinking in any AI-augmented procurement program.

I. The Identification-to-Realization Gap: Why Traditional Models Fail

How Organizations Analyze Spend

The dominant model in procurement analytics remains episodic. Spend reviews occur once or twice an year. Benchmark comparisons are rarer still. Cost models, when they exist at all, live in Excel workbooks on individual analysts' desktops. They are refreshed infrequently, and not trusted by the very practitioners that are supposed to use them. Negotiation preparation depends heavily on tribal knowledge and whoever happens to be available.

This approach leads to low accuracy today as that market conditions are less stable, while better data is available to analyze the spend.

Why This Systematically Underdelivers

The fundamental mismatch is one of cadence. The elements of operating and capital expenditures, supply markets' available capacity, and overhead structures that determine what a product or service should cost move far more dynamically. Manual refresh cycles move on a quarterly or annual basis. The mismatch causes leakage in opportunities identified.

The consequence is that by the time a model is built, validated, socialized, and acted upon, its inputs are out of date. Buyers either negotiate against yesterday's benchmark or simply do not get the actionables in time. And eventually stop trusting the modeling environment. Sourcing decisions get made with less than adequate knowledge of price and cost projections.

What Has Changed to Make AI Viable Now

Some key developments have converged to make AI-powered cost optimization continuous :

Data connectivity: Real-time data feeds via API can be fetched for commodity indices, logistics indices, labor market databases, and supplier portals, enabling cost indices to be kept live.

LLM and agentic AI capabilities: Large language models can now reason through an ask, parse unstructured supplier response documents, PDFs, spreadsheets, emails with embedded tables and normalize them for comparison in minutes rather than days.

Workflow integration: AI can now be embedded natively inside procurement platforms, rather than sitting in a separate analytics environment that buyers rarely visit.

The shift from descriptive to prescriptive analytics: While the traditional BI and MIS tools describe what happened, AI-enabled accelerators recommend what to do, with what leverage, and why.

What Has Not Changed

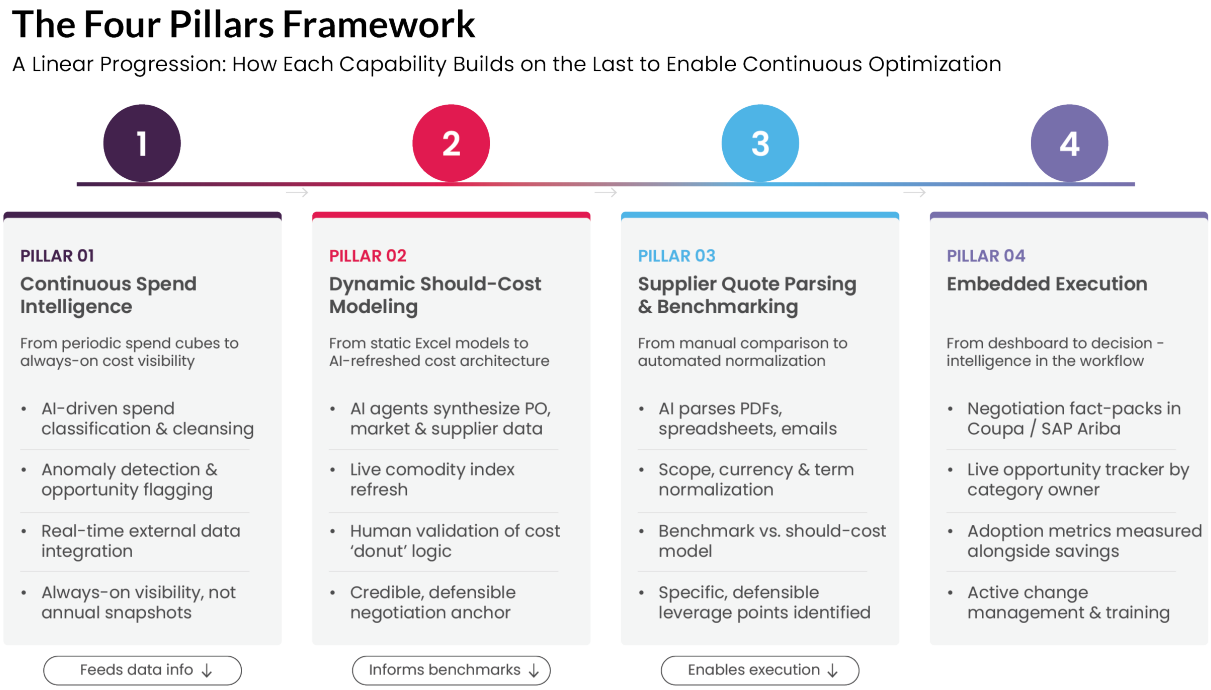
It is equally important to be clear about what AI does not replace. Two things remain human-led to ensure governance and accuracy.

