AI Vendor Assessment Guide

33 Curated Questions · 12 Domains · 3 Tiers

Aligned with ISO 42001 · NIST AI RMF · EU AI Act

DENNIS AH KING

V2.0 · 2026

AI GOVERNANCE BLUEPRINT

A Practitioner's Approach to AI Vendor Assessment

Most organisations already have a vendor assessment process : a procurement checklist, a detailed security questionnaire, a legal review, and a sign-off procedure that has served well enough for years. For software that behaves predictably and does what it is configured to do, that process is reasonable.

> ⚠ The challenge is that AI systems do not fit that mould. Applying a standard procurement lens to an AI vendor tends to produce a **false sense of assurance**.

This guide is not a replacement for your existing procurement process. It is a supplement designed for the specific risks that AI vendors introduce, written for business leaders who need to ask the right questions in the room.

The Problem with Standard Vendor Questionnaires

The Paper Trail Problem

A questionnaire answered on paper gives the vendor time to craft responses that sound credible without committing to anything substantive. **"We take AI safety seriously and have robust processes in place"** is a complete non-answer that passes a checkbox review every time.

Why Long Questionnaires Fail

Long questionnaires create vendor fatigue. You send 100 questions and get 100 responses of diminishing quality, because the person filling in question 87 has lost interest and is copying from the previous engagement. A yes to a question barely extract the insight you need to make an informed decision.

A questionnaire is also a document review, not an intelligence-gathering exercise. You learn what a vendor is *willing to write down* and not how effective is a control.

The methodology in this guide is built around **conversation with the vendor**. A vendor who can answer these questions fluently, with specifics and examples, in a live conversation is a very different proposition from one who needs two weeks and a legal team to formulate a response.

Why AI Vendor Risk Is Different

Non-Deterministic Behaviour

These systems can produce outputs their developers did not anticipate, fail in ways that do not trigger conventional monitoring alerts, and cause harm that surfaces months after initial deployment.

Data Exposure Risks

They process personal data in ways that can bypass the access controls your security team put in place ; creating regulatory exposure that standard frameworks were not built to address.

Regulatory Accountability

They make or heavily influence decisions that regulators will hold *you* accountable for, even when the decision was generated by someone else's model running on someone else's infrastructure.

Pace of Change

The AI vendor market moves faster than most procurement processes are designed to handle. A vendor with a credible governance posture six months ago may have updated their model in ways that change the risk profile entirely.

A 10-Step Methodology

A structured framework for evaluating AI vendors with confidence from strategic alignment to final due diligence. methodology so that you can use it right away.

Before You Begin: Strategic Foundations

The first five steps happen **before** you meet the vendors. Getting these right determines the quality of every conversation that follows.

Align with AI Strategy

Does this use case sit within your approved AI strategy? Does it meet a genuine organisational need, or is it being driven by enthusiasm for the technology? If the use case is not aligned with where your organisation wants to go with AI, no vendor conversation will fix that.

Calibrate to Risk Level

Not every AI use case requires the same depth of scrutiny. A low-stakes productivity tool carries different risk than an AI system influencing credit decisions, hiring outcomes, or patient triage. Spend a few minutes with the five screening questions first.

Determine Your Tier

Each of the five screening questions in the tool carries equal weight. Count your Yes answers: **0–2 = Tier 1** (basic), **3 = Tier 2** (intermediate), **4–5 = Tier 3** (comprehensive). You can always adjust based on your organization and your judgment of the specific use case.

Review Existing Documentation

Before the meeting, review any document the vendor already provided: SOC 2 Type II reports, penetration testing summaries, security whitepapers, model cards, transparency reports, and regulatory compliance documentation. This signals preparation and focuses dialogue on gaps, not basics. You want a meaningful conversation.

Customize Your Question Set

The list of curated questions in the A- is a starting point, not a fixed script. Your organization may have sector-specific requirements, internal policies, or regulatory obligations these questions do not address specifically. The Scorecard lets you add custom questions to any domain during the session.

Running the Session

The Right Mindset

At this stage, you are intentionally **not** presenting the vendor with +100 questions. You are having a focused conversation with a curated set of questions to test governance maturity and see whether there is the prospect of a lasting business relationship.

> A vendor who answers fluently with named people, described processes, and concrete timelines is telling you something. A vendor who deflects, generalizes, or goes quiet is telling you something too.

1 Run as a 2-way conversation, Not an Interrogation

Ask the question, listen carefully, and follow the thread. A focused conversation reveals governance maturity far more effectively than a rigid checklist.

2 Use the Scorecard Live, In the Room

The AI Vendor Intelligence Scorecard is a **live conversation tool**. Do not send it to vendors in advance You want unscripted responses, not rehearsed ones. Open it at the start of the meeting, work through the questions, and capture flags and notes in real time. Export a PDF at the end.

3 Treat This as Initial Qualification Only

The output of this process is a qualification decision: is this vendor worth pursuing further or not? A clean result across all most of the domains means the relationship is worth developing. Multiple concerns or unresolved follow-ups means you either need a deeper conversation or remove them from the next qualifying step.

Initial Screening
Ask 5 focused questions and select tier.

Qualification Decision
Decide to proceed or remove vendor.

Live Vendor Session
Use scorecard and note real-time flags.

This three-stage flow ensures every vendor conversation is purposeful, proportionate to risk, and produces a clear, documented outcome.

Two Companion Tools

This methodology is supported by two companion tools designed to work together. A Vendor Intelligence Scorecard tool to capture this exactly methodology described in this document so that you can use it right away. There is also full bank of questions that you can reference at the final stage of the selection process.

FOR LIVE DISCUSSIONS

AI Vendor Intelligence Scorecard

For use during vendor conversations

A self-contained HTML tool that runs in any browser. Complete the five screening questions to receive a suggested tier, customise your question set, then run the session as a live dialogue. Capture flags and notes in real time and export a PDF report at the end.

⚠ Not to be shared with or sent to vendors.

FULL DUE DILIGENCE REFERENCE

AI Vendor Question Repository

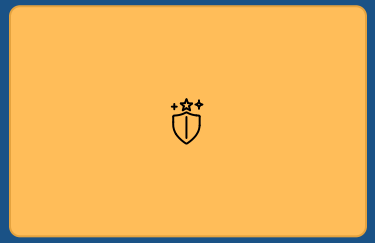
The full question bank across all **12 domains** and **3 tiers**, with complete guidance for each question.

Use this once you have selected your top vendor and are about to sign the contract.

ⓘ COMING OUT SOON. This is the next artifact in this toolkit.

The Three Tiers

Every question is assigned to a tier. Tiers are cumulative: Tier 2 covers all Tier 1 questions plus 10 additional, and Tier 3 covers all 33 questions.

