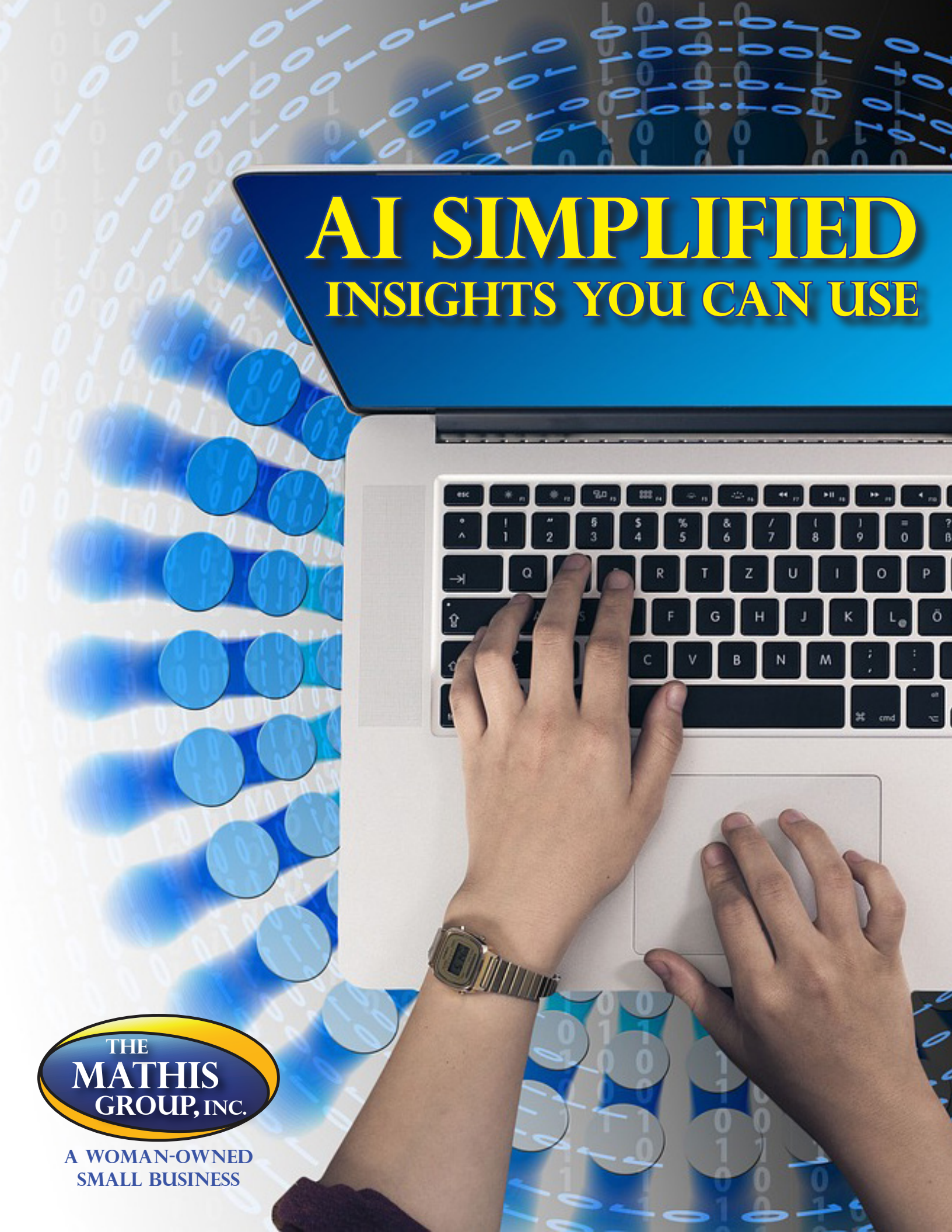

AI SIMPLIFIED
INSIGHTS YOU CAN USE
THE MATHIS GROUP, INC.
A WOMAN-OWNED SMALL BUSINESS

MARCH 2026
SELECTING AN AI VENDOR: SCORING, CONTRACT CONTROLS, AND PERFORMANCE GOVERNANCE

AI vendor selection determines control and accountability: data access, monitoring, updates, and post-go-live ownership. It determines who controls data access, who monitors model behavior, who manages updates, and who remains accountable after deployment. Most failures follow a predictable pattern: optimistic claims survive proposal review, contractual precision erodes in negotiation, and lifecycle governance is deferred until drift or incidents force it back onto the organization. This newsletter provides a practical method for scoring proposals, locking controls into contracts, and monitoring performance.

AI-ENHANCED PROPOSAL SCORING

Selection bias usually appears as inconsistent scoring, not favoritism. Different reviewers interpret the same claims differently (scope, confidence, risk). AI reduces that variance by applying a repeatable rubric to every claim and forcing evidence discipline: where does the vendor commit, where do they hedge, and where do they transfer responsibility. Reviews improve when decisions cite the same extracted text and scoring rationale.

Reducing Bias and Increasing Consistency

AI can reduce bias by applying consistent criteria across proposals. Human evaluators bring experience and judgment, but also cognitive shortcuts and organizational preferences. When trained responsibly, AI systems apply the same evaluation logic to all submissions, ensuring consistent assessment of proposals.

Bias also comes from reputation and presentation quality. A rubric reduces those effects. Different reviewers interpret identical claims differently, weigh risks inconsistently, and apply standards unevenly. AI reduces this evaluation volatility by applying uniform analytical criteria to proposals.

Transitioning From Selection to Contracting

Risk often increases during negotiation because commitments get softened. Common weakening phrases are “subject to agreement,” “estimated,” and “targets.” AI-supported contract review should run a delta analysis: what the vendor promised in the proposal versus what the contract now enforces. This is where governance moves from intention to contract language. The agreement should define measurable thresholds, logging obligations, update-approval paths, and remedies for performance degradation (National Institute of Standards and Technology [NIST], 2023).

AI tools can also assess alignment between proposals and final contracts, highlighting gaps where deliverables, timelines, or service levels may have been weakened during negotiation. This continuity

reduces downstream misunderstandings and strengthens accountability.

Monitoring Contract Performance in Real Time

During contract execution, AI-enabled analytics provide continuous visibility into vendor performance. By integrating data from project schedules, financial systems, service-level reports, and communications, AI can detect early warning signs such as schedule slippage, cost overruns, or declining service quality. Accuracy can remain high while operational reliability declines. Continuous monitoring helps detect that gap early.