First, communication, judgment, and inference. The foundational cost structures in should-cost models are the composition of the cost "donut" for a given product or solution. That donut or cost-pie must still be validated through interviews with engineers, manufacturing experts, and suppliers. AI accelerates and focuses that human effort; it does not substitute for it. By eliminating the drudgery of secondary data collection and normalization, AI elevates the quality of the analyst creativity and judgment calls.

Second, critical thinking. Agentic and generative AI are probabilistic by nature, and therefore have an inherent randomness in their reasoning. Their outputs must be reviewed by domain experts to recognize when a model is plausible and when it is not. This makes AI a powerful tool for experts, and a potentially dangerous one in the hands of novices. At Evalueserve, we begin every engagement with domain experts reengineering processes and explicitly mapping where deterministic logic should govern and where probabilistic AI adds value. The deterministic code needs to be equally valued where accuracy and reliability are critical.

II. Four Pillars of AI-Powered Cost Optimization

The following framework represents Evalueserve's approach to closing the identification-to-realization gap. The pillars together constitute a continuous operating model.



1 Continuous Spend Intelligence

From periodic spend cubes to always-on cost visibility

Moving Beyond the Spend Cube

The periodic aggregation of procurement data by supplier, category, and business unit has been the defining tool of first-generation procurement analytics. It remains useful as a baseline. But treating it as the primary source of cost visibility is like navigating by last quarter's weather forecast. AI-enabled spend intelligence replaces the snapshot with a continuously updated view. This requires solving several foundational data quality problems that many organizations have deferred for years.

AI-Driven Classification and Data Cleansing

Many organizations maintain category taxonomies whose documentation has not kept pace with their evolution. Supplier entity names proliferate through mergers, rebranding, and inconsistent data entry. Item descriptions vary wildly across business units and ERP instances. Suppliers themselves provide varying quality of item descriptions and prices on invoices. Moreover, the rulesets on what to categorize under which grouping keeps shifting due to evolution in company org chart. Roll-up and drill down of spend data across these groupings is needed to improve insights and decision making. However, the shifting rulesets necessitate recording of the trail of changes. AI can efficiently create and iterate on classification rulesets, maintain a record of the rulesets, determine exceptions, and systematically improve data classification management in a way that manual efforts cannot sustain at scale.

Anomaly Detection and Opportunity Flagging

With clean, classified spend data, AI can automatically identify the opportunities that matter prices paid above contracted rates, maverick spend outside approved channels, price drift across suppliers for equivalent items, and consolidation opportunities obscured by fragmented purchasing. These flags convert what was once a manual analytical exercise into a continuous background process.

The Role of External Data

Internal spend data answers the question of what was paid. External data ranging from commodity indices, published or derived benchmarks, spot and forward contract rates, etc., answers the question of what opportunity exists in improving the spend, what should be the timing of the spend, and where is it the fastest to implement. Connecting the two in real time transforms spend analysis from a record of history into a basis for action.

Key Takeaway:
Cost visibility is not a project. It is a capability that must be always-on.