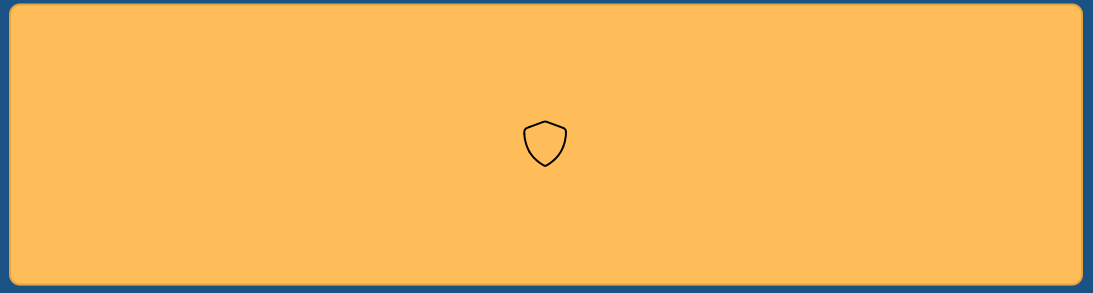


Tier 1 — Basic (12 Questions)

Low Risk. Standard AI tools with limited decision authority, no sensitive data, low business criticality. Use at every vendor conversation as the minimum baseline.

Tier 2 — Intermediate (22 Questions)

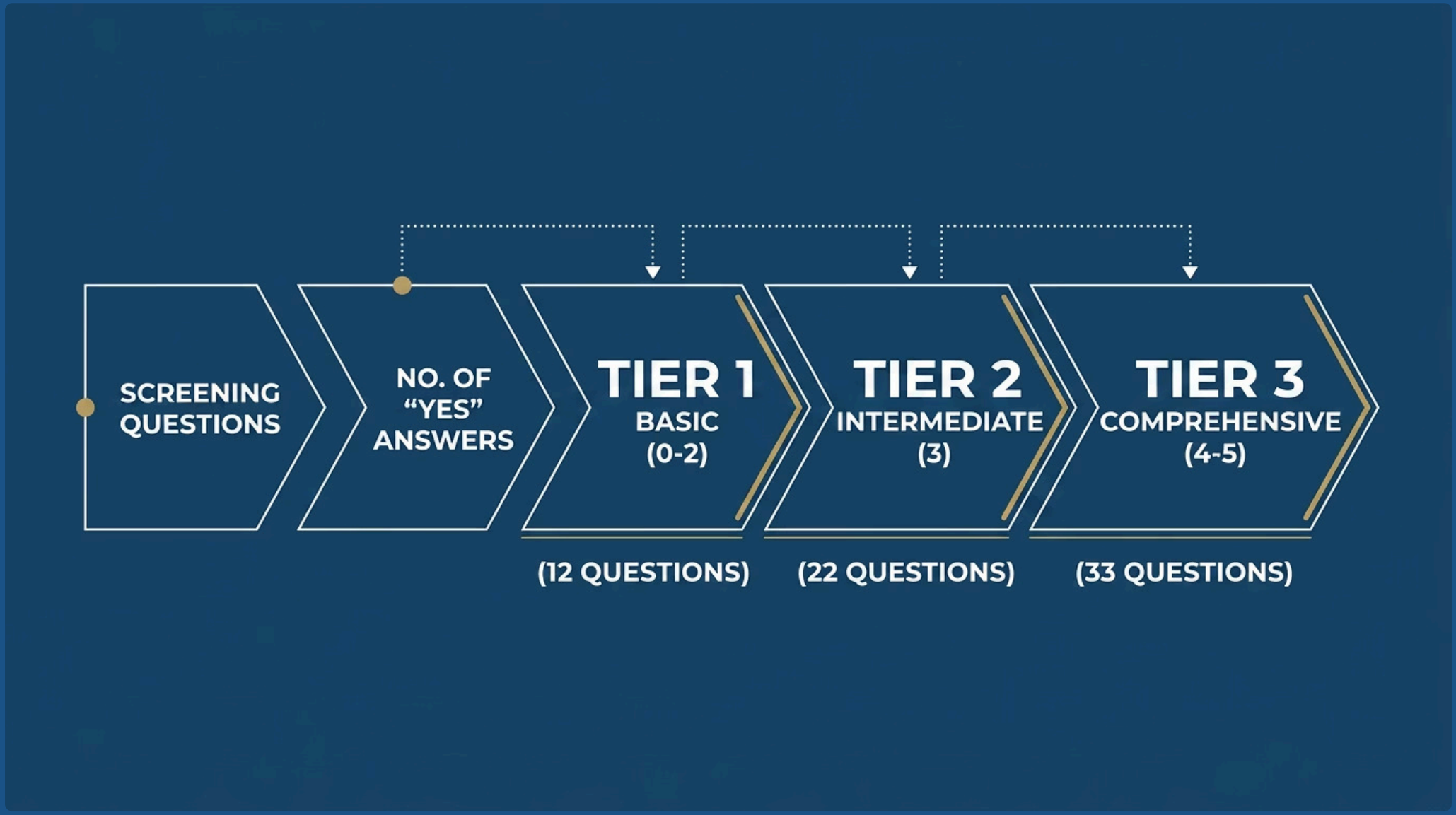
Medium Risk. AI tools processing sensitive data or influencing decisions with moderate regulatory exposure. Includes all Tier 1 questions plus 10 additional.



Tier 3 — Comprehensive (33 Questions)

High Risk. High-impact decisions, regulated data, strategic dependency, or board-level concern. Includes all questions across all 12 domains.

Scoring the Screener



Always adjust the tier up or down based on your judgement of the specific context. When in doubt, go one tier higher.

How to Score Responses

Assign one of three ratings to each vendor response as the conversation progresses. Use these consistently across the session.

✓ Satisfactory

The vendor's response demonstrates mature practice, aligns with the Look For criteria, and does not raise concerns. **No further action required** on this question.

⌑ Follow Up

The response is incomplete, vague, or requires additional evidence. **Add this to your list of items to resolve** before any contract is signed.

× Concern

The response triggers one or more Red Flags, or the vendor is unable or unwilling to answer. **Escalate and consider whether to continue** the conversation.

Question Guidance Fields

Each question includes four fields to help you evaluate what you hear in the conversation.

Why It Matters

The business or regulatory risk the question is designed to surface. Read this *before* the vendor answers so you know what you are listening for.

Look For

Indicators of a mature, trustworthy vendor response. A good answer will contain most or all of these elements.

Watch Out For

Responses that sound acceptable but should prompt further probing. These are not automatic red flags but warrant follow-up in the conversation.

Red Flags

Responses or behaviours that indicate a significant governance gap or unacceptable risk. Multiple red flags on one question is a strong signal to escalate.

The Five Screening Questions

Answer these five questions at the start of your assessment to determine the appropriate tier. Each question carries equal weight. When in doubt, go one tier higher.

