

PROCUREMENT INTELLIGENCE

The AI Vendor Selection Playbook

A disciplined evaluation framework for selecting AI vendors without getting burned. Scoring matrices, a 20-question RFP, due diligence checklists, and the contract red flags that decide who you can trust with your data.

AI CONEXIO

Built for IT leaders, procurement teams, and operators who buy AI on behalf of the business and own the consequences.

The Vendor Trap: Why AI Procurement Goes Wrong

AI vendor selection fails for a predictable reason: buyers evaluate the demo instead of the deployment. The market has fragmented into three vendor categories, each with a different risk profile, and choosing the wrong category for a given use case is the most expensive mistake in AI procurement.

VENDOR CATEGORY	PROFILE	REPRESENTATIVE PLAYERS
Platform vendors	Full-stack infrastructure and models. Broadest capability, deepest lock-in, highest switching cost.	Microsoft Azure AI, Google Vertex AI, AWS Bedrock
Point solution vendors	Narrow, purpose-built functionality. Fastest time to value, weaker for needs outside their lane.	Harvey (legal), Gong (sales), Cohere (enterprise NLP)
Services and integration	Custom builds on your terms. Highest flexibility and highest cost, with delivery risk on the partner.	Systems integrators, boutique AI consultancies

THE FIVE LOCK-IN VECTORS

Every lock-in mechanism raises the cost of leaving. Map them before you sign, because each one is cheap to accept and expensive to escape.

- ✓ **Data residency:** your data lives in their environment and cannot be exported in a usable form.
- ✓ **Proprietary API formats:** integrations are written against non-portable interfaces that do not transfer to another vendor.
- ✓ **Model fine-tuning on vendor infrastructure:** the customized model that delivers your value cannot leave their cloud.
- ✓ **Usage-scaled pricing:** cost rises with adoption, so success with the tool steadily increases your dependency and spend.
- ✓ **Auto-renewal with penalty clauses:** contracts roll over silently and impose exit penalties that deter switching.

THE SHINY DEMO PROBLEM

Vendor demos always work because they run on curated data, controlled prompts, and a happy path the vendor rehearsed. Production fails on your messy data, your edge cases, your integration constraints, and your concurrency. The demo measures the vendor's best case; the deployment measures your worst case. Never let a demo substitute for a proof of concept run on your own data.

FOUR QUESTIONS BEFORE ANY EVALUATION BEGINS

Define Before You Shop: The Requirements Taxonomy

Vendors redirect buyers who do not know what they want. A documented requirements set is your defense against the demo, the discount, and the redirection. Specify requirements across four dimensions before you take a single sales call.

DIMENSION	DEFINES	WHAT TO SPECIFY
Functional	What the system must do	Use case description, input and output types, accuracy thresholds, language requirements
Non-functional	How it must perform	Latency, throughput, availability SLA, concurrent user capacity
Compliance	What constraints it must meet	Data residency, HIPAA, GDPR, SOC 2, audit logging, PII handling
Integration	How it must connect	Existing systems, APIs, authentication protocols, data formats

USE CASE SPECIFICATION TEMPLATE

- ✓ **Problem statement:** the friction in one sentence, tied to a business outcome.
- ✓ **Current state cost:** what the problem costs today in hours, dollars, or risk.
- ✓ **Target state metrics:** the measurable result that defines success.
- ✓ **Data inputs available:** what data exists, where it lives, and its quality.
- ✓ **Human-in-the-loop requirements:** where a person must review, approve, or override.
- ✓ **Acceptable error tolerance:** the failure rate the business can absorb without harm.

THE 20-QUESTION REQUIREMENTS DOCUMENT

A single document of twenty answered questions, four or five per dimension, prevents scope creep and stops vendors from redefining your problem to fit their product. Circulate it to every shortlisted vendor and require written responses. The vendors that struggle to answer plainly are telling you something. The document also becomes the backbone of your RFP and your scoring matrix, so the work compounds rather than repeats.

The Vendor Scoring Matrix: Eight Weighted Dimensions

Score every shortlisted vendor on eight dimensions using a 1 to 5 scale. Multiply each score by its weight and sum for a weighted total. Set a minimum threshold on the critical dimensions: a vendor that scores below 3 on Capability Fit or Data Privacy is disqualified regardless of total.