Once the system is live, monitoring must move beyond status reports and be tracked continuously using service logs, incident data, user overrides, and system metrics to identify failure patterns early. These warning signs can include rising override rates, latency spikes under load, widening variance between predicted and actual outcomes, and drift signals in the incoming data stream. A challenge is maintaining high accuracy while operational reliability declines. This is why continuous monitoring helps detect problems early.

Strategic Implications for Organizations

When AI is integrated into evaluation, contracting, and performance oversight, procurement becomes a governance function rather than a one-time transaction. Decision quality improves because claims are scored consistently, contracts preserve measurable commitments, and production behavior is monitored with leading indicators rather than post-mortems. The trade-off is clear: organizations must fund data quality controls, logging, documentation, and human oversight routines that maintain accountability over time (International Organization for Standardization & International Electrotechnical Commission [ISO/IEC], 2023; NIST, 2023).

Organizations that integrate AI across proposal evaluation, contracting, and performance monitoring reduce three chronic failure drivers: misaligned expectations, delayed risk detection, and post-award information asymmetry.

However, successful adoption requires thoughtful governance. Data quality, model transparency, ethical use, and human oversight are essential to ensure AI enhances, not replaces, sound professional judgment.

The issue is not whether AI will influence vendor decisions. The issue is whether organizations will use it to enforce measurable discipline during evaluation and contracting. The real decision is whether organizations are prepared to let analytical systems challenge persuasive narratives, negotiation compromises, and long-standing evaluation habits.

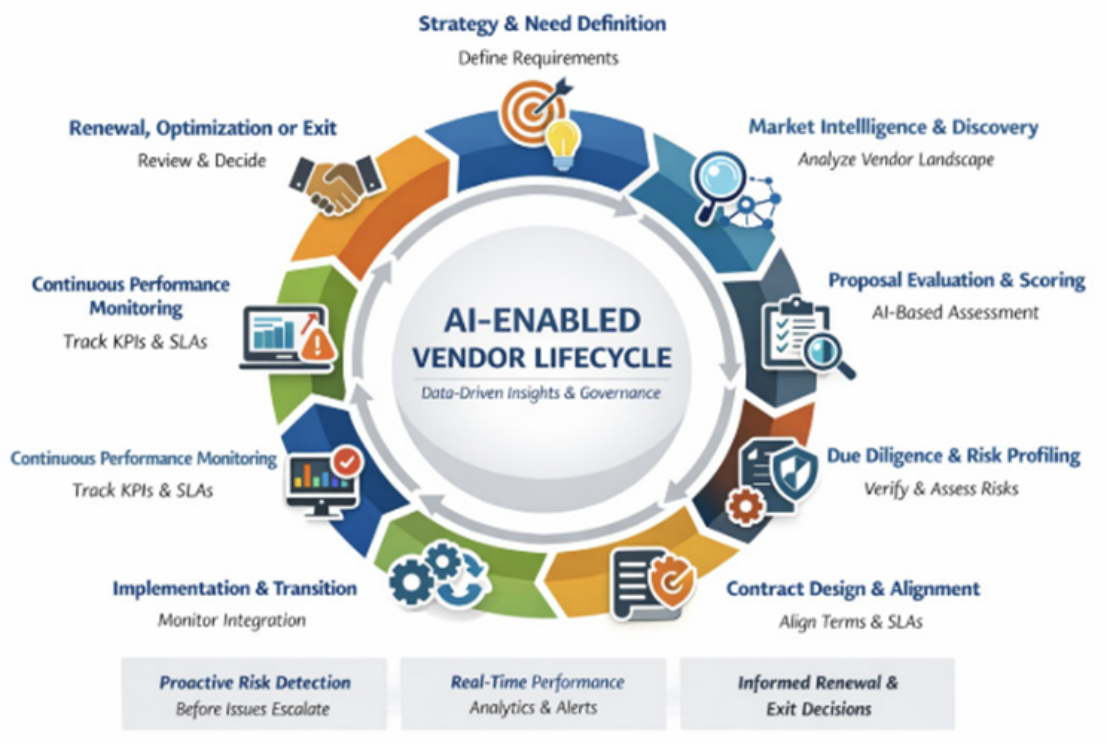
THE STRATEGIC ROLE OF THE PMO IN AI VENDOR SELECTION

AI turns vendor selection into a PMO concern because the system is never “finished” at go-live. Models drift, updates change behavior, and exceptions require decisions that must be logged and defended. The

PMO's unique value lies in its structured oversight, including stage gates for releases, change control for model updates, a decision log for exceptions, and an escalation path when performance or fairness indicators exceed thresholds (NIST, 2023). The PMO ensures structured oversight to prevent AI systems from becoming unmanaged operational risks.

A practical way to frame the PMO's expanded role is to treat AI as a product with governance, not a one-time project deliverable. NIST's AI Risk Management Framework emphasizes continuous activities across governance, context mapping, risk measurement, and risk management, all of which align with PMO disciplines such as risk registers, controls, and escalation. The PMO can translate these ideas into operating routines: decision logs, change control for model updates, and review boards for exceptions (NIST, 2023).

Finally, AI work is often cross-functional. Data owners, security, legal, operations, and business leaders all share responsibilities. ISO/IEC 42001 frames AI management as a system of policies, objectives, and processes that must be maintained and continually improved – exactly the kind of "organizational glue" PMOs are built to provide. In practice, the PMO can formalize RACI models, governance forums, and audit-ready documentation to ensure responsibilities are not overlooked (ISO/IEC, 2023).



Why AI Vendor Choice Is a PMO-Level Risk Decision

AI vendor choice is a portfolio risk decision because it shapes what you can control after deployment: how model behavior is monitored, how updates are approved, and whether the organization can audit outputs when they are challenged. Vendors also define optionality. If the platform is opaque or contract terms restrict portability, switching costs rise, and governance weakens. Model lifecycle capability matters here – ModelOps is not a "nice to have" when AI is embedded in decision workflows; it is the operating discipline that determines whether performance remains stable in production (Gartner, n.d.; NIST, 2023).

From the PMO Director seat, vendor choice shapes delivery risk in predictable ways: integration complexity, timeline realism, training quality, and operational handoff. "ModelOps" (governance and lifecycle management of AI models) makes this explicit: value depends on ongoing operationalization and control, not a one-time model build. If a vendor lacks ModelOps capability, the PMO inherits hidden work, manual monitoring, unclear ownership, and fragmented controls (Gartner, n.d.).

