

VIKRAM JHA - DEFENSIBLE AI RUBRIC

VENDOR AI-CLAIM TEARDOWN RUBRIC

Version 1.0 - 24 July 2026

<https://vikramjha.work/artifacts/vendor-ai-claim-teardown>

Eleven claims that appear on almost every AI vendor's page, what each would require to be true, and the question that settles it.

WHY IT EXISTS

Vendor diligence usually fails politely. The security questionnaire comes back complete, every box is ticked, and nobody has established what the words mean. These are the claims that recur, the specific thing each one is doing rhetorical work to avoid saying, and a single question that cannot be answered with marketing language. None of this is an accusation - most of these claims are true in some narrow sense. The problem is that the narrow sense is rarely the one the buyer heard.

WHAT IT CONTAINS

1. Eleven recurring claim patterns, with the gap between the claim and its usual meaning
2. What would have to be true for the claim to mean what a buyer assumes
3. One question per claim that cannot be answered with marketing language
4. The answer that should end the conversation
5. How to record the result so it survives into the contract rather than the sales call

THE CLAIMS

SOC 2 COMPLIANT

What it usually means: A Type I report, or a Type II whose scope excludes the AI service you are buying.

What it would take to be true: A current Type II whose system description names this service, with exceptions read rather than skipped.

The question that settles it: Send the Type II report and show me where this product sits in the system description.

EXPLAINABLE AI

What it usually means: Feature attributions such as SHAP on a model score.

What it would take to be true: A reconstruction of the decision the customer actually experienced, including retrieval and tool calls.

The question that settles it: Show me the explanation a customer would receive for one declined case.

HUMAN IN THE LOOP

What it usually means: A screen exists where a person can intervene.

What it would take to be true: Evidence of what the reviewer saw and how long they

had before deciding.

The question that settles it: What is your median reviewer dwell time, and what do they see?

YOUR DATA IS NEVER USED FOR TRAINING

What it usually means: Not used for training the shared foundation model.

What it would take to be true: No retention, no human review, no fine-tuning, no evaluation reuse - with retention windows stated.

The question that settles it: Where does my prompt and output go, for how long, and who can read it?

ENTERPRISE-GRADE GUARDRAILS

What it usually means: Prompt-level instructions and a content classifier.

What it would take to be true: Enforcement between the agent's intent and the system of record, capable of refusing the action.

The question that settles it: Show me a blocked action in your logs, not a blocked message.

FULL AUDIT TRAIL

What it usually means: Chat transcripts and API logs.

What it would take to be true: Immutable action records with authority, policy version, inputs, and model and prompt versions.

The question that settles it: Reconstruct why one action six months ago was permitted.

EU AI ACT READY

What it usually means: A blog post and a compliance page.

What it would take to be true: A stated role - provider or deployer - and the specific obligations accepted, with dates.

The question that settles it: Which role do you take under the Act, and which articles do you accept?

ZERO HALLUCINATION

What it usually means: Retrieval-augmented generation with citations.

What it would take to be true: A measured error rate on your data, with the measurement method disclosed.

The question that settles it: What is the residual rate, on what evaluation set, measured how?

GDPR COMPLIANT

What it usually means: A DPA template and EU hosting.

What it would take to be true: A named role, a lawful basis, sub-processor transparency and a working erasure path through the model estate.

The question that settles it: Delete one subject's data all the way through the estate and show me the evidence.

ISOLATED TENANT

What it usually means: Logical separation in a shared model deployment.

What it would take to be true: Named isolation boundary - dedicated inference, or shared with stated controls - and where it stops.

The question that settles it: Draw the boundary and tell me what crosses it.

MODEL AGNOSTIC

What it usually means: A provider abstraction layer.

What it would take to be true: Evidence that behavior, evaluations and controls survive a provider swap.

The question that settles it: When did you last switch a customer's underlying model, and what changed?

RECORDING THE RESULT

The point of the rubric is not to win the meeting. It is to convert an answer into something enforceable before signature.

ANSWERED WITH EVIDENCE

What to do with it: Reference the artifact in the contract, with a right to re-request it.

ANSWERED CREDIBLY, NO EVIDENCE YET

What to do with it: Make it a condition precedent with a date, not a roadmap item.

REFRAMED RATHER THAN ANSWERED

What to do with it: Record the reframing verbatim. It is usually the real scope.

REFUSED AS PROPRIETARY

What to do with it: Legitimate for some material. Convert it into an audit right and a notice obligation.

ANSWERED WITH A CERTIFICATION

What to do with it: Certifications describe a scope. Read the scope, not the badge.

HOW TO USE IT

1. Send the questions in writing before the demo. The demo is designed to answer different questions.
2. Accept artifacts, not assurances. A named report beats a confident answer.
3. Record verbatim answers, including reframings - the reframing is usually where the real scope is.
4. Convert each answer into a contract term, a condition precedent, or an accepted risk with an owner.
5. Re-run it at renewal. Most of these answers change when the underlying provider does.

Built from public standards and general practice. No client information appears anywhere in this document. Regulatory instruments move - check anything cited here

against its primary source on the day you rely on it.
Free to use and adapt; attribution welcome, not required.