DIMENSION	WEIGHT	WHAT IT MEASURES
1. Capability Fit	25%	Does it solve the defined problem at the required accuracy on your data?
2. Data Privacy & Security	20%	Data residency, encryption, access controls, breach history.
3. Integration Complexity	15%	Time and cost to integrate with your existing stack.
4. Pricing Model	10%	Per-seat vs usage vs outcome, volume discounts, escalation terms.
5. Vendor Stability	10%	Funding runway, customer concentration, key person risk, acquisition likelihood.
6. Roadmap Alignment	8%	Does their product direction match your three-year needs?
7. Support Quality	7%	SLA response times, dedicated CSM, escalation path, community.
8. Exit Cost	5%	Data portability, migration assistance, termination terms.

SAMPLE SCORING (THREE HYPOTHETICAL VENDORS)

DIMENSION (WEIGHT)	VENDOR A (PLATFORM)	VENDOR B (POINT)	VENDOR C (SERVICES)
Capability Fit (25%)	4	5	4
Data Privacy (20%)	5	3	4
Integration (15%)	3	4	3
Pricing (10%)	3	4	2
Stability (10%)	5	2	3
Roadmap (8%)	4	4	3
Support (7%)	4	3	5
Exit Cost (5%)	2	4	4
Weighted Total	4.05	3.90	3.62

Vendor A wins on weighted total, but Vendor B scored highest on Capability Fit. If privacy is a hard gate, Vendor B's score of 3 puts it on watch. The matrix surfaces the tradeoff; the threshold rules force the decision.

The RFP Template: 20 Questions That Separate Vendors

These twenty questions, grouped into five categories, expose the gaps that demos hide. Require written answers. Evasive or templated responses are themselves a finding.

Architecture (4)

- ✓ Where is data processed, by region and provider?
- ✓ How is model inference isolated per customer?
- ✓ What is the disaster recovery architecture and RTO/RPO?
- ✓ How are model updates managed and version-controlled?

Data Handling (5)

- ✓ Does our data train your models, ever?
- ✓ Where is data stored and for how long?
- ✓ How do you detect and handle PII?
- ✓ What certifications do you hold?
- ✓ What happens to our data when we terminate?

Performance (4)

- ✓ What are your contractual SLA uptime commitments?
- ✓ What is P95 latency at our expected query volume?
- ✓ How do you handle degraded performance?
- ✓ What monitoring and dashboards do you provide?

Pricing (4)

- ✓ What exactly is the pricing model?
- ✓ What triggers a price increase?
- ✓ How are usage overages billed?
- ✓ What are the auto-renewal terms?

References (3)

- ✓ Provide three customers in our industry with a similar use case.
- ✓ What is your average customer tenure?
- ✓ What are the top three reasons customers have churned?

The reference questions are the most revealing. A vendor that cannot name three same-industry customers, or that dodges the churn question, is showing you the risk before you buy it.

Due Diligence Checklist:

Verify, Do Not Trust

The vendor's claims are the start of diligence, not the end. Verify the certifications, stress-test the financial health, and run the reference calls before any signature. Each item below is a verification, not a question.

SECURITY CERTIFICATIONS TO VERIFY

- ✓ **SOC 2 Type II** report, current and unqualified.
- ✓ **ISO 27001** certification.
- ✓ **FedRAMP** if any government data is involved.
- ✓ **HIPAA BAA** signed if health data is involved.
- ✓ **GDPR DPA** in place if any EU data is involved.

FINANCIAL HEALTH INDICATORS

- ✓ Funding runway of at least 18 months.
- ✓ Positive revenue growth trend.
- ✓ No single customer above 20% of revenue.
- ✓ Stable key personnel and low founder churn.

REFERENCE CHECK PROTOCOL

Ask every reference these five questions, and listen as much for hesitation as for content.

1. What did the deployment actually cost versus the original quote?
2. How long did it take to reach production value?
3. What broke, and how fast did support respond?
4. What do you wish you had known before signing?
5. Would you buy again today, and why or why not?

Red flag signals: a reference that is vague on cost, cannot quantify the result, was hand-picked and coached, or hesitates on "would you buy again."