1

VQ1 — Decision Impact & Data

Does this AI system make or significantly influence decisions that affect your employees, operations, finances, or safety — or will it process personal, confidential, commercially sensitive, or regulated data?

2

VQ2 — Regulatory Scope

Is your organisation subject to specific industry regulations for this use case, or does this system fall within the scope of frameworks such as the EU AI Act, GDPR, HIPAA, or equivalent?

3

VQ3 — External Visibility

Will the outputs of this system be visible to or directly affect customers, regulators, or the public in a way that could create reputational or legal accountability for your organisation?

4

VQ4 — Board-Level Exposure

Would a failure, security incident, or compliance breach involving this system create board-level concern, significant reputational damage, or material legal exposure for your organisation?

5

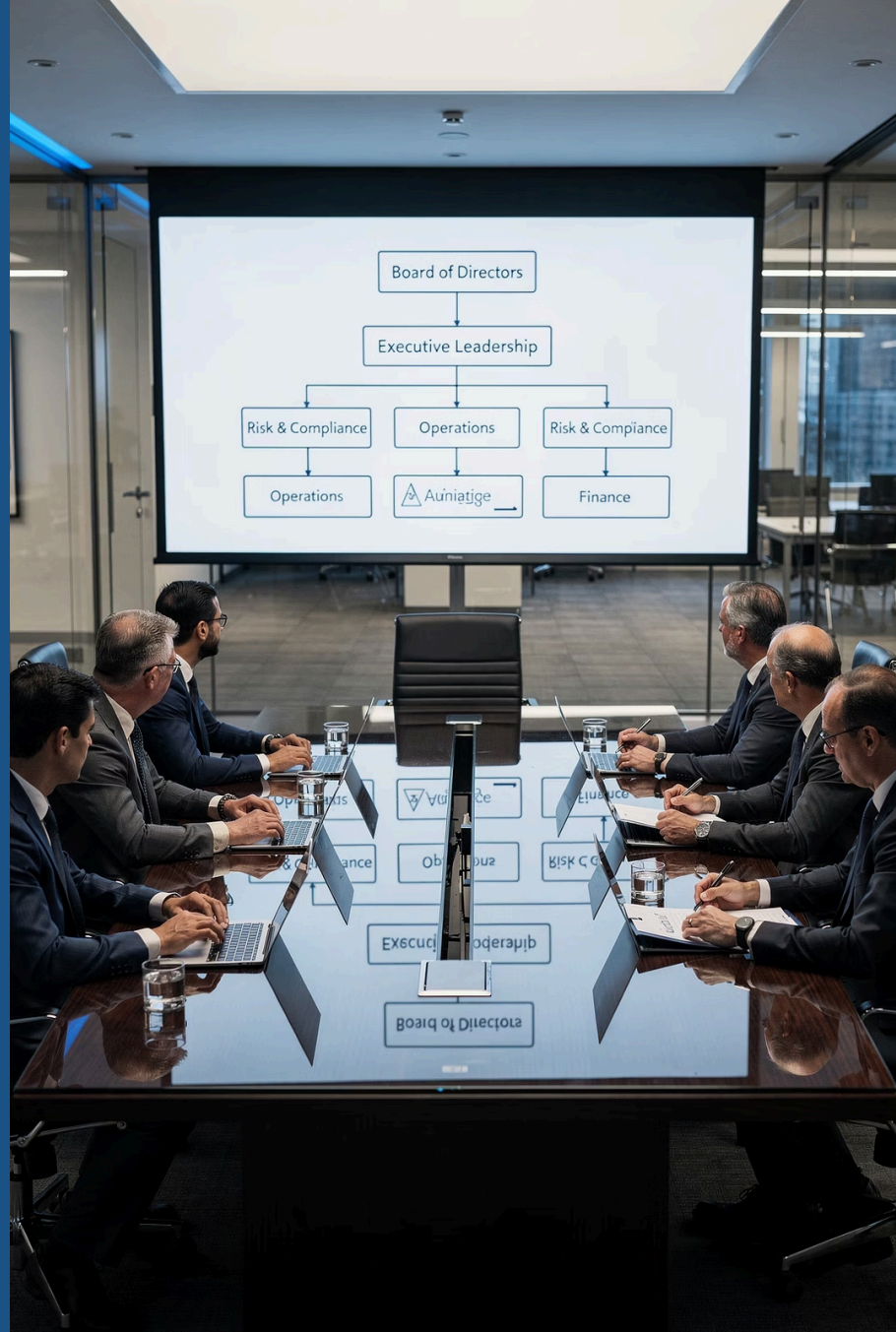
VQ5 — Agentic Capability

Does this AI system involve autonomous agents, agentic workflows, or any capability to execute actions without a human approving each individual step?

DOMAIN 01

AI Governance & Oversight

Accountability, oversight structures, and governance maturity · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**



Q1 · TIER 1 · BASIC

AI Governance & Oversight — Q1

> Is there a formal AI governance body or committee overseeing this system — who sits on it, what authority does it have, and how does it operate in practice?

Why It Matters

A vendor with genuine AI governance has a named owner and a real structure behind them. A vendor without it will give you a committee answer or deflect to engineering. Knowing who is accountable before deployment tells you a great deal about how problems will be handled after it.

Look For

Named individual with defined accountability, supported by a governance committee with a charter, meeting cadence, and documented decisions. Role is not shared or vague.

Watch Out For

"Our AI ethics board oversees all AI development" without a named individual or described mandate.

Red Flags

- No clear owner
- Governance described as a future initiative
- Accountability deflected to a team rather than a person

Q13 · TIER 2 · INTERMEDIATE

AI Governance & Oversight — Q13

> Do your developers and relevant personnel receive formal training on responsible AI, security, and privacy, and what ongoing training is required to maintain that standard?

Why It Matters

Governance policies only work if the people building the system have been trained on them. A vendor whose engineers have not received formal responsible AI training is producing a governance posture that exists on paper only.

Look For

Formal training programme described with named curriculum, frequency of mandatory refreshers confirmed, training records available for audit.

Watch Out For

"All our staff are committed to responsible AI development."

Red Flags

- No formal training programme
- Responsible AI training described as optional or informal
- Training records not maintained

Q23 · TIER 3 · COMPREHENSIVE

AI Governance & Oversight — Q23

> If we needed to explain a specific decision this system made to a regulator, what could we actually show them — and can you demonstrate that capability now with a real example?

Why It Matters

Explainability is not just a technical feature — it is a regulatory requirement for consequential decisions under GDPR Article 22 and the EU AI Act. "The algorithm decided" is not a defensible position in front of a regulator.

Look For

Live demonstration of audit trail artefact, format appropriate for regulatory submission, human-readable rationale available for individual decisions.