Vendor choice also affects the organization's strategic flexibility. If the vendor's platform is a black box or if contract terms restrict portability, future upgrades, audits, or vendor changes become expensive and time-consuming (ISO/IEC, 2023).

ISO/IEC 42001 explicitly highlights lifecycle management and third-party oversight as part of a structured management system approach, which is a strong signal that vendor governance is not optional if the organization intends to scale AI responsibly (ISO/IEC, 2023).

ORGANIZATIONAL READINESS AS A PRECONDITION FOR VENDOR SELECTION

The organization must define the decision or workflow that the AI system will support. Without a clear use case, vendor proposals turn into marketing comparisons rather than solution comparisons. The PMO can require each business sponsor to state a measurable outcome, a current baseline, and the operational boundary conditions (who uses it, when, and what happens if it is wrong). This aligns with NIST's "Map" function: understand context, intended use, and potential impacts (NIST, 2023). Without that clarity, proposals become marketing contests. With it, evaluation becomes defensible: every scored requirement traces back to an articulated operational need and risk profile (ISO/IEC, 2023; University of Pittsburgh Institute for Cyber Law, Policy, and Security, 2025).

Finally, clarity improves procurement defensibility. When evaluators can trace every scored requirement back to a documented business need and risk profile, selection decisions are easier to explain to leadership, auditors, and stakeholders. Public-sector procurement guidance similarly stresses surfacing data limitations and context early so vendors do not bid on assumptions that collapse during delivery. Even in private-sector settings, this principle reduces rework and renegotiation (University of Pittsburgh Institute for Cyber Law, Policy, and Security, 2025).

Assessing Data Readiness and Infrastructure Maturity

Data readiness must be confirmed before selection. If you cannot document data lineage, ownership, and quality standards, the vendor will either expand the scope or deliver a fragile system. Data can degrade over time, requiring retraining periodically. Infrastructure maturity is the parallel constraint: identity, access control, logging, and integration patterns determine whether the system is observable and auditable in real workflows. Monitoring must be designed from day one because drift is not an edge case; it is a routine condition of deployed models (Bayram et al., 2022; Hinder et al., 2024).

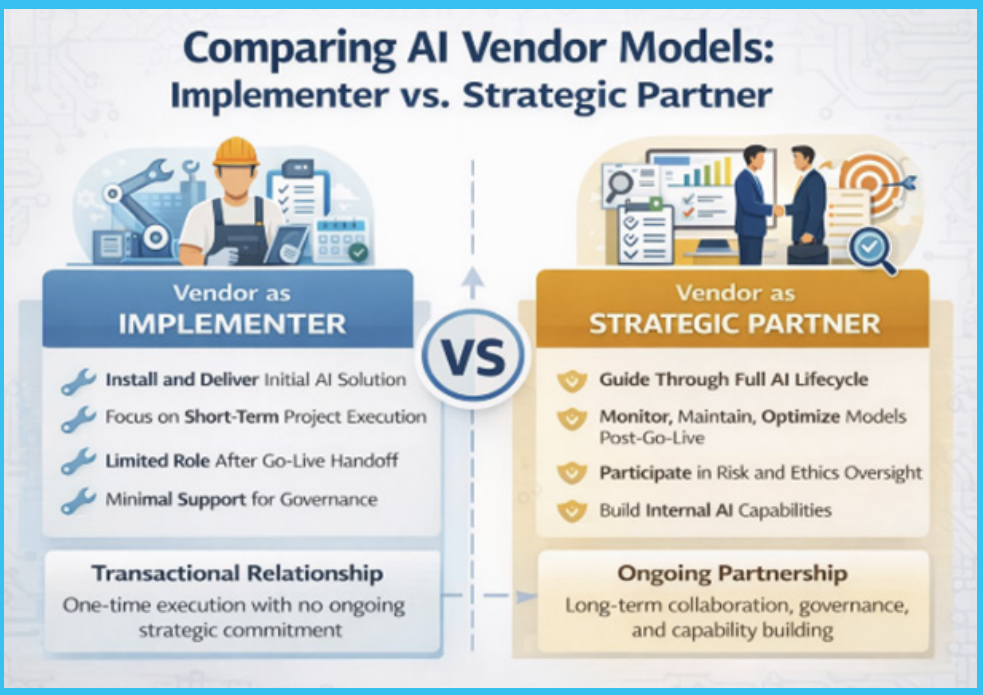
Infrastructure maturity matters just as much as data. An AI system often depends on stable integrations with identity management, logging, security tooling, and operational systems. PMOs can reduce

delivery risk by requiring vendors to demonstrate how their systems fit the organization's architecture (cloud/on-prem/, or hybrid), how logs support auditing, and how controls operate in real workflows. ISO/IEC 42001's "management system" view supports that AI governance requires repeatable processes, not ad hoc scripts (ISO/IEC, 2023).

Readiness also needs a monitoring plan from day one. Models can degrade due to concept and data drift after deployment, making early monitoring a core requirement rather than a "phase two" add-on. Research surveys show that drift is a well-established operational problem and underscore the need for detection and performance-aware monitoring. PMOs should require vendors to specify drift detection, thresholds, and response workflows during selection (Bayram et al., 2022).

DEFINING THE AI VENDOR RELATIONSHIP MODEL

Building relationships with the vendor is imperative regardless of the type of project one is running, but especially when implementing AI into the organizational culture. There are two main models in the vendor relationship that connect in different ways: utilizing the vendor as the implementer or as the strategic partner. The implementer role allows organizations to design, plan, and verify their internal insights and planning, while a strategic partner begins on day one, guiding the organization toward AI implementation.



Vendor as Implementer vs. Strategic Partner

Before the RFP is issued, the PMO should choose the relationship model: implementer or strategic partner. For higher-impact systems, the partner model can reduce risk by funding monitoring, incident response, retraining workflows, and controlled updates. The constraint is non-negotiable: the partnership must build internal capability. Contracts should require runbooks, documentation, role-based training, and an exit-ready operating model so the organization remains accountable even when a supplier is involved (ISO/IEC, 2023).

The PMO Director should determine the relationship model before issuing an RFP, as it sets the evaluation criteria and contract structure. If the vendor is expected to remain involved, proposals should be scored on operational support, incident response, monitoring coverage, and governance participation, not just on implementation milestones. Procurement-oriented AI guidance for public agencies similarly treats procured AI as creating ongoing governance obligations that must be planned from the start (University of Pittsburgh Institute for Cyber Law, Policy, and Security, 2025).

