UNITOOLX

Enterprise AI Procurement Stress Test

Fictional sample assessment for “Northstar AI”

Purpose
Demonstrate the structure, evidence traceability and remediation style of a UniToolx deliverable. Northstar AI is fictional. Every policy, control, test result and finding in this document is invented for demonstration.

Sample version 1.0 · 18 September 2026
unitoolx.com · reports@unitoolx.com

EXECUTIVE SUMMARY

Northstar AI — procurement readiness at a glance

Northstar AI is a fictional B2B knowledge assistant that connects to enterprise document sources and routes prompts through third-party model providers. The hypothetical company is preparing for a large financial-services buyer review.

Result	Count	Meaning in this sample
Likely blocker	1	A material claim is broader than the evidence available to defend it.
Escalation	2	A buyer is likely to require clarification, evidence or a remediation commitment.
Hygiene	1	Unlikely to block alone, but weakens answer quality and increases follow-up.

Highest-friction issue

The public statement “Customer data is never used to train AI models” is not supported across every processing path represented by the wording. The fictional DPA covers Northstar’s own training activity, but the evidence set does not establish an equivalent binding commitment for one optional analytics path routed through a downstream provider.

What we would recommend first

- Normalize the training-use claim across marketing, privacy, DPA and product documentation.
- Document retention by data class and storage path instead of using one headline duration.
- Create an evidence record for model-provider changes and customer-impact review.
- Separate conventional web-security testing from AI-layer testing claims.

Important: This is a demonstration of report mechanics. The labels above are not legal conclusions, certifications, vulnerability ratings or claims about a real company.

SCOPE & EVIDENCE

What was reviewed in the fictional engagement

A real report starts by making scope explicit so “not found” is not confused with “does not exist.” In this sample, the following evidence is assumed to have been provided or publicly available.

Source	Example material	Status
Public website	Homepage, enterprise page, security page, FAQ	Reviewed
Privacy / contractual	Privacy notice, fictional DPA v3.2, terms	Reviewed
Subprocessors	Provider list dated 2026-08-30	Reviewed
Product docs	Admin guide, retention article, connector docs	Reviewed
Test environment	Two test tenants, admin + standard-user accounts	Limited checks
Internal policies	Model change procedure, incident runbook	Partially provided

Out of scope

Formal penetration testing, source-code review, legal interpretation, SOC 2 / ISO certification, production destructive testing, employee interviews, and independent confirmation from downstream vendors.

Assessment language

Supported = evidence matches claim within scope. **Partially supported** = evidence covers only part of the representation. **Unverified** = sufficient evidence was not available. **Inconsistent** = reviewed sources materially conflict. **Test-dependent** = documentary review cannot resolve the issue.

EVIDENCE MAP

Claim-to-evidence matrix

The matrix is the working core of the assessment. It shows what the vendor says, what was checked and where the chain breaks.

Claim	Evidence	Result	Buyer consequence
Customer data is never used for training.	Website; DPA §4.2; provider terms; analytics docs	**Partially supported**	Likely blocker / contractual follow-up
Data can be deleted on request.	Privacy policy; deletion runbook; admin docs	**Partially supported**	Escalation: logs/embeddings not scoped
Model changes are evaluated before release.	Internal procedure; release notes	**Unverified**	Escalation: no repeatable evidence record
The product is red-teamed.	Sales deck; pen-test executive summary	**Inconsistent**	Hygiene: AI-layer scope unclear
Tenant isolation is enforced.	Architecture diagram; two-tenant retrieval check	**Supported in limited test**	Answerable with scope caveat

Observation

The matrix does not attempt to compress these findings into a universal maturity score. Procurement impact depends on the buyer, data sensitivity, deployment model and contractual context.

FINDING UTX-01

Training-use claim is broader than the demonstrated commitment

LIKELY BLOCKER

Field	Assessment
Public claim	“Customer data is never used to train AI models.”
Evidence checked	Website, DPA §4.2, subprocessor list, model-provider terms, optional analytics documentation.
Result	**Partially supported.** The DPA prohibits Northstar from using customer content for its own training. The supplied evidence does not establish an equivalent commitment for one optional analytics path whose metadata is transmitted to a downstream provider.
Why a buyer cares	An absolute “never” statement implies all tiers, providers, features and telemetry paths. A reviewer can ask for the binding source that covers every path.
Likely follow-up	“Does the commitment cover every model/provider route, support workflow and optional analytics setting? Where is that commitment documented?”
Recommended action	Either narrow the public claim to the evidenced scope, or extend contractual/technical controls so the broader claim becomes accurate. Add a single evidence record listing each processing path and the applicable commitment.

Evidence note

This finding does **not** assert that downstream training occurs. The problem is that the reviewed evidence does not prove the absolute claim across the complete represented scope.

FINDING UTX-02

Retention language does not distinguish storage paths

ESCALATION

Field	Assessment
Representation	“Customer data is deleted within 30 days of account deletion.”
Evidence checked	Privacy policy, fictional deletion runbook, admin docs, architecture data-flow notes.
Result	**Partially supported.** Primary application records have a 30-day deletion process. The evidence does not define the corresponding treatment of operational logs, vector embeddings and backup copies.
Likely follow-up	“Does the 30-day commitment include logs, backups and embeddings? If not, provide the retention schedule by data class.”
Recommended action	Publish or maintain a retention matrix by data class and processing path. Use the same terminology in privacy, DPA and technical documentation.