Watch Out For

Description of the capability without a demonstration.

Red Flags

- No live demonstration available
- Audit artefact is a system log rather than a human-readable explanation
- Explainability described as a roadmap item

DOMAIN 02

Regulatory Compliance & Standards

Frameworks, certifications, and regulatory positioning · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**

Q2 · TIER 1 · BASIC

Regulatory Compliance — Q2

> Does your organisation follow AI-focused standards or frameworks such as ISO/IEC 42001 or the NIST AI RMF, and how does this system sit within that framework?

Why It Matters

Frameworks like ISO 42001 and NIST AI RMF are the clearest signal that a vendor has structured their governance around recognised standards rather than internal policy alone. A vendor who cannot place their product within a named framework has likely not done the work.

Look For

Named framework adopted with evidence, system mapped to specific framework controls, certification or assessment evidence available.

Watch Out For

"We follow industry best practices" without naming a framework or showing how the system maps to it.

Red Flags

- No named framework
- Compliance described as aspirational
- Cannot describe how the product maps to any published standard

Q14 · TIER 2 · INTERMEDIATE

Regulatory Compliance — Q14

> How does this product actively support our compliance obligations under applicable AI regulations and the sector-specific frameworks that govern our industry?

Why It Matters

As a deployer, you carry compliance obligations that the vendor's product either helps or hinders you to meet. A vendor who has not thought about your sector's regulatory environment will leave the compliance mapping entirely to you.

Look For

Sector-specific regulatory requirements identified, described controls that support your compliance posture, compliance documentation available for your records.

Watch Out For

"We comply with all applicable regulations" without identifying which ones or how the product supports your specific obligations.

Red Flags

- Vendor cannot name the regulatory frameworks relevant to your sector
- No compliance documentation available
- Compliance described as entirely your responsibility

Q24 · TIER 3 · COMPREHENSIVE

Regulatory Compliance — Q24

> How have you assessed this system's risk classification under applicable AI regulations across the jurisdictions where we operate, what documentation supports that classification, and how is it kept current as regulations evolve?

Why It Matters

EU AI Act classification determines the legal obligations that apply to both the vendor and your organisation as a deployer. A vendor who cannot produce a classification and supporting documentation either has not done the work or does not believe the regulation applies.

Look For

Written classification provided with rationale, risk management file available, technical documentation ready for regulatory inspection, update process described.

Watch Out For

"We are monitoring the regulatory landscape and will update our documentation as requirements become clearer."

Red Flags

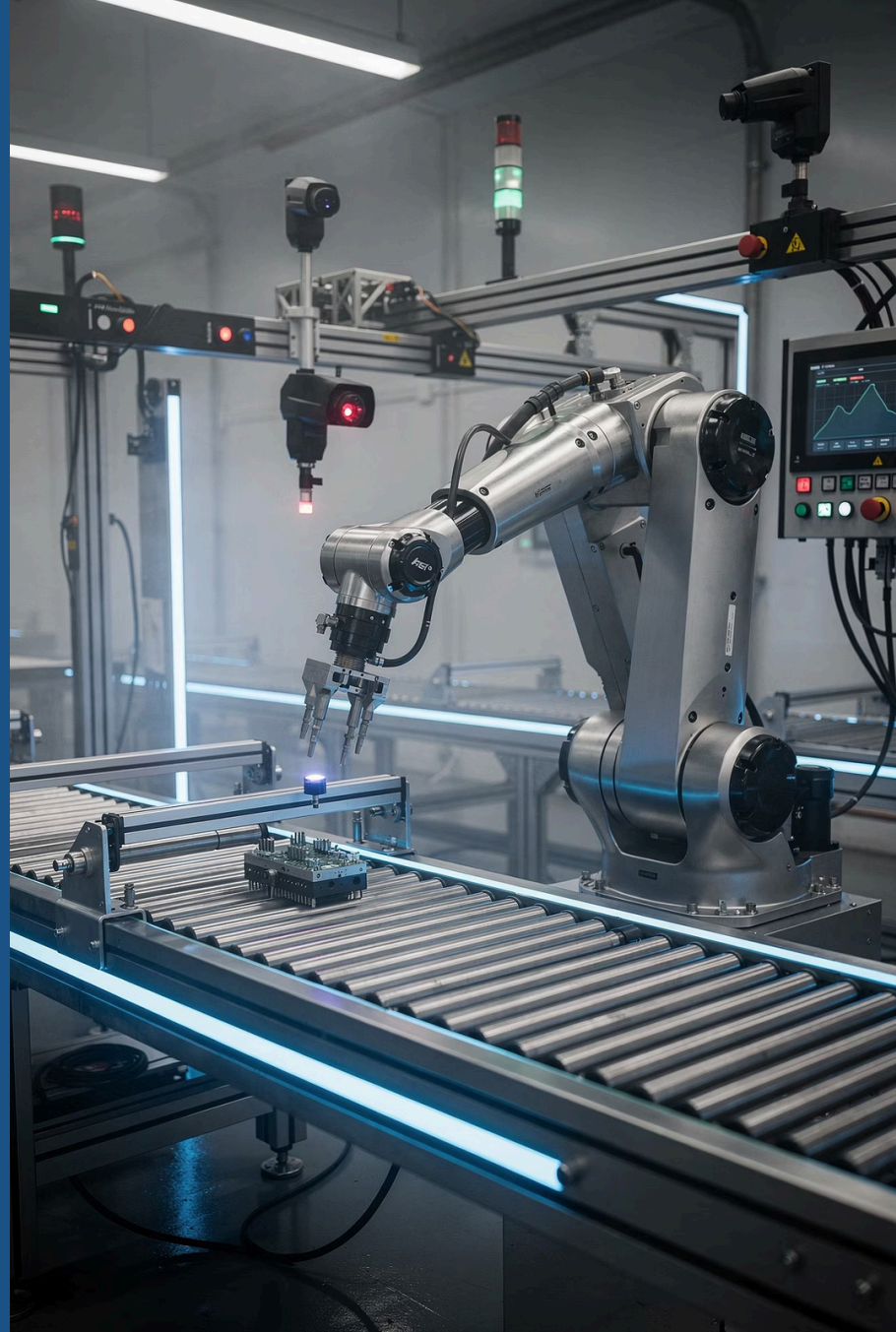
- No classification completed
- No risk management file
- Vendor does not distinguish between provider and deployer obligations under the Act

DOMAIN 03

AI Agent Use Case, Scope & Autonomy

Conditional Domain — Agentic Systems Only · 1 Tier 1 · 1 Tier 3

This domain applies only when the system involves autonomous agents, agentic workflows, or any capability to execute actions without a human approving each individual step (see VQ5).



Q3 · TIER 1 · BASIC · CONDITIONAL

AI Agent Autonomy — Q3

> What level of autonomy does this system operate at — which decisions does it make entirely on its own, which does it influence without final authority, and where is human judgement still required before an outcome is determined?