Partnership does not mean dependency. The PMO should insist on a partnership that includes internal capability-building: documentation, training, and defined handoffs. ISO/IEC 42001's focus on continual improvement and supplier oversight supports this stance: the organization remains accountable for its AI management system even when it uses third-party suppliers. A good vendor helps the organization mature rather than remain reliant (ISO/IEC, 2023).

EVALUATING AI VENDOR CAPABILITIES

Evaluating the capabilities of an AI vendor can be difficult because there is limited AI expertise in the consulting industry. One way to evaluate the AI vendor is to look at other organizations, both private and public, to see which companies they use and, even more, what level of success they are experiencing. Stay alert to any major problems with other companies in the AI project and with the companies they used for their implementation. Every organization wants a technically competent organization that is also an architectural fit.

Technical Competence and Architectural Fit

Technical competence should be scored based on how it fits the environment, not on feature breadth. Require reference architectures, data flow diagrams, logging designs, and security control descriptions as proposal artifacts; these artifacts expose integration risk early and determine audit readiness later (NIST, 2023). Explainability should carry explicit weight because leaders must be able to interpret outputs, challenge them, and document how decisions were made – an expectation echoed in emerging regulatory guidance that emphasizes transparency and documentation for higher-risk uses (European Commission, n.d.). Domain experience reduces ambiguity: vendors who understand the constraints estimate more accurately, integrate faster, and document more realistically.

Explainability and transparency deserve explicit weight in scoring. For many organizations, leaders must be able to interpret outputs and challenge outcomes, especially where AI influences high-stakes decisions. The EU's AI regulatory framework emphasizes transparency obligations for high-risk systems and instructions for use that communicate capabilities and limitations – ideas that translate well into vendor requirements even if the organization is not EU-based (European Commission, n.d.).

Finally, technical maturity must include lifecycle operations. Vendors should demonstrate how models are versioned, tested, monitored, and updated. This is central to ModelOps thinking, which frames AI value as a function of operational governance across the full lifecycle. A PMO that selects a vendor without lifecycle strength effectively agrees to bear a hidden operational burden later (Gartner, n.d.).

Domain Experience and Organizational Fit

Domain experience reduces ambiguity. Vendors who understand the industry can anticipate constraints, data quirks, and compliance requirements, thereby improving estimation accuracy and reducing rework. This is especially important because AI systems can have downstream impacts on people and processes that are difficult to predict from a purely technical perspective. NIST underscores that risk management must be grounded in context and intended use, where domain experience is critical (NIST, 2023).

Organizational fit matters alongside domain experience. A vendor can be technically strong but still fail if their delivery style clashes with governance expectations (e.g., weak documentation, unclear escalation procedures, or insufficient stakeholder engagement). The Information Commissioner's Office (ICO) guidance emphasizes accountability practices and documentation requirements for organizations using AI, requiring vendors to operate under disciplined governance rather than merely agile experimentation (Information Commissioner's Office, 2023).

PMOs can operationalize "fit" through structured proofs: scenario-based demos using your data shape, implementation plans that name required internal roles, and reference checks focused on change management and adoption outcomes. Public-sector procurement guidance also recommends addressing limitations and practical realities upfront to prevent vendors from proposing solutions built on unrealistic assumptions. This is a useful PMO lesson across sectors (University of Pittsburgh Institute for Cyber Law, Policy, and Security, 2025).

ETHICAL GOVERNANCE AND RESPONSIBLE AI

Ethical issues are one of the main reasons people fear AI. Creating guidelines will help address ethical risks and provide direction on how to use data within the organization. This guidance is essential for protecting the organization and data.

Ethical Risk as a Formal Project Risk

Ethical risk should be treated like any other project risk. If unmanaged, it can lead to operational disruption and the need to implement trusted mitigation plans. During the selection process, require vendors to provide evidence of bias testing methods, override pathways, audit trails, and mitigation

plans (NIST, 2023; OECD, 2020). For privacy and security, demand explicit data-handling terms: where the data lives, who can access it, how long it is retained, and how incidents are handled. The ICO's guidance emphasizes accountability and documentation requirements for AI processing of personal data; requirements that must be built into vendor governance routines, not treated as post hoc compliance work (ICO, 2023). NIST provides a practical way to embed ethics into standard controls by treating fairness, accountability, and explainability as measurable outcomes supported by governance routines. For vendor selection, PMOs can require evidence of bias testing methods, audit trails, and documented mitigation plans. This converts "ethics" from a values statement into deliverable, testable work products (NIST, 2023).

The PMO Director should also require a clear stance on human oversight. Governance plans should specify who can override AI outputs, when overrides are expected, and how the organization learns from overrides. This is consistent with major frameworks and regulations that treat transparency and oversight as essential to safe deployment. Even when not legally required, it is operationally wise and reduces reputational risk (European Commission, n.d.).

Data Privacy and Security Responsibilities

AI vendor selection must treat data as a controlled asset. Vendors may need access to personal, confidential, or proprietary information for training and operation. The PMO should require clear data-handling descriptions: which data is used, where it is stored, how long it is retained, and how it is protected. The ICO outlines accountability measures for AI systems that process personal data, including DPIA considerations and governance controls (ICO, 2023).

Security expectations should be tied to concrete mechanisms: encryption, access controls, logging, incident response, and third-party risk management. ISO/IEC 42001 reinforces the idea that AI governance includes risk management and supplier oversight as part of a structured management system. PMOs can translate this into contract clauses and acceptance criteria (ISO/IEC, 2023).

PMOs should also plan for "security as operations," not "security as a checklist." AI systems can change as models update and data pipelines evolve, so controls must be continuously monitored. NIST's emphasis on ongoing measurement and management aligns well here: risk posture can shift over time, and governance must be able to detect and respond to these shifts. This belongs in the vendor selection scorecard, not just implementation planning (NIST, 2023).

CONTRACTING FOR AI SYSTEM IMPLEMENTATION

AI projects differ from typical software development projects. They do not have fixed start and end points with set deliverables. Instead, AI projects require ongoing training and retraining. Additionally, the implementation team must share project knowledge with the organization or operational team.