2 Dynamic Should-Cost Modeling

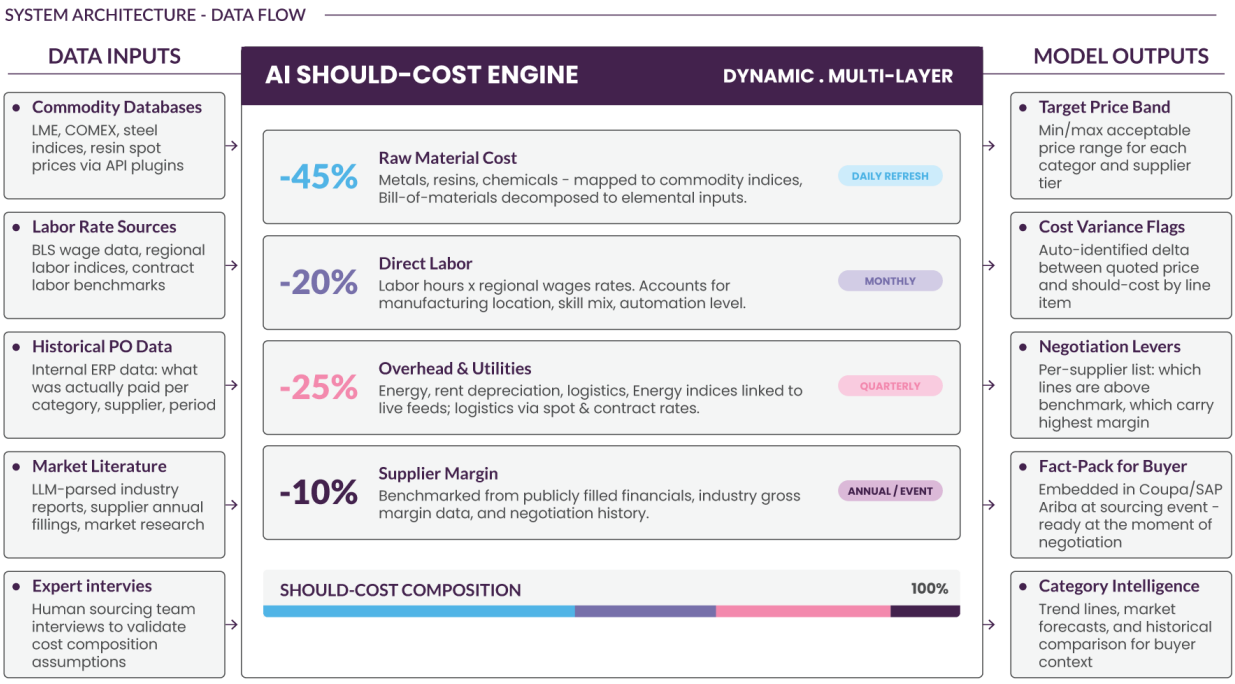
From static Excel models to AI-refreshed cost architecture

The Should-Cost Imperative

A should cost model tracks what an item or service or package or bundle costs, and how the costs have moved over time. Done right, it helps deliver assumptions and leverage points that can be used well in a data-backed negotiation. Done with the right cadence, the tools become a proactive approach to augment the otherwise intuition and relationship-based negotiations that get executed.

How AI Builds and Maintains the Model

AI agents construct should-cost models by drawing from historical purchase order data, market literature, commodity databases, and supplier interviews. They synthesize inputs that would take an analyst days or weeks to compile. The model architecture itself remains multilayered: Underlying indices and utilizations are each tracked separately and updated at different frequencies reflecting their underlying volatility.



Critically, this does not eliminate the need for human validation. AI generates the initial model; a procurement analyst validates the donut and the overall logic through direct engagement with engineers and suppliers before it is used in a negotiation. The validation step converts the probabilistic estimate into a credible negotiating position.

Key Takeaway:
A credible, current should-cost model is the single most powerful tool a buyer can bring into a negotiation. AI makes it accessible and accurate thus scaling the impact.

3 Supplier Quote Parsing and Benchmarking

From manual comparison to automated normalization and insight

The RFP Response Problem

Suppliers respond to RFPs in their own formats, with their fine print on scope assumptions, exclusions, payment terms, and lead time conventions. Producing a genuine apples-to-apples comparison across four or five supplier responses consumes days of analyst time.

Automated Extraction and Normalization

AI agents can parse supplier responses from documents, PDFs, spreadsheets, and structured emails, extracting relevant line-items and qualifications automatically. A normalization engine then adjusts for scope differences, currency, incoterms, and volume assumptions to produce a genuinely comparable view across all respondents.

Benchmarking and Negotiation Lever Generation

The normalized comparison is then benchmarked against other suppliers, the should-cost model for the category, and against historical data for equivalent purchases. For each supplier, the AI identifies where pricing is in line with expectations, where it exceeds benchmark, which line items carry the highest margin, and what terms to push on.

This output does not replace the buyer's judgment or communication influence. It ensures the buyer negotiates from a position of data-backed intelligence, with specific, defensible leverage points rather than general impressions.

Key Takeaway:

AI does not replace the buyer's judgment in a negotiation. It ensures the buyer negotiates from a position of data-backed intelligence.