Why It Matters

The line between AI-assisted and AI-decided is the line between augmentation and accountability transfer. Knowing precisely where each type of decision falls before deployment defines your governance and oversight obligations for this system.

Look For

Clear taxonomy of decision types with autonomy level stated per category, human-in-the-loop checkpoints named, technical enforcement of those checkpoints described.

Watch Out For

"The system assists your team in making better decisions" without specifying which decisions are fully autonomous.

Red Flags

- Cannot distinguish between AI-assisted and AI-decided outcomes
- Autonomy scope not technically bounded
- Human checkpoints described in policy but not enforced in code

Q25 · TIER 3 · COMPREHENSIVE · CONDITIONAL

AI Agent Autonomy — Q25

> For this specific use case, at what points in the workflow is a human required to review, approve, or override the system's output before any action is taken — and how is that requirement technically enforced rather than left to user discretion?

Why It Matters

A human-in-the-loop described in policy but not enforced in code is not a control. For consequential or irreversible actions, the distinction between policy-level and technical enforcement defines your actual risk exposure.

Look For

Human review points named per workflow step, technical enforcement described, override capability confirmed with audit trail.

Watch Out For

"Users can always review outputs before acting on them" without describing enforcement.

Red Flags

- Human review described in policy only
- Technical enforcement not in place
- Users can bypass checkpoint without audit trail

DOMAIN 04

Model Performance & Explainability

Hallucination, drift detection, and decision traceability · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**

Q4 · TIER 1 · BASIC

Model Performance — Q4

> What is your process for detecting incorrect or hallucinated outputs from this system — what mechanisms identify them, how quickly, and what triggers an alert?

Why It Matters

Hallucination is not a theoretical risk. It is a documented failure mode of every large language model currently deployed. A vendor who cannot describe their detection process is relying on your users to catch mistakes, which is not a governance control.

Look For

Described detection mechanism with named triggers, alert routing to customer administrators, remediation SLA for confirmed hallucination events.

Watch Out For

"Our model is highly accurate and we monitor performance continuously."

Red Flags

- No systematic hallucination detection
- Customers expected to report errors through support
- No committed response time for confirmed incorrect outputs

Q15 · TIER 2 · INTERMEDIATE

Model Performance — Q15

> When this system's behaviour deviates from its validated performance baseline — in accuracy, output quality, or decision patterns — how are we informed, how quickly, and what information do you provide so we can assess the impact?

Why It Matters

Performance drift is one of the most common ways AI systems quietly fail. A vendor who only notifies you when a system goes down has a very different governance posture from one who proactively surfaces deviation from validated baselines.

Look For

Defined performance baseline documented at onboarding, deviation thresholds stated, notification triggered at threshold breach, impact assessment information provided with notification.

Watch Out For

"We continuously monitor performance and will contact you if there is an issue."

Red Flags

- No defined performance baseline
- No notification for drift events
- Performance reporting limited to uptime rather than output quality

Q26 · TIER 3 · COMPREHENSIVE

Model Performance — Q26

> Can you explain how this model arrived at a specific decision or output — and can you produce the supporting documentation, reasoning chain, and input factors that drove that result in a form that a business leader or regulator could review?

Why It Matters

Explainability at the individual decision level is required for high-risk AI systems under the EU AI Act and for automated decision-making under GDPR. A vendor who cannot produce this on request cannot support your compliance obligations.

Look For

Reasoning chain available per decision, input factors documented, output in business-readable format, format suitable for regulatory or legal submission.

Watch Out For

"The model is a black box but it is very accurate." Or explainability described as a roadmap item.

Red Flags

- No per-decision audit trail
- Explanation requires technical expertise to interpret
- Logs not retained long enough for regulatory purposes

DOMAIN 05

Bias, Fairness & Ethical AI

Testing methodology, independent review, and protected characteristics · **1 Tier 1 · 1 Tier 3**

Q5 · TIER 1 · BASIC

Bias & Fairness — Q5

> What specific methods do you use to measure bias in this system's outputs before it goes live, and can you share the metrics and thresholds that must be met before deployment is approved?

Why It Matters

Bias in AI systems creates legal exposure, reputational risk, and real harm to the people affected by the decisions. Understanding how the vendor detects and remediates bias before deployment is part of your own due diligence.

Look For

Described bias testing methodology, named protected characteristics tested, clear thresholds that gate deployment approval, results shareable under NDA.

Watch Out For

"We are committed to fairness" without methodology. Bias testing described as a one-time pre-launch activity.

Red Flags

- No ongoing bias monitoring
- Results unavailable or confidential
- Vendor shifts responsibility for bias outcomes to the customer

Q27 · TIER 3 · COMPREHENSIVE

Bias & Fairness — Q27

> Has this system undergone an independent ethical review or algorithmic impact assessment by a third party, and can you share the findings or a summary of outcomes?

Why It Matters

Self-assessed ethics is a low bar. An independent review conducted by a third party provides scrutiny that internal testing cannot replicate, and is increasingly expected by regulators and institutional procurement teams.

Look For

Independent review conducted, third-party reviewer named, findings shared or available under NDA, remediation steps documented.

Watch Out For

"We conduct internal ethics reviews" without independent verification.

Red Flags

- No independent review conducted
- Results of any assessment not available
- Vendor has no process for external ethical scrutiny

DOMAIN 06

Data Management Practices

Provenance, residency, retention, and deletion · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**

Q6 · TIER 1 · BASIC

Data Management — Q6

> How do you establish and document the origin of the data this model was trained on — and can you confirm that the sources, consent status, and licensing of that training data are fully traceable?

Why It Matters

Training data provenance is a legal and governance requirement. Unlicensed training data creates IP liability. Unconsented personal data creates regulatory exposure. A vendor who cannot trace their training data sources cannot give you a clean risk picture.

Look For

Training data sources documented, consent and licensing status confirmed per source, traceability available for audit, any gaps identified and described.

Watch Out For

"We use publicly available data and proprietary datasets" without specifics on consent or licensing.

Red Flags

- Training data sources undisclosed
- Consent status unknown
- Licensing not confirmed
- Vendor cannot support an IP or privacy audit of training data

Q16 · TIER 2 · INTERMEDIATE

Data Management — Q16

> Where is our data processed and stored, and can you provide a full data flow diagram showing every location where our data resides, including all sub-processors?

Why It Matters

Data residency is a legal obligation in many jurisdictions. If your data sits in a region your regulations prohibit, or passes through a sub-processor you were not aware of, you carry the compliance exposure regardless of what your vendor's marketing says.

Look For

Named storage regions, data flow diagram provided covering all sub-processors, cross-border transfer mechanisms documented, configurable data residency available.