Moving Beyond Fixed Deliverables

AI contracts fail when they treat performance as static. Acceptance metrics at deployment mean little if contracts are silent on post-production degradation. A stronger pattern pairs outcome measures with governance obligations: monitoring dashboards, drift thresholds, escalation timelines, and remedies when performance breaches occur. Update governance must be explicit (who approves changes, what testing is required, how rollbacks work, and how incidents are reported) because models evolve even when the contract pretends they do not (Bayram et al., 2022; NIST, 2023).

A better contract pattern is performance-based milestones paired with governance deliverables. Performance milestones define measurable outcomes (accuracy, false positive rates, response times, uptime, and user adoption targets). Governance deliverables define monitoring dashboards, drift detection, audit logs, and periodic review cadences. NIST's framework supports this approach by emphasizing measurement and management throughout the lifecycle rather than the "ship and forget" approach (NIST, 2023).

PMOs should also ensure contracts specify how the system will be updated safely. Who approves model changes? What testing is required before deployment? What happens if the system's outputs cause harm or policy violations? Major regulatory frameworks emphasize transparency and instructions for use; contracts should mirror these expectations by requiring clear documentation of limits and proper operational use (European Commission, n.d.).

Intellectual Property and Knowledge Transfer

IP and knowledge transfer determine whether the organization owns its future. If the vendor trains models on your data, the contract must clarify ownership of the trained models, any derivative improvements, and the rights to portability. Knowledge transfer is what turns that ownership into operational control: documentation, runbooks, model inventories, and training that enable a third party – or your team – to operate and audit the system. Without exit readiness, governance becomes fragile because accountability depends on a supplier's availability and willingness to explain system behavior (ISO/IEC, 2023; NIST, 2023).

Knowledge transfer is not a training activity; it is a control mechanism. When vendors retain exclusive operational knowledge, organizations inherit dependency risk, audit limitations, and governance blind spots. Structured documentation, model transparency artifacts, and operational runbooks preserve institutional control.

Finally, knowledge transfer serves as a control mechanism to prevent lock-in. If the vendor is the only party that can operate, audit, or improve the system, governance becomes fragile. NIST emphasizes accountability and transparency – goals that are difficult to achieve when the organization lacks internal understanding. Strong PMOs bake knowledge transfer into acceptance criteria, not optional training sessions (NIST, 2023).

IMPLEMENTATION, OVERSIGHT, AND CHANGE MANAGEMENT

AI requires an organization to do business differently by working with organizational learning and supporting the adoption of various ideas and techniques. Creating a continuous improvement plan for organizational learning changes the future of any company. However, future learning requires openness not only from management but also from workers themselves.

Supporting Adoption and Organizational Learning

Adoption is where AI projects often succeed or fail. Users must trust outputs, understand limits, and know how to respond when the system flags uncertainty or risk. The PMO should require the vendor to include adoption planning for training paths, communication, workflow redesign, and user feedback loops. AI reliability depends not only on model performance but also on training, oversight, and workflow integration (NIST, 2023).

Organizational learning should be designed into the rollout. Early phases should include controlled pilots, monitoring of user decisions (including overrides), and structured lessons-learned cycles. This is where PMO rigor helps: stage gates can require evidence that adoption metrics and operational readiness criteria are met before expansion. Public procurement guidance also encourages actionable steps and oversight routines for procured AI; these translate well to enterprise PMO playbooks (University of Pittsburgh Institute for Cyber Law, Policy, and Security, 2025).

Finally, adoption must be paired with accountability. If AI supports decisions, teams need clarity on who is responsible for final judgments and how exceptions are handled. Regulatory frameworks emphasize transparency and appropriate use; operationally, this means training users not only on "how to use it" but also on when not to use it. PMOs should require vendors to provide usage guidelines and limit statements in plain language (European Commission, n.d.).

PERFORMANCE MANAGEMENT AND VENDOR ACCOUNTABILITY

Measuring vendor performance and holding them accountable might differ in an AI project from other projects. Looking for deliverables that support clear AI outcomes and align with the organization's goals will require adjustments and evolution throughout the project.

Continuous Monitoring and Improvement

Continuous monitoring also changes vendor dynamics. Performance disputes move from retrospective arguments ("we met contractual terms") to evidence-based conversations grounded in observed delivery behavior. Vendors unfamiliar with this level of visibility often resist it.

Monitoring must cover more than accuracy. It should include bias indicators, security signals, latency, uptime, and data pipeline health. NIST's emphasis on "Measure" and "Manage" supports broad

monitoring across trustworthiness dimensions and reinforces that monitoring must drive action, not passive dashboards. PMOs can translate this into operational KPIs and service credits tied to response times and remediation (NIST, 2023).

Continuous improvement should also be governed. A PMO Director can require a quarterly review cadence, model update approvals, and audit-ready documentation of changes. ISO/IEC 42001's continual-improvement orientation aligns closely with this expectation. The most effective vendor relationships treat improvements as planned, controlled releases rather than informal tweaks (ISO/IEC, 2023).

Avoiding Vendor Lock-In

Vendor lock-in in AI is rarely accidental. It emerges from proprietary architectures, opaque model artifacts, restrictive licensing terms, and undocumented dependencies. Portability requirements, exit-readiness criteria, and artifact ownership clauses function as governance safeguards rather than negotiation preferences.

PMOs can mitigate this by requiring portability: open APIs, exportable artifacts, clear documentation, and the ability to transition model operations to another vendor or in-house team. This is a governance and resilience requirement, not just a negotiation detail (ISO/IEC, 2023).

A practical PMO approach is to define "exit readiness" as part of acceptance. The vendor must provide documentation, training, and data/model inventories sufficient for a third party to take over operations. NIST's emphasis on accountability and transparency becomes difficult to meet if the organization cannot access critical information about system behavior, assumptions, and limitations. Exit readiness protects governance continuity (NIST, 2023).

Finally, lock-in avoidance supports strategic agility. AI strategy changes quickly due to regulation, competition, and evolving needs. The EU AI Act underscores that compliance expectations can evolve; even organizations outside the EU often align with these standards to reduce risk. PMOs should prioritize vendors who design for change and interoperability, not dependency (European Commission, n.d.).

VENDOR RED FLAGS EXECUTIVES ROUTINELY MISS

Vendor proposals rarely fail in obvious ways. They fail through quiet transfers of responsibility: "client will provide clean data," "performance depends on integration," "timelines assume stakeholder availability." That language is not deception; it is structural risk shifting. The danger is that capability narratives remain crisp while operational obligations remain vague. A PMO should treat that imbalance as a scoring penalty because it predicts downstream volatility – scope growth, delayed handoffs, and performance disputes once production conditions diverge from proposal assumptions.