Example remediation artifact

Data class	Primary store	Logs	Embeddings	Backups
Customer content	30d after deletion	Define	Define	Define

FINDING UTX-03

Model-change review exists as a statement, not as reproducible evidence

ESCALATION

Field	Assessment
Representation	“Model updates are tested before production rollout.”
Evidence checked	Internal procedure, sample release notes, product documentation, change tickets supplied for two releases.
Result	**Unverified as a repeatable control.** The procedure describes evaluation, but the supplied release records do not consistently capture model version, test set, acceptance criteria, approver and customer-impact decision.
Likely follow-up	“Show us how the last material model change was evaluated and approved.”
Recommended action	Create a lightweight model-change record with: affected product path, old/new model, test suite, acceptance criteria, material behavior change, risk owner, approval and customer-notification decision.

Minimum evidence record

A one-page change record can be stronger than a 20-page policy if it shows what happened for the actual release. The procurement goal is answerability and traceability, not paperwork volume.

FINDING UTX-04

“Red-teamed” claim does not match the supplied test scope

HYGIENE

Field	Assessment
Representation	Sales deck: “Northstar is independently red-teamed for enterprise AI threats.”
Evidence checked	Third-party penetration-test executive summary, sales deck, internal test notes.
Result	**Inconsistent.** The third-party report covers conventional web/API testing. Internal notes include limited prompt tests, but no independent AI-layer test scope was supplied.
Likely follow-up	“Which AI-specific threats were included, who performed the testing, when, and against which product version?”
Recommended action	Change the sales wording to match the evidence or commission an explicitly scoped AI-layer review. Keep web/API penetration testing and AI-behavior testing as separate evidence items.

Reference note

When relevant, AI-layer security questions can be structured using recognized references such as the OWASP GenAI / LLM risks. Use of a reference does not make the review an OWASP certification.

CONTROL CHECK

Example of a supported claim with an explicit scope

SUPPORTED

Not every assessment item should become a “finding.” Recording supported claims matters because it creates buyer-ready evidence and prevents future teams from re-investigating the same question.

Claim	“Users cannot retrieve documents they do not have source-system permission to access.”
Evidence	Architecture diagram + connector permission documentation + two-tenant/two-role test environment.
Check performed	Attempted retrieval of a document available to Tenant A/admin but not Tenant B/standard user across agreed test cases.
Observed result	The restricted document was not returned in the tested retrieval paths. Audit logs recorded the requesting identity and denied retrieval path.
Conclusion	**Supported in the supplied test environment and cases.** This is not a proof of universal absence of authorization defects.

Buyer-ready answer

“Retrieval is evaluated against source-system permissions at query time. We can provide the architecture note and the latest scoped verification record. The current verification covered [defined connectors / roles / version].”

REMEDIATION

Prioritized 14-day plan

The recommended sequence resolves the highest-friction evidence issue first, then creates maintainable records so the same gaps do not reappear in the next questionnaire.

Day	Action	Owner	Output
1–2	Normalize training-use scope across website, DPA and security FAQ.	Legal/Product	One approved statement + path matrix
2–4	Create retention matrix by data class and storage path.	Engineering/Privacy	Retention evidence record
4–7	Backfill model-change record for last two material releases.	ML/Product	Repeatable change template
7–10	Correct “red-teamed” wording or scope an AI-layer assessment.	Security/Marketing	Evidence-aligned security claim
10–14	Assemble buyer evidence index and assign evidence owners.	Security/Sales	Reusable procurement pack

Verification option

A remediation verification pass should retest only the claims and evidence that changed. It should not silently expand into a new full assessment unless scope is explicitly changed.

BUYER EVIDENCE PACK

Recommended index for future questionnaires

The final objective is not one good report. It is a compact system your team can keep current.

#	Evidence item	Owner / trigger
01	AI product architecture & data-flow overview	Engineering · architecture change
02	Training-use / provider-path matrix	Product/Legal · provider/feature change
03	Retention & deletion matrix	Privacy/Engineering · storage change
04	Model and subprocessor inventory	Product/Security · provider change
05	AI-layer test evidence	Security · material release/test cycle
06	Model-change records	ML/Product · every material model change
07	Incident classification & notification path	Security/Legal · annual/incident review
08	Approved buyer-facing answers + source links	Sales/Security · source changes

Operational rule

A buyer-facing answer should never be “owned” only by the salesperson who last filled in a spreadsheet. Material answers need an evidence owner and a change trigger.

METHODOLOGY & LIMITATIONS

How to interpret this sample

UniToolx assessments are evidence-readiness reviews. They can use recognized frameworks as references where useful, but they do not convert those references into certifications.

Reference	How it may be used
NIST AI RMF / NIST AI 600-1	Risk-management structure and generative-AI risk considerations.
OWASP GenAI / Top 10 for LLM Applications	Security-oriented questions and AI-application threat coverage.
EU AI Act / European Commission guidance	Context for documentation and value-chain questions where relevant; legal applicability remains for qualified counsel.

Limitations that belong in a real report

- Conclusions are limited to the evidence, access, product version and dates stated in scope.
- A documentary review cannot prove the absence of undisclosed behavior or vulnerabilities.
- A limited technical test demonstrates behavior only for the agreed cases and environment.
- “Unverified” is not equivalent to “false,” and “supported” is not a universal guarantee.
- Legal, regulatory and certification determinations require appropriately qualified professionals or bodies.

Contact

UniToolx
sales@unitoolx.com · research@unitoolx.com · reports@unitoolx.com
https://unitoolx.com

FICTIONAL SAMPLE — NOT AN ASSESSMENT OF A REAL COMPANY