Watch Out For

"Data is stored in the cloud" without named regions. Sub-processor list unavailable or described as case-by-case.

Red Flags

- Data storage location unknown or variable
- No data flow diagram available
- Sub-processor list restricted or not maintained

Q28 · TIER 3 · COMPREHENSIVE

Data Management — Q28

> What controls govern how long our data is retained within this system, what is the verified deletion process when we request removal, and can you confirm that deleted data is also purged from any model fine-tuning or training pipelines it may have entered?

Why It Matters

Right to erasure obligations under GDPR and equivalent regulations apply to AI vendors as processors. The fine-tuning pipeline question is the one most vendors are not prepared for.

Look For

Retention periods defined per data category, deletion process described with committed timeline, backup deletion included, fine-tuning data deletion confirmed, deletion confirmation mechanism provided.

Watch Out For

"We delete data upon contract termination" without specifics on backups or training data.

Red Flags

- Data used in model training cannot be fully deleted
- Backup deletion not included in standard terms
- No deletion confirmation mechanism

DOMAIN 07

Security, Robustness & Resilience

Access controls, adversarial testing, and AI Bill of Materials · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**

Q7 · TIER 1 · BASIC

Security & Resilience — Q7

> What access controls govern who can query this system and under what conditions, and can AI functionality be disabled or restricted at the user, group, or organisational level without requiring vendor involvement?

Why It Matters

Granular access control is a basic security requirement for any enterprise system. A vendor who requires their own involvement to adjust access creates an operational dependency your IT governance cannot accommodate.

Look For

Role-based access controls described, customer administrators can restrict or disable AI features without vendor involvement, audit log of access events maintained.

Watch Out For

"Access is managed through your SSO integration" without describing what can be restricted at the AI feature level.

Red Flags

- No granular AI feature controls
- Disabling AI functionality requires vendor-side action
- No audit log of access events available to customer administrators

Q17 · TIER 2 · INTERMEDIATE

Security & Resilience — Q17

> What technical controls do you have against prompt injection and adversarial inputs, and how frequently do you conduct red team testing on this system?

Why It Matters

Prompt injection is the primary attack vector for AI systems. Red team testing frequency tells you whether the vendor treats adversarial security as an ongoing practice or a launch activity.

Look For

Described runtime controls against prompt injection, named red team methodology with frequency, willingness to share summary findings.

Watch Out For

"Our model has safety training" without described runtime controls.

Red Flags

- No described prompt injection controls
- Red teaming never conducted or results entirely confidential
- Vendor unfamiliar with adversarial testing methodology

Q29 · TIER 3 · COMPREHENSIVE

Security & Resilience — Q29

> Which open-source components or models are used in this service, can you provide a current AI Bill of Materials, and how do you manage the security and licensing risks they introduce?

Why It Matters

AI systems built on open-source models can carry licence obligations that flow downstream to deployers. A vendor who cannot produce an AI Bill of Materials has not assessed this risk, and that gap becomes your risk.

Look For

AI Bill of Materials available and current, open-source licence review conducted, security patching process described.

Watch Out For

"We believe our use of open-source components is compliant with applicable licences."

Red Flags

- No AI Bill of Materials
- No open-source licence review conducted
- No IP indemnification for claims arising from open-source components

DOMAIN 08

Privacy Compliance & Legal Basis

Encryption, jurisdictional compliance, and model-layer data leakage · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**

Q8 · TIER 1 · BASIC

Privacy Compliance — Q8

> How is personal and sensitive information protected within this AI system — including controls over access, encryption, and retention?

Why It Matters

AI systems can surface, combine, and expose personal data in ways that conventional access controls were not designed to address. Knowing the specific controls in place is the minimum required before processing any personal data through this system.

Look For

Named encryption standards at rest and in transit, described access controls for personal data, retention periods defined per data category, data minimisation principles applied.

Watch Out For

"We comply with GDPR and all applicable privacy laws" without describing specific technical controls.

Red Flags

- No named encryption standard
- Personal data retention period undefined
- No access log for personal data processed by the AI

Q18 · TIER 2 · INTERMEDIATE

Privacy Compliance — Q18

> How does this system's design and operation ensure compliance with the privacy and data protection laws that apply in the jurisdictions where we operate — and what evidence can you provide that those requirements are actively met, not just acknowledged?

Why It Matters

Acknowledging a regulation is not the same as meeting it. A vendor who can describe specific design decisions made to satisfy GDPR, CCPA, or equivalent requirements, and can produce evidence, is a fundamentally different compliance partner.

Look For

Specific design decisions referenced to named regulatory requirements, evidence available (certifications, DPIAs, legal review documentation), DPA executable within your timeline.

Watch Out For

"We comply with GDPR and all applicable privacy laws" without described design decisions or evidence.

Red Flags

- Evidence requires follow-up
- Compliance described as the customer's responsibility to verify
- No named regulatory frameworks addressed in product design

Q30 · TIER 3 · COMPREHENSIVE

Privacy Compliance — Q30

> How do you ensure that personal and sensitive information processed by this system cannot be memorised, reconstructed, or surfaced through model outputs — and what technical controls and testing support that assurance?

Why It Matters

Large language models can memorise and reproduce training data, including personal information. A vendor who relies on access controls alone has not addressed the model-layer leakage risk that sits beneath them.

Look For

Technical controls described for preventing memorisation, differential privacy or equivalent applied where relevant, red team testing conducted for data extraction attacks.

Watch Out For

"Our model does not retain personal information" without describing technical controls or testing.

Red Flags

- No technical controls for model-layer data leakage
- No testing for data extraction attacks
- Assurance based on policy only

DOMAIN 09

Model Cards & Technical Documentation

Transparency artefacts and use-case validation · **1 Tier 1 · 1 Tier 2**

Q9 · TIER 1 · BASIC

Model Cards — Q9

> Do you publish model cards or equivalent documentation describing training data, intended use, known limitations, and performance characteristics?

Why It Matters

A model card is the baseline of technical transparency. Without it you are going in blind, with no informed basis for the decision, no baseline to measure drift against, and no documentation to show a regulator if asked what you knew before you deployed.

Look For

Model card or equivalent published covering training data provenance, intended use cases, known limitations, and performance characteristics. Updated when material changes occur.

Watch Out For

No model card. Or a capability overview framed as documentation that contains no technical substance.

Red Flags

- Cannot describe training data sources or known failure modes
- No documentation updated since initial release
- Vendor unfamiliar with model card as a concept

Q19 · TIER 2 · INTERMEDIATE

Model Cards — Q19

> Can you provide detailed technical documentation covering the evaluation methodology, datasets used for testing, known failure modes, performance thresholds, and the specific conditions under which this model was validated for our use case?

Why It Matters

A model card describes the product. Validation documentation describes whether it works for your specific use case. The gap between the two is where deployment risk lives.