One of the most common warning signs is precision imbalance. Vendors describe capabilities in impressive detail yet remain vague about constraints. The system sounds powerful. Boundaries,

however, remain undefined. What data conditions degrade performance? What failure scenarios exist? What decisions should the AI not support? Silence here is rarely accidental.

Another red flag emerges in effort asymmetry. Implementation timelines appear aggressive, while organizational responsibilities remain understated. Integration work is minimized. Data preparation is assumed. Change management is implied rather than specified. The proposal projects velocity. The risk, however, migrates internally.

Then there is the subtle but dangerous indicator: governance ambiguity. Who monitors drift? Who validates model updates? Who owns performance accountability post-deployment? When lifecycle responsibilities blur, operational risk expands.

Executives often approve vendors based on capability narratives. Failures usually originate in operational realities.

Example Scenario

A logistics firm deploys an AI routing engine with strong predictive performance. Accuracy metrics remain consistently high, and executive dashboards show stable results. However, operational friction grows.

Drivers increasingly override recommendations and latency spikes during peak scheduling windows. Variance patterns emerge between predicted and actual delivery times while drift indicators remain unmonitored.

Accuracy may reassure leadership; however, user behavior may indicate instability. Governance metrics, such as override rates and latency trends, often reveal issues that accuracy alone conceals. Behavior contradicts stability.

When broader governance metrics are introduced (override frequency, rate of performance change, or latency deviation), early instability becomes visible. The issue is not model correctness, but system reliability under real-world operating conditions. Governance metrics reveal what accuracy is concealing.

QUESTIONS PMOS SHOULD FORCE VENDORS TO ANSWER

Strong PMOs evaluate vendors through friction. Ask how the system behaves when data quality degrades, what signals indicate drift, what thresholds trigger retraining, and who approves model changes? Then move beyond the technical: who approves model changes, how exceptions are logged, and what response times apply when the model causes operational harm. Finally, ask the question that reveals maturity fast: describe a recent deployment where the model underperformed – what failed, how it was detected, and what controls were changed afterward. Mature vendors answer with monitoring evidence. Immature vendors answer with stories. These are not technical curiosities; they are governance controls.

PMOs should also probe integration realism. Which systems carry the highest coupling risk? What dependencies threaten schedule stability? Where have prior implementations failed? Vendors fluent in delivery risk provide specific answers. Vendors fluent only in sales provide abstractions.

Then comes the most revealing question:

"Describe a recent deployment where your model underperformed. What failed, and how was it corrected?"

Mature vendors discuss variance, monitoring, remediation, and lessons learned.
Immature vendors redirect toward success stories. The quality of answers often predicts the quality of outcomes.

FAILURE PATTERNS IN AI CONTRACTS

AI contracts rarely fail because deliverables are missing. They fail because responsibilities, monitoring, and update governance are not clearly defined.

A frequent failure pattern is the static-performance trap. Contracts define acceptance metrics at deployment, yet remain silent on post-production degradation. The model meets requirements, but performance drifts while accountability dissolves.

Another recurring issue involves update ambiguity. Even though models evolve, data shifts, and algorithms change, contracts treat the system as fixed. Who approves updates? What testing validates changes? What liabilities emerge from modified behavior? Undefined update governance creates continuous exposure.

Monitoring dilution is just as problematic. Vendors promise analytics, and contracts specify reporting, but neither clarifies the responsibilities for intervention. Dashboards are available, but decisions are still delayed, and variances increase.

Well-structured AI contracts recognize a fundamental truth: AI systems require ongoing monitoring, retraining, and approval of updates after deployment. They are governed continuously through retraining of the data and the model.

Why AI Projects Fail After Go-Live

Most AI failures do not occur during development. They surface months later, after deployment celebrations fade and operational complexity asserts itself.

One cause is model drift invisibility. No alarms trigger as performance gradually degrades. Output remains plausible, but decisions subtly deteriorate. Without explicit monitoring mechanisms, degradation masquerades as normal variability.

Another driver is ownership diffusion. During implementation, accountability is clear. Post-deployment, responsibilities fragment. IT assumes stability, operations assume accuracy, and vendors assume oversight; however, no single function has performance integrity.

Then comes the most underestimated factor: behavioral resistance. Because users override outputs inconsistently, trust erodes selectively, and workarounds emerge quietly. The system technically functions, yet it weakens operationally.

AI systems rarely fail in a single event. They degrade gradually as data, usage patterns, and ownership responsibilities shift.

Executive Decision Errors in AI Investments

Executive decision errors in AI investments occur when leaders overestimate vendor capability, misunderstand performance metrics, or underestimate ongoing governance requirements.

For Project Managers

From a project manager's perspective, executive decision errors often materialize as downstream delivery instability. Aggressive timelines, optimistic capability assumptions, and underestimated integration complexity create schedules that appear achievable but lack structural resilience. The project plan reflects optimistic assumptions rather than realistic operational constraints.

These errors also reshape execution risk. When investment decisions are framed primarily around expected outcomes rather than operational constraints, project managers inherit variance drivers that were never explicitly acknowledged. Scope volatility increases, dependency risks surface late, and change management demands expand beyond initial estimates.

For PMO Directors

For PMO directors, executive decision errors represent portfolio-level governance exposure. AI investments approved under incomplete risk modeling often disrupt resource allocation, control structures, and performance-monitoring frameworks. What begins as a single initiative often evolves into a multi-year governance commitment.

More critically, decision errors create systemic imbalance in risk. Vendors set capabilities. Business leaders set expectations. The PMO takes on stabilization responsibilities. Without a disciplined investment review, the PMO shifts from a governance body to a variance-control role.

For Executives

At the executive level, decision errors often stem from framing distortions. AI initiatives are evaluated as technology acquisitions rather than as operational systems that require sustained oversight. Demonstrations appear deterministic. Performance metrics appear definitive. Governance demands remain largely invisible at approval stages.

The consequence is rarely immediate failure. Instead, organizations experience gradual erosion due to escalating operational costs, extended stabilization cycles, and persistent ambiguities in accountability. AI investments succeed not through capability selection, but through lifecycle governance discipline.

HOW VENDORS QUIETLY TRANSFER RISK TO CLIENTS

Vendor risk transfer refers to contractual, operational, and structural mechanisms through which AI vendors shift performance, integration, monitoring, and governance responsibilities onto client organizations, often through conditional language, dependency assumptions, and ambiguous accountability constructs.

