
FEDERAL BUYER'S GUIDE

How to evaluate an AI vendor for a regulated environment.

Most AI evaluations test the demo. Very few test the things that decide whether a system can actually be authorized, operated, and eventually replaced. This guide is the list of questions we would want asked of us.

See our acquisition posture (<https://wagonerenterprise.com/acquisition.html>)

Review our compliance posture (<https://wagonerenterprise.com/proof.html>)

Written for program offices, Information System Security Officers (ISSOs), contracting officers, and AI governance leads. Vendor-neutral. Use it on us too. To save this guide as a PDF, use your browser print command (Ctrl+P on Windows, Cmd+P on macOS).

01

Deployment reality

Can it run where your data actually lives?

02

Authorization evidence

Who assembles the ATO package, and when?

03

Data rights

Where does your data go, and who trains on it?

04

Exit

What happens when you want to leave?

SECTION 01: DEPLOYMENT REALITY

“Air-gapped” is a claim, not a feature.

Many vendors describe a system as deployable on-premises when what they mean is that a thin client runs locally while inference, telemetry, or licensing still calls out to a commercial cloud. Ask questions that make the distinction unavoidable.

Ask: what breaks with no route to the internet?

Do not stop at “can it run offline.” Ask what specific functionality degrades. Licensing checks, model updates, telemetry, error reporting, and authentication are the usual dependencies. A vendor who has genuinely done this will answer immediately and specifically.

Ask: how does an update reach a disconnected enclave?

There should be a described, repeatable transfer process with signed bundles and integrity verification on the receiving side, not “we would work that out during implementation.”

Ask: what is the deployment artifact?

A container image with a Helm chart or equivalent is a real answer. “We will provision it for you” usually means a managed service that cannot cross into your enclave.

Ask: has this been run at your classification level?

If the answer is no, that is not automatically disqualifying, but it should change your pilot scope and your schedule assumptions. Be suspicious of a yes with no detail attached.

SECTION 02: AUTHORIZATION EVIDENCE

Ask who does the paperwork, and when it gets done.

The gap between “our product is secure” and “here is the evidence package your Authorizing Official (AO) needs” is usually measured in months of your staff's time. Find out early whose months those are.

QUESTIONS THAT REVEAL THE ANSWER

- ✓ What artifacts does the product generate on its own, without our staff assembling them?
- ✓ Do you produce an SBOM? In which format, and is it generated per build or per release?
- ✓ Can you show a control mapping to our baseline, or only to a generic one?
- ✓ Are findings delivered in a format our POA&M tracker can import?
- ✓ Is the evidence cryptographically bound to the artifact we actually deploy?
- ✓ Can evidence be re-verified independently, by us, without your involvement?

WHAT A STRONG ANSWER LOOKS LIKE

The vendor hands you a sample evidence package before you ask twice. It contains a manifest, content hashes, an SBOM in a standard format, normalized findings with ownership, and a mapping to named controls. You can verify the hashes yourself.

The vendor is also clear about what the evidence does *not* do: notably, that continuous evidence generation supports ongoing authorization but does not replace an initial ATO. A vendor claiming their tool “gets you an ATO” is overselling.

SECTION 03: DATA RIGHTS AND AI GOVERNANCE

Where your data goes is a contract question, not a technical one.

Get “we do not train on your data” in writing

Marketing pages say it; contracts often do not. Ask for the specific clause. Ask separately about telemetry, logs, prompts, and error reports. These are frequently carved out of a no-training commitment.

Ask for an AI Bill of Materials

You should be able to see what models are in the system, where they came from, what they were trained on at a provenance level, and what changes between versions. If the vendor cannot inventory their own AI components, they cannot help you govern yours.

Ask for model cards and data cards

These should describe intended use, known limitations, evaluation results, and failure modes. A model card that lists no limitations is a marketing document.

Ask what adversarial testing has been done

Prompt injection, data poisoning, and model extraction are the baseline. Ask for the assessment report, the test cases, and, most tellingly, what the testing found and what was done about it.

Confirm the governance framework being claimed is current

Federal AI policy has moved repeatedly. A vendor citing a superseded memo may simply have stale collateral, but check that their compliance work tracks the policy actually in force for your system type.

SECTION 04: LOCK-IN AND EXIT

Plan the exit during the evaluation, not during the dispute.

1

Ask for the exit path in writing

A written description of how you extract your data, configurations, and generated artifacts, in open formats, without vendor assistance.

2

Test the export during the pilot

Export everything once while you still have leverage. Confirm you can read it with tools you already own.

Check what the artifacts depend on

Evidence packages and policy configurations only the vendor's product can read are lock-in, whatever the contract calls them.

RED FLAGS

Signals worth slowing down for.

IN THE PITCH

- ✓ Accuracy or time-savings percentages with no methodology behind them
- ✓ A compliance certification implied rather than stated, or listed without scope
- ✓ “FedRAMP” used loosely; ask for the impact level and the authorization status
- ✓ A claim that the tool produces or guarantees an ATO
- ✓ Customer logos the vendor will not let you contact
- ✓ Blanket “patent pending” across an entire product line

IN THE EVALUATION

- ✓ Demo environments the vendor controls and will not let you drive
- ✓ Reluctance to support government-led independent testing
- ✓ No written limitations section anywhere in the documentation
- ✓ Security documentation that arrives only after contract award
- ✓ An architecture diagram that stops at the trust boundary
- ✓ Pricing that changes materially once air-gapped deployment is mentioned

SCORING

A simple evaluation matrix.

Score each vendor 0 to 3 per criterion: 0 no answer, 1 claim only, 2 documented, 3 demonstrated in your environment. Anything scoring 0 or 1 on the first four lines is not

ready for a regulated deployment regardless of how good the demo was.

Disconnected operation

Runs with no route to the internet; update path is defined, signed, and verifiable on receipt.

Weight: critical

Evidence generation

Produces SBOM, findings, control mappings, and a verifiable manifest with no work by your staff.

Weight: critical

Data rights

Contractual no-retention and no-training terms covering prompts, logs, and telemetry.

Weight: critical

Documented exit path

Open-format export of data, configuration, and artifacts, testable during the pilot.

Weight: critical

AI transparency artifacts

AI-BOM, model cards, data cards, and adversarial testing results with findings disclosed.

Weight: high

Interoperability

Integrates with existing SIEM, SOAR, ITSM, and DevSecOps tooling rather than replacing it.

Weight: high

Independent testability

Supports government-led evaluation, red teaming, and source review under agreement.

Weight: high

Stated limitations

Documentation says plainly what the product does not do and where human judgment is required.

Weight: moderate

We publish this guide knowing it will be used on us. That is the point. If a criterion here disqualifies us on your program, you have saved us both a procurement cycle. Score us the way you would score anyone else: nothing published on a vendor website, ours included, counts as more than a claim until you have seen the artifact behind it. Ask us for the evidence pack, the written exit path, the data rights clause, and the disconnected update procedure, and score what we actually hand you.