Look For

Evaluation methodology described, test datasets named, known failure modes listed, performance thresholds stated, use-case-specific validation evidence available or offered under NDA.

Watch Out For

"We can share our internal benchmark results" without use-case-specific validation.

Red Flags

- No use-case-specific validation
- Evaluation methodology undisclosed
- Known failure modes not documented or not available to customers

DOMAIN 10

Incident Response & Continuous Monitoring

Detection, notification timelines, and post-incident review · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**

Q10 · TIER 1 · BASIC

Incident Response — Q10

> What monitoring capabilities are in place to detect abnormal AI system behaviour, adversarial inputs, or policy violations in real time, and do our administrators have direct visibility into those alerts?

Why It Matters

Being told about an incident after the fact is not the same as being able to see a problem developing in real time. If your only visibility comes through vendor-prepared reports, you are dependent on their judgement about what you need to know and when.

Look For

Customer-accessible monitoring available, covers system health, performance, and security signals, alert thresholds configurable by your administrators, demonstrated during onboarding.

Watch Out For

"We have internal monitoring and will notify you of significant events." Internal visibility is not customer visibility.

Red Flags

- No customer-facing monitoring access
- Visibility limited to periodic vendor reports
- Alert thresholds set and controlled entirely by the vendor

Q20 · TIER 2 · INTERMEDIATE

Incident Response — Q20

> How would we find out if something goes wrong with this system, what does your incident notification process look like in practice, and what are your committed timelines by incident type?

Why It Matters

How quickly you find out about a problem determines how quickly you can contain it. Committed timelines by incident severity, in writing, are what separate a governance control from a vague reassurance.

Look For

Incident classification described with named severity levels, committed notification timelines per severity, named contact for escalation, post-incident report committed.

Watch Out For

"We will notify you of any significant issues as soon as possible."

Red Flags

- No defined incident notification timeline
- Notification method informal
- No post-incident reporting commitment
- Severity classification not described

Q31 · TIER 3 · COMPREHENSIVE

Incident Response — Q31

> After an AI-specific incident is resolved, what structured post-incident review process do you follow, what documentation is produced, and how do you demonstrate to us that identified gaps have been closed?

Why It Matters

The quality of a vendor's post-incident review determines whether the same incident happens again. A vendor who resolves incidents without structured review and documented closure is repeating the same risk cycle.

Look For

Structured post-incident review process described with named outputs, root cause analysis committed, gap closure evidence provided, remediation timeline committed.

Watch Out For

"We review all incidents internally and take steps to prevent recurrence."

Red Flags

- No formal post-incident review process
- Root cause analysis not shared
- Gap closure not evidenced or committed

DOMAIN 11

Commercials, SLAs & Change Management

Uptime commitments, model update governance, and exit terms · **1 Tier 1 · 2 Tier 2**

Q11 · TIER 1 · BASIC

Commercials & SLAs — Q11

> What SLAs apply to this AI system, covering availability, uptime guarantees, and response time commitments for AI-specific components?

Why It Matters

Standard software SLAs often exclude AI-specific components, leaving a gap between what is committed and what you actually depend on. Understanding exactly what is and is not covered before you sign is the time to negotiate.

Look For

SLA covers AI-specific components by name, uptime commitment stated with penalty structure, degraded performance thresholds defined, AI-specific exclusions listed explicitly.

Watch Out For

"We offer 99.9% uptime on our platform" without confirming this covers AI inference and model response time.

Red Flags

- SLA explicitly excludes AI model performance
- No penalty structure for downtime
- No definition of degraded performance for AI-specific functions

Q21 · TIER 2 · INTERMEDIATE

Commercials & SLAs — Q21

> When you update, retrain, or replace an AI model underpinning this service, what is your change management process — how far in advance are we notified, what testing or validation must occur before rollout, and what recourse do we have if a model change degrades performance or breaks our use case?

Why It Matters

Model updates are the most common source of unexpected behaviour change in deployed AI systems. A vendor with a defined change management process, with advance notice and rollback capability, is a fundamentally different operational partner.

Look For

Advance notification period committed in contract, staged rollout process described, rollback capability confirmed, regression testing against customer use cases offered, recourse mechanism defined.

Watch Out For

"We test updates thoroughly before release" without describing notification or rollback options.

Red Flags

- No advance notification for model updates
- No rollback capability
- Customer has no recourse if a model change degrades their use case

Q32 · TIER 2 · INTERMEDIATE

Commercials & SLAs — Q32

> If we decide to exit, what specific artefacts do you provide to support migration — model outputs, configurations, audit logs — and do you offer defined exit assistance services with scope, timeframes, and rates?

Why It Matters

Exit friction is a strategic risk. A vendor who makes it difficult to leave has changed the commercial dynamics of your relationship from the moment you sign. Knowing the exit terms before you enter is basic procurement governance.