Example Scenario

An enterprise healthcare organization contracts with an AI vendor to optimize patient flow. The proposal emphasizes automation efficiency, predictive accuracy, and workflow acceleration. Contract language appears standard.

“Client will provide clean, structured historical data.”
“Client is responsible for internal system integrations.”
“Performance is dependent on data quality.”

During deployment, data inconsistencies emerge, and integration dependencies multiply. Model outputs require frequent manual interpretation. Monitoring dashboards exist, but no intervention obligations are defined in the contract.

Operational burden shifts internally. The vendor delivered the software. The organization absorbed the integration, data quality, monitoring risks, and variability. Risk was not shifted through deception. It was shifted through contract language that conditioned performance on client responsibilities.

For Project Managers

Project managers primarily transfer risk through execution friction. Responsibilities framed as “client support obligations” frequently conceal significant workload expansion. Data preparation, integration coordination, user training, and exception handling are often migrated internally, without corresponding schedule or resource adjustments.

This dynamic destabilizes delivery control. Variance arises not from technical failure but from expanded responsibility. When plans become reactive, teams absorb unanticipated operational burdens. The project shifts from implementation management to expectation reconciliation.

For PMO Directors

For PMO directors, vendor risk transfer is a governance pattern, not a delivery anomaly. Contract structures that lack explicit accountability boundaries create systemic weaknesses in control. Vendors provide tooling. Monitoring exists. Intervention ownership remains undefined.

More concerning is structural dependency risk. Proprietary architectures, opaque model artifacts, and vendor-controlled updates reduce organizational autonomy. The PMO assumes oversight responsibilities without corresponding operational authority.

For Executives

Executives often experience risk transfer indirectly through escalating costs, extended timelines, and performance disputes. Vendor commitments conditioned on idealized assumptions introduce latent exposure. Systems operate within defined parameters. Organizational realities rarely remain static.

Effective vendor governance requires recognizing that risk transfer is rarely malicious; it is structural. Vendors minimize uncertainty. Organizations absorb variability. Leadership discipline determines whether this exchange remains balanced or destabilizing.

Example Scenario

A global financial services firm approves an AI-based forecasting platform after an impressive executive demonstration. The vendor presents strong accuracy metrics, visually compelling dashboards, and a projected ROI within twelve months. The decision appears rational. The risks, however, remain hidden.

Six months into implementation, integration complexity expands. Data normalization requires unexpected engineering effort. Forecast accuracy fluctuates across regions due to inconsistent historical data structures. Change management demands escalate as analysts struggle to interpret model outputs. The system technically functions, but operational instability continues to increase.

What failed was not the AI; it was the investment framing, the executive decision prioritized capability demonstration over operational readiness, lifecycle governance, and environmental variability.

AI GOVERNANCE METRICS THAT ACTUALLY MATTER

AI governance metrics are performance indicators that assess not only predictive accuracy but also system stability, operational integrity, risk exposure, fairness, transparency, and behavioral reliability across the AI lifecycle.

For Project Managers

For project managers, governance metrics function as execution stabilizers. Accuracy alone rarely predicts operational reliability. Variance trends, latency fluctuations, data pipeline integrity, and override frequency provide earlier signals of emerging instability.

These measures reshape risk management. Instead of tracking milestone completion alone, project managers monitor system behavior. Deviations become detectable before failures mature. Control improves, and escalations become preventative rather than corrective.

For PMO Directors

PMO directors rely on governance metrics to support portfolio oversight. AI systems influence decision flows, risk posture, and operational consistency. Monitoring must therefore extend beyond technical outputs to organizational impact indicators.

Bias signals, drift velocity, incident triggers, and decision variance patterns offer governance visibility. Without such measures, AI oversight becomes episodic, reactive, and structurally fragile.

For Executives

Executives require governance metrics as decision assurance mechanisms. AI systems increasingly influence capital allocation, operational optimization, and strategic analysis. Leadership must evaluate not only correctness but also reliability, stability, and risk symmetry.

Metrics should trigger action, not just reporting. They illuminate exposure early, contextualize performance shifts, and transform ambiguity into measurable signals.

THE ILLUSION OF ACCURACY: WHEN GOOD MODELS MISLEAD

The illusion of accuracy occurs when a model shows strong statistical performance but performs inconsistently in real operations.

For Project Managers

Project managers encounter this illusion when systems technically "perform," yet operational friction increases. Outputs appear credible. Users hesitate. Exceptions multiply. Decision cycles are slow. Accuracy masks usability and trust erosion.

The risk is subtle. Performance dashboards reassure stakeholders, but field behavior contradicts metrics. Without behavioral and operational indicators, degradation remains undetected until consequences accumulate.

For PMO Directors

For PMO directors, the illusion of accuracy represents governance vulnerability. AI systems integrated into workflows may meet acceptance thresholds while generating systemic distortions (biased outcomes, inconsistent decisions, unpredictable variance).

Oversight maturity requires distinguishing between statistical success and operational trustworthiness. Governance failures rarely originate in visible breakdowns. They emerge from plausible yet flawed system behavior.

For Executives

Executives are most vulnerable to the illusion at the decision-making level. AI outputs delivered with statistical confidence create persuasive authority. Confidence appears equivalent to certainty. Certainty appears equivalent to correctness.

Strategic governance demands skepticism without resistance. High accuracy is necessary, but it is never sufficient. Reliable AI systems are defined not by correctness alone, but by transparency, stability, and predictable behavior under variability.

Example Scenario

A retail organization deploys an AI demand prediction model. Validation testing shows 94% accuracy. Confidence across leadership teams rises quickly. However, in production, distortions emerge.

The model performs well under stable purchasing patterns but struggles during promotional cycles and seasonal shifts. While outputs remain statistically confident and predictions remain numerically precise, decisions increasingly misalign with market behavior.

The illusion persists because performance metrics lack contextual qualifiers. Accuracy reflects historical data conformity, not adaptability to evolving conditions. Confidence levels reinforce trust even as errors compound.

The model met statistical benchmarks but failed under changing market conditions. Because monitoring did not account for seasonal variability, leadership interpreted historical accuracy as ongoing reliability.

CONCLUSION

In closing, all AI vendors are not the same. Check them out with their previous customers and reduce the chances of hiring a bad vendor or one who has spent their time hyping their abilities and outputs. Understanding the performance metrics and the specific deliverables for each sprint helps set expectations and improve communication between the vendor and the customer.

