

The AI Vendor Evaluation Checklist

40 questions to ask before signing with any AI or analytics vendor — with the red flags each answer can hide.

Yash Bhatt

AI Strategy & Enterprise Solutions Consultant
bhattiyash.com · contact@bhattiyash.com

© 2026 Yash Bhatt · Free to use with attribution

HOW TO USE THIS CHECKLIST

Buy the operating reality, not the demo

AI vendor demos are the best-rehearsed thirty minutes in enterprise software. This checklist exists for minutes thirty-one onward: eight categories, forty questions, and the red flags that live inside evasive answers. It is vendor-neutral by design — the goal is not to catch vendors lying but to surface the operating reality you are actually buying: how the product fails, what your data trains, what exit costs, and who picks up the phone.

Use it in three passes. Pass 1: send categories 1–4 as written questions before any demo; written answers create accountability. Pass 2: ask categories 5–8 live, and watch which questions get deferred to 'a follow-up with our solutions engineer.' Pass 3: score the weighting matrix on the final page with your own priorities before comparing vendors — otherwise the most polished deck wins by default.

Scoring. Rate each answer 0–2 (0 = unacceptable/evasive · 1 = adequate with caveats · 2 = strong, evidenced). Category subtotals feed the weighting matrix on the last page.

1 · COMPANY & VIABILITY

01 How long has the product existed in its current architecture, and how many production customers run it at our scale? 0 1 2

RED FLAG · Logos on a slide is not production at scale. Ask for count, not names.

02 What is your funding runway or profitability position? 0 1 2

RED FLAG · An AI vendor that dies mid-contract is a data-repatriation crisis, not just a procurement one.

03 Which three customers most resemble us, and will two take a reference call? 0 1 2

RED FLAG · Refusal to produce comparable references is a finding in itself.

04 What percentage of revenue comes from your largest customer? 0 1 2

RED FLAG · Concentration risk becomes your roadmap risk.

05 What was deprecated or sunset in the last 18 months, and how were customers migrated? 0 1 2

RED FLAG · Past deprecation behavior predicts your future migration experience.

Category subtotal: _____ / 10

2 · SECURITY & COMPLIANCE

06 Which certifications do you hold — SOC 2 Type II, ISO 27001, ISO/IEC 42001 — and will you share the reports? 0 1 2

RED FLAG · 'In progress' means 'no.' Ask for the actual report under NDA.

07 When was your last third-party penetration test, and can we see the summary? 0 1 2

RED FLAG · A vendor who won't share a pentest summary is telling you something.

08 How is customer data segregated — logically or physically — from other tenants? 0 1 2

RED FLAG · Get the architecture answer, not the marketing answer.

09 What is your breach notification commitment in hours, and is it in the contract? 0 1 2

RED FLAG · If it isn't contractual, it isn't a commitment.

10 Who at your company can access our data, under what controls and audit trail? 0 1 2

RED FLAG · Support-access backdoors are the most common quiet exposure.

Category subtotal: _____ / 10

3 · DATA HANDLING & MODEL TRAINING

11	Is our data — including prompts and outputs — used to train or fine-tune models serving other customers? RED FLAG · The single most important question on this list. Get it in writing, with the default setting named.	0 1 2
12	Where is our data processed and stored geographically, and can we pin regions? RED FLAG · Residency claims must survive subprocessor scrutiny.	0 1 2
13	What are the retention defaults for inputs, outputs, and logs — and can we set them to zero? RED FLAG · Ephemerality should be a configuration, not a promise.	0 1 2
14	Which subprocessors and foundation-model providers sit under your product? RED FLAG · You inherit their terms. Ask for the full chain.	0 1 2
15	On termination, how is our data returned and destroyed — in what format, and certified how? RED FLAG · Exit mechanics negotiated at signing cost nothing; negotiated at exit they cost leverage.	0 1 2

Category subtotal: _____ / 10**4 · MODEL & TECHNICAL CAPABILITY**

16	Which underlying models power the product, and what happens when a provider deprecates one? RED FLAG · Model churn is now routine; their migration plan is your continuity plan.	0 1 2
17	How do you measure output quality, and what accuracy/hallucination data can you share for our use case? RED FLAG · Demand evaluation methodology, not adjectives.	0 1 2
18	Can we run a structured proof-of-value on our data, with success criteria we define? RED FLAG · A vendor who resists your success criteria expects to fail them.	0 1 2
19	How does the product behave when the model is uncertain or wrong — what does graceful failure look like? RED FLAG · Every AI system fails; the design question is how visibly.	0 1 2
20	What is your documented latency and uptime SLA, with credits? RED FLAG · Ask for the historical attainment, not just the target.	0 1 2