4 Embedded Execution

From dashboard to decision — intelligence in the workflow

The Last-Mile Problem

Even the most analytically sophisticated procurement tool will deliver no value if buyers do not use it. And buyers, under pressure from competing priorities and time constraints, will not seek out insights that require them to leave their daily work environment. This is the last-mile problem in procurement AI: most tools generate insights in dashboards that practitioners do not access.

Workflow-Embedded Intelligence

The solution is to reduce the UI friction and bring the insight to the buyer.. Negotiation fact-packs synthesizing should-cost analysis, supplier benchmarks, anomaly flags, and recommended leverage points can be served directly inside Coupa, SAP Ariba, or Jaggaer where a sourcing event is initiated. The intelligence, the recommendation, and the sequence of actions to be taken appears in the tool the buyer already uses, weeks before they need it.

The Opportunity Tracker and Change Management

Closing the identification-to-realization loop also requires tracking identified opportunities through to execution. An opportunity tracker that serves as a live repository linking savings opportunities to specific sourcing events, category owners, and timelines ensures that no identified saving is lost to organizational inertia.

The other critical human element to design is to acknowledge that technology alone is insufficient. Adoption requires active change management: regular training, accessible helpdesks, and in some cases individualized support for high-value buyers whose engagement is critical to the system's learning and improvement. Embedded AI and the change management around it ensure improvement in adoption metrics and usage.

Key Takeaway:
Embedded AI closes the loop from savings identified to savings realized. Adoption metrics matter as much as identification metrics.

III. Case Studies

The following case studies illustrate how some of these pillars have been implemented in practice at Fortune 500 manufacturers. Details have been structured to reflect the implementation approach and lessons learned for parts of the above framework.

Case Study 1: F500 Home Appliance Manufacturer

Situation

A large North American manufacturer of home appliances wanted to streamline its cost modeling workflow. The company has moderate analytical maturity across its procurement organization with several existing platforms for procurement and cost modeling.

The company has significant direct materials spend across steel parts, plastics parts, electrical parts, and sensors. It had invested in a large cost modeling software, and all the legacy models for mechanical parts were built in that software. However, the back-end libraries for many of the parts had become outdated, due to which there was high mismatch between cost model predictions and supplier quotes.

Challenge

The company had asked suppliers to present extensive details including detailed part drawings and cost line items. The idea was to compare by line-items and determine discrepancies between the costing software models and supplier's quotes. However, the intricate details and different ways of presenting the information by suppliers resulted in buyers in the company taking weeks to compare and contrast supplier responses.

Missing information in quotes that was discovered late meant further iterations with the suppliers. Eventually, the procedure resulted in delays in sourcing events. For NPI programs, it resulted in slower time-to-market, and overall a sub-optimal negotiation process.

What Was Implemented

We started with a PoC where the goal was to prove out a) high accuracy parsing of unstructured responses; b) quick comparisons of supplier quotes against each other and cost models; c) recommendations for further requests of data from suppliers or questions for the supplier for deeper negotiations.

a) Parsing – The main challenge was not just the text, but extracting the data linked to graphics and parts drawings as well. This was accomplished through Evalueserve

proprietary Hybrid-X platform that uses state of the art multi-modal LLM's specifically designed to handle text and graphics at scale. Data extracted from text, charts, and drawings and fed into standardized, structured templates for easy comparison. Adversarial testing was used to determine AI's confidence in deriving specific outputs in the templates. Domain experts and human inputs were used to validate the datasets extracted.

b) Response correlation and comparison with cost model software – Extracted datasets were checked for completeness compared to the output expected. For responses with persistent data gaps, the AI was instructed to request further data from suppliers along the missing lines. For responses where adequate breadth and depth was present, the line-items were correlated against cost modeling line items and the discrepancies were flagged.

c) Recommendations for discussions with suppliers - Using the discrepancies identified, the comparison agent flagged the potential assumptions used by supplier vis-à-vis the computations from the cost models. The agent highlighted rationale for negotiation levers for further discussions and the questions that should be asked from the suppliers.

The PoC was successfully executed for 15 different supplier responses for steel and plastic parts. Utilizing the Agentic architecture built, it took minutes to process supplier responses. All the outputs still needed validation by sourcing experts and domain analysts, due to which the human effort was about the same during the process. Overall, productivity gains of 50-80% in turnaround were noticed in all-encompassing time comparison.