Look For

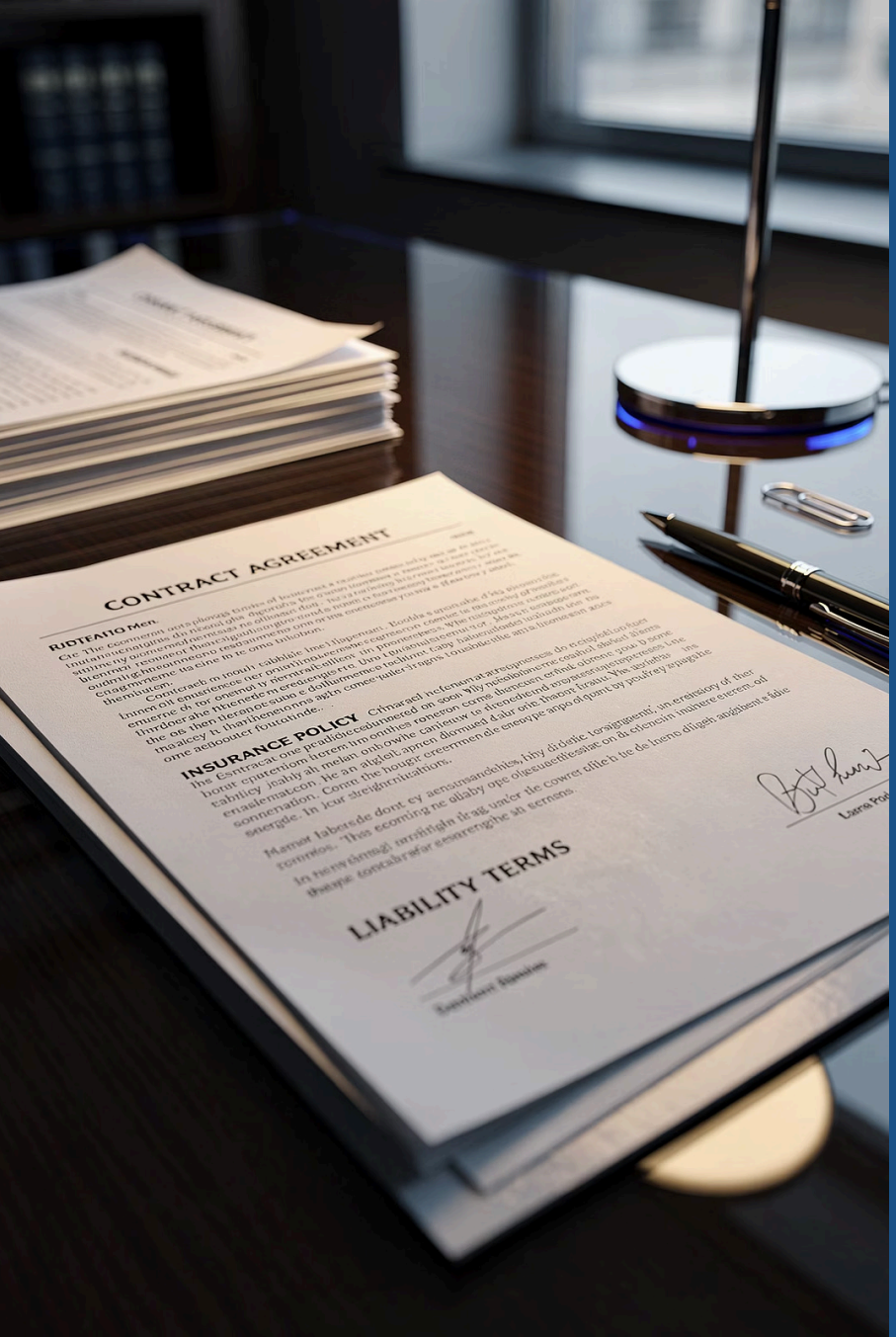
Named exit artefacts listed in contract, data export in portable format confirmed, audit log export included, exit assistance services described with scope and rates, migration period defined.

Watch Out For

"We will work with you to ensure a smooth transition" without contractual specifics.

Red Flags

- No defined exit artefacts
- Data export format proprietary
- No exit assistance committed
- Migration timeline not described



DOMAIN 12

Legal, Liability & Insurance

Contractual liability, AI-specific coverage, and financial ceiling · **1 Tier 1 · 1 Tier 2 · 1 Tier 3**

Q12 · TIER 1 · BASIC

Legal & Liability — Q12

> What liability limitations exist in your standard contracts for AI-specific failures or regulatory penalties?

Why It Matters

Most standard software contracts cap liability at fees paid — a figure that may bear no relationship to the harm an AI failure could cause. Understanding the contractual ceiling before deployment is the only time you have leverage to negotiate different terms.

Look For

Liability position clearly described, cap amount stated relative to realistic loss scenarios, AI-specific failures addressed separately from general software failures.

Watch Out For

"Our standard terms limit liability to fees paid in the prior twelve months" without discussing whether that covers an AI-specific incident scenario.

Red Flags

- Liability capped at a figure that would not cover a minor incident
- No acknowledgment of AI-specific failure modes in the contract
- Vendor cannot describe their own contractual position

Q22 · TIER 2 · INTERMEDIATE

Legal & Liability — Q22

> Do you carry insurance coverage for AI-related incidents, and what are the coverage limits and specific risk categories that policy covers?

Why It Matters

A vendor's liability position in your contract tells you what they are willing to accept. Their insurance coverage tells you whether they can pay if something goes wrong at scale.

Look For

Cyber and AI-specific insurance confirmed, coverage limit stated, named risk categories covered including data breach, model failure, and regulatory penalty exposure.

Watch Out For

"We carry comprehensive insurance" without coverage limits or named risk categories.

Red Flags

- No insurance for AI-specific incidents
- Coverage limited to infrastructure events
- Vendor cannot produce evidence of coverage on request

Q33 · TIER 3 · COMPREHENSIVE

Legal & Liability — Q33

> If this system causes a measurable business or legal problem for us, what is your contractual liability position, what does your standard contract actually commit to, and what is the realistic financial ceiling on any claim we could make?

Why It Matters

Liability for AI-caused harm varies enormously between vendors. Understanding exactly where the contract leaves you before something goes wrong is the only time you have leverage to negotiate different terms.

Look For

Liability position clearly described, cap amount stated relative to realistic loss scenarios, indemnification clauses cover AI-specific failures, vendor can walk through a hypothetical without deflecting to legal.

Watch Out For

"Our legal team would need to review any specific claim scenario." At a meeting, a vendor should know their own contract.

Red Flags

- Liability capped at a figure that would not cover a minor incident
- No indemnification for regulatory penalties
- Vendor cannot describe their own contractual position

Assessment Tracking Summary

Use this page to record the overall outcome of your vendor conversation. Transfer ratings from individual question blocks to produce a domain-by-domain view.

Domain	✓ Satisfactory	¤ Follow Up	□ Concern
AI Governance & Oversight			
Regulatory Compliance & Standards			
AI Agent Use Case, Scope & Autonomy			
Model Performance & Explainability			
Bias, Fairness & Ethical AI			
Data Management Practices			
Security, Robustness & Resilience			
Privacy Compliance & Legal Basis			
Model Cards & Technical Documentation			
Incident Response & Continuous Monitoring			
Commercials, SLAs & Change Management			
Legal, Liability & Insurance			
TOTAL			

Decision Framework

Use the outcome of your assessment tracking summary to determine the appropriate next step.

✓ All Satisfactory

Proceed to the next stage. Ensure all assurances given during the conversation are reflected as **contractual commitments** before anything is signed.

⌑ 1–3 Follow Ups, No Concerns

Continue the conversation or schedule a follow-up. **Resolve outstanding items** before moving to contract review.

× Any Concern Responses

Escalate before proceeding. Each concern requires resolution or a documented decision to discontinue.

×× Multiple Concerns

Pause. Present findings to relevant stakeholders before any further commitment is made.

⊗ This assessment covers vendor governance maturity in conversation. It does not replace a full written evidence review or a technical security assessment of the deployed solution. **Both are required before committing to a high-risk AI deployment.**

A Foundation for Lasting Partnership

The output of this process is a qualification decision: is this vendor worth pursuing further or not? A clean result across all 12 domains means the relationship is worth developing.

Next: Full Assessment

A structured conversation covering all 12 domains in depth — conducted after vendor selection but before contract signing. Full question bank coming soon as the next component of this toolkit.

Next: AI Security Assessment

A separate assessment covering architecture review, integration risks, data flows, access controls, and your own configuration decisions. A strong vendor does not automatically mean a secure deployment.

THIS IS PART OF DENNIS AH KING AI GOVERNANCE BLUEPRINT