Category subtotal: _____ / 10

5 · INTEGRATION & ARCHITECTURE

21	Which of our core systems do you integrate with natively, via supported APIs — not 'roadmap' or services work? RED FLAG · 'Integration' quoted as services revenue is a build, not a buy.	0 1 2
22	Is there a full, versioned, documented API covering everything the UI does? RED FLAG · UI-only features become automation dead-ends.	0 1 2
23	How do you handle authentication — SSO/SAML/SCIM — and role-based access at our granularity? RED FLAG · Identity integration is where 'two-week deployments' go to die.	0 1 2
24	What does a typical deployment actually require from our engineering team, in person-days? RED FLAG · Get the number, then call a reference and check it.	0 1 2
25	Can the product operate inside our VPC / on-prem if policy requires it? RED FLAG · Even if you don't need it now, the answer reveals architectural flexibility.	0 1 2

Category subtotal: _____ / 10

6 · COMMERCIALS & PRICING

26	Walk us through total cost at our realistic usage — licenses, tokens/consumption, implementation, support tiers. RED FLAG · Consumption pricing needs a modeled scenario, not a rate card.	0 1 2
27	What happens commercially when usage spikes 3× — caps, alerts, overage rates? RED FLAG · Uncapped consumption is an unbudgetable line item.	0 1 2
28	Which fees rise at renewal, and what is your historical renewal uplift? RED FLAG · Ask for the number; silence is an answer.	0 1 2
29	What is included in support versus billed as professional services? RED FLAG · The gap between these two lines is where budgets die.	0 1 2
30	Is there a pilot-to-production pricing path, or does production repricing start from zero? RED FLAG · Pilot pricing that doesn't convert is a negotiation trap.	0 1 2

Category subtotal: _____ / 10

7 · SUPPORT & CUSTOMER SUCCESS

31	What are your support tiers, response times by severity, and escalation path to engineering? RED FLAG · Sev-1 response time is a contract term, not a culture claim.	0 1 2
32	Who owns our success post-sale — a named human or a portal? RED FLAG · The handoff from sales to success is where accountability evaporates.	0 1 2
33	How are model updates and product changes communicated, and can we defer breaking changes? RED FLAG · Silent model swaps can change your outputs overnight.	0 1 2
34	What onboarding and enablement is included — and what did comparable customers actually use? RED FLAG · Unused enablement hours are a sign the product needs services to work.	0 1 2
35	Show us your status page and last two incident post-mortems. RED FLAG · Post-mortem quality is the best proxy for engineering culture.	0 1 2

Category subtotal: _____ / 10**8 · EXIT & LOCK-IN**

36	What is the contractual exit process — notice period, data export format, transition assistance? RED FLAG · Read this section before the pricing section, not after.	0 1 2
37	Which artifacts leave with us: prompts, fine-tunes, embeddings, workflows, configurations? RED FLAG · Work you paid to create should not be hostage to the platform.	0 1 2
38	Are our fine-tuned models or custom weights portable, and in what format? RED FLAG · 'Your model' sometimes means 'our model, your rental.'	0 1 2
39	What have past customers' migrations off your platform looked like? RED FLAG · A vendor who claims no one has ever left is not being candid.	0 1 2
40	If we reduce scope or seats mid-term, what flexibility exists? RED FLAG · Down-flex terms reveal how the relationship behaves under stress.	0 1 2

Category subtotal: _____ / 10

SCORING

The weighting matrix

Equal weighting is a decision to let the vendor's strengths choose your priorities. Before comparing vendors, distribute 100 weighting points across the eight categories according to your context — a regulated enterprise might put 40 points on categories 2–3; a startup buying velocity might weight 4–5. Then: weighted score = (category subtotal ÷ 10) × category weight. Sum for a comparable 0–100 vendor score.

Company & Viability	_____	_____	_____
Security & Compliance	_____	_____	_____
Data Handling & Model Training	_____	_____	_____
Model & Technical Capability	_____	_____	_____
Integration & Architecture	_____	_____	_____
Commercials & Pricing	_____	_____	_____
Support & Customer Success	_____	_____	_____
Exit & Lock-In	_____	_____	_____
Total		100	____ / 100

Decision rule. Any single question scored 0 in categories 2, 3, or 8 is a conditional disqualifier regardless of total score — security, data training rights, and exit terms are where total scores hide fatal specifics.

References & anchors

SOC 2 (AICPA Trust Services Criteria) · ISO/IEC 27001 — information security management · ISO/IEC 42001:2023 — AI management systems · NIST AI Risk Management Framework 1.0 (2023) · EU Artificial Intelligence Act (2024).

Running an evaluation now? I facilitate vendor evaluations end-to-end — question set tailoring, demo interrogation, weighted scoring, and negotiation support. → bhattyash.com/contact · contact@bhattyash.com