Next Steps

The PoC is now ready for production phase, where the Agentic layer will be deployed on a GCP tech-stack. Gemini Enterprise Agent platform was utilized to build the agents. Data pipelines are planned to be connected to the RFP response platform to ingest the supplier responses and to their cost modeling application to ingest cost modeling outputs. The recommendations for negotiation actions will be presented to the practitioners in both the sourcing platform as well as the Gemini setup.

The setup's biggest win has been re-establishing organizational trust in the should-cost model. Category managers who had stopped trusting the modeling output have started investigating and trusting the tool's output. The improvement in adoption and trust was executed by involving buyers in the model validation process, and addressing their concerns.

Case Study 2: F500 Semiconductor Capital Equipment Manufacturer

Situation

The client in this case is a manufacturer of semiconductor capital equipment. One arm of this manufacturer sources specialized services to extend the life of capital equipment. Detailed proprietary manuals have been created to determine the nuances and specs associated with each aspect of the service. The service packages sourced include thousands of combinations of reagents and servicing equipment across several regions of the world. Given the high volatility in the global economy, and current shift in semi-conductor value chains, the company wanted to set-up a cost modeling application to a) accurately determine implications of moving parts of their supply chain or sourcing; and, b) execute a data-backed discussion with both their suppliers and customers,than have a gut-feel based interaction.

Challenge

The company had thousands of manuals that needed to be read, assimilated and converted to models. Traditional analyst way of scaling this service would have led to extremely high costs and inordinate delays in setting up.

The company decided to utilize Agentic AI to parse the documents for cost modeling, and then utilize the manual models to train agents to model the parsed documents.

What Was Implemented

A cost-intelligence prototype was developed using the Microsoft AI stack to help organizations unlock cost insights embedded in complex service manuals. These documents typically require significant manual effort to interpret, slowing down cost analysis and limiting scalability. The solution addressed this challenge by enabling faster, more consistent, and auditable conversion of service documentation into cost-ready data.

The approach uses a structured AI workflow to automatically extract key cost drivers—such as parts, activities, and service steps—from semi-structured PDF manuals. AI-based extraction is combined with clearly defined business rules and standardization logic to ensure outputs are consistent across documents and fit for should-cost modelling and benchmarking.

To support enterprise adoption, the solution was designed with governance and transparency at its core. Every extracted data element is traceable to its original document location and includes confidence indicators, allowing cost engineers and analysts to quickly validate assumptions and focus their effort where it matters most.

The final outputs are delivered in business-friendly formats (CSV and Excel) and system-ready JSON files, enabling rapid review, cross-document comparison, and integration into cost models. As a result, the organization can significantly reduce manual analysis time, improve the speed and quality of should-cost estimates, and scale cost intelligence efforts across a broader set of assets and suppliers—without compromising control or auditability.

Next Steps

The next phase focuses on scaling cost intelligence into an enterprise capability. The solution will be productionized by expanding coverage across all service manuals and training agents to handle complex service configurations and regional variations at scale.

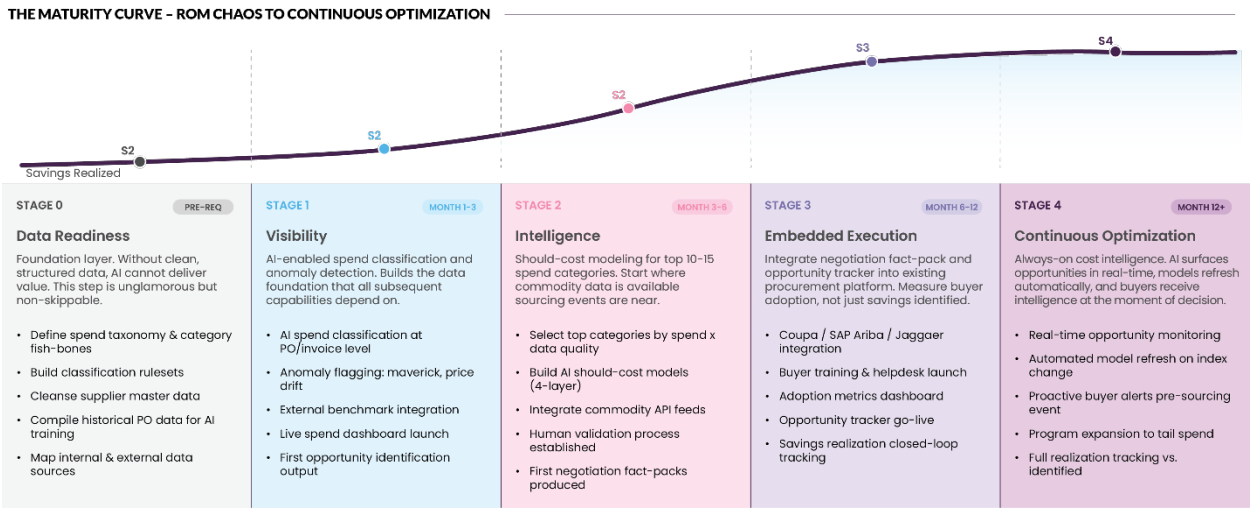
Structured cost outputs will be embedded directly into procurement and supply-chain workflows, ensuring should-cost insights are used consistently in sourcing decisions and supplier negotiations. To drive adoption, human-in-the-loop validation will be formalized, allowing cost engineers and category managers to co-review assumptions and build trust in model outputs.

