

DOCUMENT 1 OF 3 · 12 PAGES

Foundation-Model Vendor Scorecard

Procurement-grade evaluation of foundation-model and AI vendors — capability, data handling, security posture, IP & indemnification, exit terms and weighted scoring with named sign-off.

Audience · Procurement, Security, Legal, Engineering, AI governance

Companion documents · Industrial AI Model Card & Validation Pack · Agent Incident-Response Runbook (industrial)

Use as · Audit-grade artefact requiring named human sign-off

Version · v1.0 · 2026-05

§ 1

How to use this scorecard

This scorecard is the document a procurement team hands a vendor before contract. Each section defines the evidence required, the scoring scale, and the named approver. The scorecard travels with the contract through legal review and into ongoing vendor management.

Scoring scale

Score	Definition
5	Exceeds expectations; evidence provided proactively; auditable
4	Meets expectations with strong evidence
3	Meets expectations with adequate evidence
2	Partial; remediation plan accepted
1	Does not meet expectations; conditions for use undefined
0	Disqualifying

Weighting

Default weighting is shown per section. Adjust by use case (e.g. safety-critical control = +50% weight on security and incident); document any adjustment with rationale.

§ 2

Capability fit · evidence required

Dimension	Evidence	Score
Task fit	Benchmarks on representative workload; not vendor-curated only	
Domain coverage	Coverage of OT vocabulary, sensor data, manufacturing context	
Modality coverage	Time-series, image, structured, free text as required	
Latency & throughput	Measured on the buyer's payload distribution	
Multilingual / multi-locale	Where workforce or supplier base requires	
Fine-tuning / customisation	Method, data isolation guarantees, cost	
Retrieval / grounding	Mechanism, freshness, citation accuracy	
Determinism / repeatability	Seed control, temperature, stable-prompt behaviour	
Output quality monitoring	Vendor's own eval transparency; published model cards	

Default weight: 25%. Capability without security or operability is uninvestable; capability alone scores low if the rest of the scorecard is empty.

§ 3

Data handling & privacy

Dimension	Evidence	Score
Customer-data use for training	Default off; opt-in only; documented controls	
Logging & retention	What is retained, for how long, who can access	
Data residency	Region commitment; sub-processor regions	
Cross-border transfers	SCCs, TIA, DPF or equivalent	
Right to delete / portability	Mechanism, SLA, evidence	
Customer-controlled keys (BYOK)	Available? KMS integration?	
Confidentiality of prompts	Contractual not used outside service delivery	
PII / IP scrubbing	Filters, audit, leakage tests	

Default weight: 20%. Data-handling gaps are typically uncurable later — fix them in contract.

§ 4

Security posture

Dimension	Evidence	Score
Certifications	SOC 2 Type II, ISO 27001, FedRAMP / HITRUST where applicable	
Penetration testing	Frequency, scope, executive summary shared	
Vulnerability management	SLAs by CVSS; patch cadence; supply-chain (SLSA / SBOM)	
Identity & access	SSO, MFA, conditional access; NIST SP 800-63 alignment	
Transport & storage	TLS 1.3 (RFC 8446), AES-256 at rest, key rotation	
Logging & SIEM	Customer access to relevant logs; retention	
Incident notification	SLA, channel, content; recent history	
Insurance	Cyber liability limits, breach response coverage	
Personnel	Background checks, security training, sanction monitoring	

Default weight: 20%. Security gaps in foundation-model vendors compound: an exposed inference endpoint plus PII-leaky model can equal a multi-jurisdictional breach.

§ 5

Model governance & transparency

Dimension	Evidence	Score
Model card	Public; covers intended use, training, evals, limitations	
Eval methodology	Published methodology + sample evals; independent replications	
Update notice	Advance notice on model version change; opt-out window	
Deprecation policy	Notice period; migration assistance	
Safety filters	Documented, configurable; bypass policy	
Bias / fairness testing	Vendor's own testing + permission for buyer testing	
Hallucination / refusal posture	Stated rates with methodology	
Provenance attestation	Where training data came from; permissions; opt-outs	

Default weight: 10%.

