Local AI and the test of sovereignty

A practical guide to security, authority and independence

Anwesh Rath
Founder & CTO and Chief Scientist, AI Standards

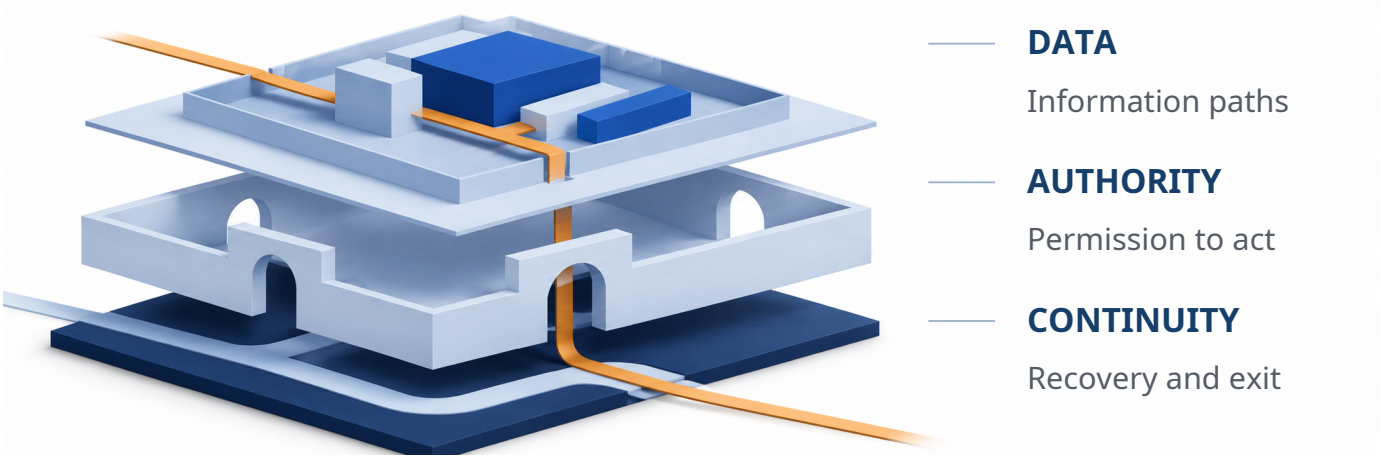
In a hypothetical case, an employee asks a local AI assistant for a supplier summary. It returns confidential acquisition details the employee was never allowed to read.

Nothing leaves the building. The information still reaches the wrong person.

Local processing can reduce exposure to external services and support disconnected work.

It does not establish who may access data, change the model, decrypt backups or authorize actions.

Each of those claims needs evidence.

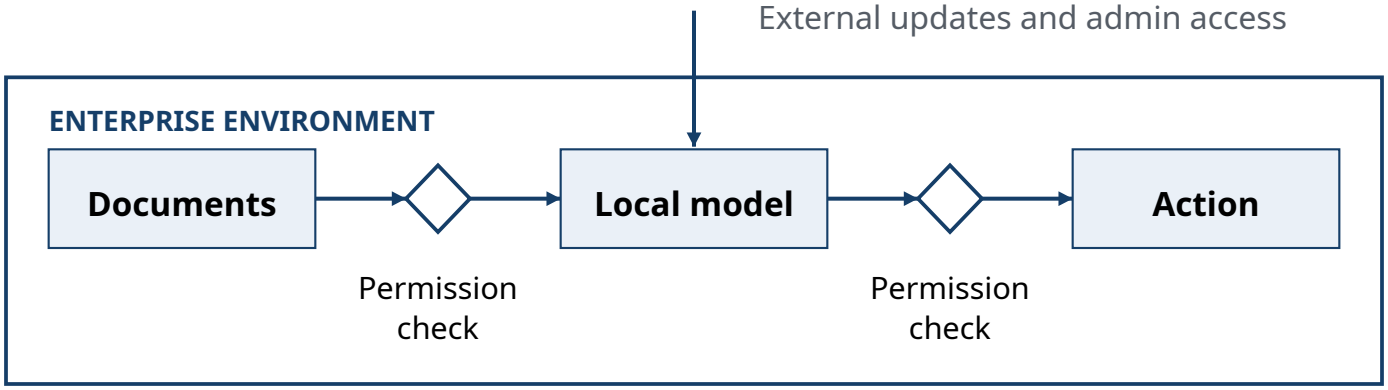


Conceptual view of data, authority and continuity. The amber route crosses the local footprint.

Residency locates processing. **Security** concerns resistance to threats. **Sovereignty** concerns enforceable control over the system and its dependencies. NIST rejects location or ownership as a basis for implicit trust. [1]

Follow control beyond the model

The system includes retrieval, identity, keys, tools, updates and support. Local inference is only one part of that system.