The platform will then be extended to support rapid scenario analysis, enabling data-backed evaluation of supply-chain shifts and sourcing strategies in a volatile semiconductor environment. Together, these steps position the solution as a system of record for should-cost intelligence, delivering faster decisions and stronger commercial outcomes.

IV. A Roadmap for CPOs: Building the Capability

The Maturity Spectrum

The path from traditional to AI-powered cost optimization is a journey, not a switch. The following stages represent a practical sequence, with each stage building the data and organizational foundation for the next.



Stage	Capability	Priority Action
0	**Data Readiness**	Define and align on spend taxonomy, classification rulesets, and data cleansing standards. Identify and remediate historical data quality gaps. This work is unglamorous, but is also non-negotiable. Getting the data AI ready is the foundation for all subsequent analytics on the data.
1	**Visibility**	Deploy AI-enabled spend classification and anomaly detection. Build the data foundation that all subsequent capabilities depend on.
2	**Intelligence**	Build should-cost modeling capability for top 10–15 categories by spend. Prioritize categories where commodity index data is available and sourcing events are planned.
3	**Embedded Execution**	Integrate negotiation fact-packs and the opportunity tracker into the existing procurement platform. Measure buyer adoption, not just savings identified. Scale the model to achieve higher RoI and wider adoption.

Three Things That Determine Success

Across every implementation, three factors separate programs that deliver sustained value from those that stall after initial proof-of-concept:

- **Data quality**. AI is only as good as the underlying spend, contract, and supplier data. A data remediation workstream is required before AI can deliver full value in most organizations. This is not a reason to delay starting, but it is a reason to start the data work in parallel with early-stage AI deployment.
- **Human-in-the-loop design**. The programs that have succeeded have designed AI as an augmentation of the procurement professional. Analyst validation of cost models and buyer ownership of negotiations are critical elements of the process design.
- **Platform integration.** Insights that live outside the buyer's daily workflow do not get used. Integration with the procurement platform is a must-have. Adoption metrics improve much more through both integration and change management.

Build vs. Buy vs. Partner

The right operating model for most organizations combines a specialist partner with deep domain expertise in the relevant procurement categories, genuine AI engineering capability, and existing integration experience with major procurement platforms. Building this capability in-house is possible but slow; buying a point solution typically lacks the domain depth required for credible should-cost modeling in complex categories. The partner model, properly structured, delivers the integration of domain knowledge, data science, and procurement expertise that drives results.

V. Conclusion

The identification-to-realization gap in procurement is a structure and scale problem that technology can now solve. The tools exist today to make cost optimization a continuous organizational capability economically.

The organizations that will capture this advantage are not necessarily those with the most sophisticated technology stacks. They are the ones that pair AI capability with domain expertise, embed insights in workflow and equip their buyers with an accelerator.

Start with one category where you have high spend, reasonable data quality, and a sourcing event in the next 90 days. Use it as the proof-of-concept for AI-powered cost modeling and negotiation support. The results will make the case for the rest of the program.

About Evalueserve

Evalueserve helps Fortune 500 procurement and supply chain organizations monetize AI through the combination of domain expertise, data science, and platform integration. Our AI-powered cost optimization practice covers spend intelligence, should-cost modeling, supplier quote analytics, and embedded workflow execution across direct and indirect spend categories.