§ 6

IP & legal

Dimension	Evidence	Score
IP ownership of outputs	Buyer owns; vendor disclaims; carve-outs explicit	
IP indemnification	Defence + damages for third-party IP claims; limit clear	
Patent & trade secret	Warranty against infringing training data	
Confidentiality	Mutual; survives termination; specific carve-outs	
Limitation of liability	Symmetric vs asymmetric; cap relative to risk	
Acceptable use policy	Reasonable; not unilaterally changeable to detriment	
Governing law / forum	Acceptable to buyer; arbitration clauses scrutinised	
Audit rights	Right to audit security / compliance; conditions reasonable	

Default weight: 10%.

§ 7

Operational & commercial

Dimension	Evidence	Score
SLA	Availability, latency, support response; meaningful credits	
Capacity & rate limits	Stated; burst handling; quota negotiation path	
Cost model transparency	Per-token / per-request; bulk pricing; observability	
Total cost of ownership	Hidden costs (egress, fine-tuning, support tiers)	
Roadmap visibility	Briefing cadence; ability to influence priorities	
Reference customers	Verifiable; comparable use case & scale	
Support	Hours, channels, named TAM at right tier	
Financial stability	Funding, runway, customer concentration, public financials if any	

Default weight: 10%.

§ 8

Exit & portability

Dimension	Evidence	Score
Data export	Format, completeness, time-bound right after termination	
Fine-tuned model export	Available? Weights or only via inference endpoint?	
Model-derived assets	Embeddings, indexes, transformations exportable	
Transition assistance	Defined; included or chargeable; SLA	
Destruction certification	Within stated time; covering sub-processors	
Renewal terms	Auto-renew constrained; notice periods symmetric	
Termination for convenience	Available with reasonable notice	

Default weight: 5%. Underweight at procurement; high regret weight at exit.

Aggregate scoring template

Section	Weight	Score (0-5)	Weighted
§ 2 Capability fit	25%		
§ 3 Data handling & privacy	20%		
§ 4 Security posture	20%		
§ 5 Model governance & transparency	10%		
§ 6 IP & legal	10%		
§ 7 Operational & commercial	10%		
§ 8 Exit & portability	5%		
Total	100%		

Decision rule

- ≥ 4.0 weighted average and no section $< 3 \rightarrow$ Approve
- 3.0–3.9 weighted average $\rightarrow$ Approve with conditions and remediation plan
- < 3.0 OR any section ≤ 1 OR any disqualifying 0 $\rightarrow$ Reject

Disqualifiers. Customer data used to train shared models without opt-out · refusal of right to audit · no IP indemnification · no incident-notification SLA · no exit clause. Any disqualifier in isolation is a stop, regardless of headline score.

§ 10

Comparative table (vendor short-list)

Section	Vendor A	Vendor B	Vendor C
Capability			
Data handling			
Security			
Governance			
IP / legal			
Operations			
Exit			
Weighted total			
Disqualifier present?			
Recommendation			

§ 11

Ongoing vendor management

Activity	Cadence	Owner	Trigger
SLA & cost review	Quarterly	Vendor mgr + Finance	Always
Security re-certification check	Annual	Security	Cert expiry
Model version change review	Per change	AI governance	Vendor notice
Sub-processor change review	Per change	Privacy + Vendor mgr	Vendor notice
Incident debrief	Per incident	Security + Vendor mgr	SEV-1/2
Re-scorecard	Annual or material change	Vendor mgr	Anniversary

Sign-off & attestation

Sign-off block

Role	Name	Date	Signature
Procurement lead			
Security			
Legal			
Privacy			
AI governance / model risk			
Business sponsor			

A signed scorecard is the entry artefact for contract execution. Re-scorecard at material change to the contract, model version, sub-processor list, or vendor financial standing.