Track every copy

Trace documents through text extraction, embeddings, caches, prompts, memory, outputs, diagnostics and backups. Separate inference from training and evaluation. A ‘not used for training’ promise does not settle retention or support access. Embeddings are not automatically anonymous. [3, 8]

Keep permissions outside the model

Use distinct user, service and agent identities. Enforce source permissions before retrieval and tool permissions before execution. NIST’s October 2026 proposed tenets separate policy from reasoning and emphasize limited, revocable authority. This proposal is not a mandatory standard. [4,5]

Test the hidden dependencies

A local service may need external identity, a registry or a license token to restart. Test renewal, cold start, restore and migration. Identify who can approve updates, reject a supplier change and recover keys. Local does not necessarily mean offline or independently maintainable.

Limit every assurance claim

A file hash does not prove which weights ran. Encryption at rest does not prevent authorized runtime access to plaintext. A tamper-evident log does not guarantee completeness. Each claim needs a defined scope and trusted verifier. [6] Legal exposure also depends on entities and remote access, not location alone. [7]

Four tests before production

Use synthetic or authorized data. Define pass criteria first and record configuration, behavior, timestamps and exceptions. Approve only when critical controls pass; otherwise remediate and retest.

01 The wrong reader

Index a synthetic restricted record, then query it as an unauthorized user. Change access during a live session. Retrieval, caches and shared memory must respect the boundary. Keep the denied request and policy decision. Do not give protected text to the model and merely instruct it to hide it.

02 The hostile document

Insert a test instruction that asks the assistant to change a record outside its authority. The model may propose the action; an independent permission check must refuse it. Preserve the proposal, denial reason and tool result. Passing one test does not eliminate all prompt-injection risk.

03 The absent supplier

Remove a declared external dependency in a controlled environment. Attempt cold start, renewal and restore. The service should continue, degrade or stop safely as agreed. A silent external fallback fails a restricted workload. Record recovery times and any supplier intervention still needed.

04 The revoked session

Withdraw authority while work is in flight. Check active sessions, cached permissions, queued tool calls and new retrieval. Further actions must stop within the agreed window. Revocation cannot erase information already disclosed; already-fetched data requires separate containment and retention controls.

A pass supports a bounded claim under recorded conditions. Repeat affected tests after changes to models, data, tools, permissions or suppliers.

Choose control you can sustain

Choose the deployment by workload, not label. Local operation transfers patching, staffing and recovery work to the enterprise. A well-run managed service may be safer than a weak internal stack. Hybrid routing needs enforceable boundaries. The European Commission's 2026 framework examines legal, operational and supply-chain control beyond residency. [2] For each group below, identify an owner and collect dated evidence, not a verbal assurance.

Accountability

Record permitted uses, prohibited actions, named business and data owners, escalation routes and who can stop the service.

Data and access

Map every copy, retention rule and deletion path. Identify who can retrieve data, decrypt backups and open support sessions.

Models and infrastructure

Record model provenance, license rights, runtime versions and approved updates. Test host hardening, patching, egress and rollback.

Operations and recovery

Set capacity, latency and recovery targets. Restore models, policies, indexes, keys and logs, not just database backups.

Vendors and law

List entities, subprocessors and access locations. Verify audit rights, incident deadlines, license changes and cross-border access obligations.

Economics and exit

Price hardware, energy, staffing, updates and downtime. Demonstrate usable exports, migration to an alternative and deletion on supplier offboarding.

Mark each requirement **demonstrated, partial or unverified**. A critical control without evidence blocks approval; strengths elsewhere do not cancel it. Accept only explicit, time-limited exceptions with an owner, compensating controls and a retest date. Reassess after material system or supplier changes.

When ‘local’ meets the real world

Consider a hypothetical procurement evaluation for a local contract assistant. The model runs on company hardware, but the assessment uncovers three different kinds of dependency. Each finding calls for a different decision.

01 The index crosses a team boundary

The search index includes restricted deal documents. Require source-level access checks before retrieval, rebuild the affected index and repeat the wrong-reader test. A local network cannot substitute for document permissions.

02 Restart still needs the supplier

A license service is required after a cold start. Agree an acceptable outage window and demonstrate recovery. If disconnected operation is mandatory, this dependency must be removed or the claim rejected.

03 Support exports more than expected

Diagnostic bundles contain prompts and retrieved passages. Inspect an actual bundle, minimize its contents and control access, destination and retention. ‘Not used for training’ is not an answer to this finding.

Bring one workload and its evidence to AI Standards. Related filed provisional applications address client-controlled deployment, opt-in sharing and workload boundaries. A filing is not a grant or proof of implementation or security.

Primary sources Checked 5 November 2026

[1] NIST. SP 800-207, Zero Trust Architecture (2020).
[2] European Commission. Sovereign Cloud Framework explained (2026).
[3] NSA and partners. AI Data Security (2025).
[4] NIST. Agent identity management design tenets (proposal, October 2026).
[5] NCSC. Prompt injection is not SQL injection (2025).
[6] NCSC. Secure AI system development, secure deployment (2023).
[7] EDPB. Recommendations 01/2020 on transfer measures (final, 2021).
[8] Morris et al. Text Embeddings Reveal (Almost) As Much As Text (2023).