CONTRACT RED FLAGS

- ✓ Unilateral pricing change clauses that let the vendor raise rates at will.
- ✓ Data usage rights that permit training models on your data.
- ✓ Automatic renewal with no notification requirement.
- ✓ Liability caps set below the contract value.
- ✓ Ambiguous IP ownership for custom fine-tuned models.

Build, Buy, or Partner: The Decision Framework

Most AI procurement decisions are settled by four branch questions. Answer them honestly before you compare vendors, because the right answer may be to build, not to buy.

BRANCH QUESTION	SIGNAL
1. Is this use case core to competitive differentiation?	Yes → strong Build
2. Does a mature point solution already exist?	Yes → strong Buy
3. Do you have sufficient AI engineering capacity?	No → strong Partner
4. Is time to value critical, under six months?	Yes → strong Buy or Partner

3-YEAR TCO COMPARISON (BUILD VS BUY)

COST COMPONENT	BUILD	BUY
Upfront engineering / license	Engineer time	License fees
Infrastructure	You own it	Included
Integration and customization	Internal	Integration + config
Maintenance and iteration	Ongoing internal	Vendor support
Who carries the risk	You	Shared

When Each Option Wins

Build wins when the capability is your moat and you have the talent. **Buy** wins when a proven solution exists and speed matters. **Partner** wins when the need is custom but your team lacks AI engineering depth.

The Hybrid Model

Buy a foundation, such as a platform or point solution, and build your differentiation on top of it. This captures vendor speed for the commodity layer while keeping the proprietary edge in-house, where it belongs.

Contract Negotiation Guide: Where the Real Value Is Won

Price is the most visible term and rarely the most important. Negotiate the levers, lock the service levels to criticality, secure data ownership, and write your exit before you ever need it.

SIX PRICING LEVERS

- ✓ **Volume commit discounts** in exchange for a usage floor.
- ✓ **Multi-year terms** traded for a locked rate against price escalation.
- ✓ **Pilot pricing** for the first 90 days while you prove value.
- ✓ **Payment timing flexibility** to align cost with realized benefit.
- ✓ **Free professional services credits** bundled into the deal.
- ✓ **Training and onboarding** included rather than billed separately.

SLA MINIMUMS BY CRITICALITY

Core business process	99.9% max 8.7 hrs/yr down
Supporting process	99.5% max 43.8 hrs/yr
Non-critical	99% max 87.6 hrs/yr

FIVE DATA OWNERSHIP CLAUSES

- ✓ No model training on customer data.
- ✓ Data deletion within 30 days of termination.
- ✓ Data export in a standard format on request.
- ✓ Customer retains all IP in outputs.
- ✓ DPA signed before any data access.

EXIT PROVISIONS TO REQUIRE

- ✓ **90-day transition assistance** period after notice of termination.
- ✓ **Data migration support** to move your data to a new provider.
- ✓ **Knowledge transfer documentation** for configurations and integrations.
- ✓ **Prorated refund** for any pre-paid periods not consumed.

Vendor Review Cadence: Selection Never Stops

The selection decision is renewed every quarter, whether you manage it or not. A standing review cadence catches drift early and gives you leverage at renewal instead of surprise.

Quarterly Business Review

- ✓ Performance against SLA.
- ✓ Usage and adoption metrics.
- ✓ Roadmap update and impact.
- ✓ Open escalation items.
- ✓ Next-quarter success criteria.

Switching Triggers

- ✓ Three consecutive QBR misses.
- ✓ A security incident.
- ✓ Acquisition by a competitor.
- ✓ A price increase of 30% or more.
- ✓ Key integration partner incompatibility.

MONTHLY PERFORMANCE SCORECARD

METRIC	GREEN	YELLOW	RED
Uptime vs SLA	Meets SLA	Within 0.2%	Below SLA
P95 latency	At target	+20%	+50%
Support resolution time	Within SLA	1.5x SLA	2x SLA
Feature request completion	On plan	One slip	Repeated slips
User satisfaction score	4.0+	3.0 to 3.9	Below 3.0
Cost per unit of value	Flat or down	Up to 10%	Up 10%+

RENEWAL NEGOTIATION TIMELINE

Start **120 days** before renewal with a performance review against the original case. Run an alternatives assessment at **90 days**, so you negotiate with a credible walk-away. Exchange a term sheet at **60 days**, and sign at **30 days**. A vendor that knows you have evaluated alternatives negotiates very differently from one that knows you have not.