RUBRIC TABLE (CRITERIA, WHAT TO LOOK FOR, WEIGHT)

Category (PMO View)	Weight	What "Good" Looks Like (Evidence-based)	Typical Evidence Request
Strategic Alignment & Value Realization	15%	Clear use-case fit, outcome metrics, benefit tracking, realistic ROI narrative, alignment to org. goals	Value hypothesis, KPI tree, benefits realization plan
Delivery Approach & Implementation Capability	15%	Credible plan, milestones, dependencies, integrated change + training, PM discipline, risk/issue routines	Project plan, RAID sample, staffing plan, training plan
Technical Architecture & Integration Fit	12%	Works with your stack, clean integration, observability, performance, scalability, clear interfaces/APIs	Reference architecture, integration diagrams, API docs
Data Readiness Support & Data Engineering	10%	Practical data onboarding, data quality plan, lineage, feature engineering approach, and data governance collaboration	Data assessment method, migration plan, data model
Responsible AI (Fairness, Explainability, Controls)	10%	Explainable outputs for users, bias testing, human oversight, documentation, and governance routines	Model cards, bias test methods, override workflow
Security, Privacy, and Compliance	12%	Strong controls, clear privacy posture, audit readiness, least privilege, secure SDLC, and incident response	SOC2/(International Organization for Standardization & International Electrotechnical Commission, 2023) reports, IR plan, access controls, pen test summary
Vendor Viability & Support Model	8%	Financial stability, clear escalation, support SLAs, upgrade cadence, roadmap transparency	Support SLAs, escalation tree, roadmap, and financials.
Contracting Readiness & Commercials	8%	Terms aligned to outcomes, clear pricing/TCO, avoids hidden fees, fair liability, and workable warranty terms.	Pricing sheet, TCO model, contract redlines
Knowledge Transfer & Capability Building	6%	Enables internal ownership, role-based training, documentation, and a handoff plan, while avoiding "black box" dependencies	Training plan, documentation set, KT milestones
Performance Monitoring & ModelOps / Lifecycle Ops	4%	Monitoring, drift management, retraining plan, change control, reporting cadence, runbooks	Monitoring dashboard examples, runbooks, and retraining policy

Total: 100%

SCORING WORKSHEET (COPY/PASTE)

Use this for each vendor.

Category	Weight	Score (1–5)	Weighted Score (Score × Weight)	Notes / Evidence
Strategic Alignment & Value	15%			
Delivery & Implementation	15%			
Technical Fit & Integration	12%			
Data Readiness & Engineering	10%			
Responsible AI	10%			
Security/Privacy/ Compliance	12%			
Viability & Support	8%			
Commercials & Contracting	8%			
Knowledge Transfer	6%			
Lifecycle Ops (ModelOps)	4%			
Total	100%		/500	

Total possible points (5 × 100) = 500

REFERENCES

Bayram, F., Ahmed, B. S., & Kassler, A. (2022). From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems, 245*, 108632. https://doi.org/10.1016/j.knosys.2022.108632

European Commission. (n.d.). *AI Act: Shaping Europe's digital future*. Retrieved February 18, 2026, from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Gartner. (n.d.). *Artificial intelligence model operationalization (ModelOps)*. Retrieved February 18, 2026, from https://www.gartner.com/en/information-technology/glossary/artificial-intelligence-model-operationalization-modelops

Hinder, F., Vaquet, V., & Hammer, B. (2024). One or two things we know about concept drift—A survey on monitoring in evolving environments. Part A: Detecting concept drift. *Frontiers in Artificial Intelligence, 7*, 1330257. https://doi.org/10.3389/frai.2024.1330257

Information Commissioner's Office. (2023). *Guidance on AI and data protection*. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/

Information Commissioner's Office. (n.d.). *Documentation* (Explaining decisions made with AI). Retrieved February 18, 2026, from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence/part-3-what-explaining-ai-means-for-your-organisation/documentation/

ISO/IEC. (2023). *ISO/IEC 42001:2023—Information technology—Artificial intelligence—Management system*. https://www.iso.org/standard/42001

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1)*. https://doi.org/10.6028/NIST.AI.100-1

Organisation for Economic Co-operation and Development. (2020). *What are the OECD Principles on AI?* https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/what-are-the-oecd-principles-on-ai_f5a9a903/6ff2a1c4-en.pdf

Reuters. (2025, July 18). AI models with systemic risks are given pointers on how to comply with EU AI rules.

University of Pittsburgh Institute for Cyber Law, Policy, and Security. (2025). *Procuring public-sector AI: Guidance for local governments* [White paper]. https://www.cyber.pitt.edu/sites/default/files/AI/Procuring%20Public-Sector%20AI.pdf

THE MATHIS GROUP, INC.

A WOMAN-OWNED SMALL BUSINESS (WOSB)

Providing quality, customized training and consulting services that inspire, educate, and equip organizations to be better tomorrow than they are today.

DR. KEITH MATHIS, PMP, PMI-ACP, CSP-SM, CSP-PO
WANDA MATHIS, M.ED. PMI-ACP

PROJECT MANAGEMENT TRAINING

OVER 60 PROJECT MANAGEMENT COURSES REGISTERED WITH PMI

PRESENTATIONS THAT EDUCATE, MOTIVATE, AND INSPIRE

Since 1993, The Mathis Group has been helping organizations change worker productivity and behavior.

PROJECT MANAGEMENT
MARKETING
MOTIVATION
ORGANIZATIONAL BEHAVIOR
LEADERSHIP
CUSTOMER SERVICE

COMPANY MANDATE

The Mathis Group provides training and consulting that will impact the organization and individual while maintaining an outstanding reputation for success and integrity.

VALUES STATEMENT

Every person has worth and should be treated with respect.

AREAS OF EXPERTISE

- Curriculum Design
- Project Management
- Organizational Behavior and Development
- Management
- Agile Project Management
- Strategic Planning
- Executive Coaching
- Performance
- Team Building
- Emotional Intelligence
- Leadership
- Customer Service
- Supervisory Leadership
- Hybrid Project Management

9515 N Spring Valley Dr
Pleasant Hope, MO 65725
800-224-3731
417-759-9110
(voice/fax)

www.themathisgroup.com

keith@themathisgroup.com
wanda@themathisgroup.com

DUNS Number:
007722098
CAGE: 3C1N9
GSA Contractor Number:
GS-02F-0010V

GSA Contract Holder

ACCREDITED BUSINESS