

CONTROL COSTS. KEEP ASSESSMENT WORK MOVING.

Cyber assurance without external AI token dependency

DyDox Govern and Auto-Govern | A briefing for Australian Government CIOs and CISOs

Could your agency continue AI-assisted assessment work if its external AI service reached a spending limit or became unavailable?

Joint guidance published by ASD's Australian Cyber Security Centre on 23 October 2026 identifies cost, capacity and availability risks from AI token consumption. It encourages organisations to consider local models where appropriate. [1]

Tools that use external frontier AI APIs can incur usage charges when processing large evidence sets and generating reports. They also introduce an external data-handling boundary. The cost and exposure depend on the service, workload and controls; API use is not itself data exfiltration.

The DyDox value

In its locally hosted configuration, DyDox Govern and Auto-Govern can support assessment workflows without external frontier AI APIs or external AI token charges. Local models still process tokens, but there is no external token allowance to purchase or exhaust. Supported offline operation also removes reliance on live connectivity to an AI provider.

Less processing time. More time for review.

11½ hours → 7 hours

DyDox reports reducing the readiness and assessment workflow time for a medium-sized system by optimising input prompts, prompt caching and output prompts. This is a reduction of 4½ hours, or approximately 39%.

This is a reported workflow result, not a universal completion guarantee. Scope, evidence quality, model and hardware affect results. Drafts still require human review and validation; formal risk acceptance and authorisation remain separate decisions. [2]

An illustrative hosting case: approximately \$500 per month

DyDox reports that a small GPU-backed Azure virtual machine can support 20 medium-sized, end-to-end readiness and assessment jobs per month at about \$500 in monthly hosting costs: approximately \$25 per job. This is hosting cost allocation, not the full cost of an assessment. [2]

Validate currency, VM specification, region, operating hours, utilisation and included charges before budgeting. Licensing, storage, support, taxes and reviewer effort may add costs. The example is supplied by DyDox, not a verified Azure quotation.

MAP THE DEPLOYMENT TO THE RISK

How DyDox addresses AI token-related risks

The table covers every risk category in the guidance’s “Risks related to AI token usage” section, with the two resource-consumption attacks shown separately. Responses apply to locally hosted DyDox; residual controls remain necessary. [1, 2]

Risk identified in guidance	DyDox response	Controls still required
Rising and unpredictable costs	No external per-token bill. Prompt optimisation and caching can reduce repeated processing.	Budget for hosting and licences; validate workload assumptions and total operating costs.
Reduced access to critical services	Local execution removes dependence on an external AI provider’s availability.	Prioritise jobs and reserve capacity. Maintain power, recovery arrangements and manual fallback.
Limited visibility and oversight	An agency-controlled environment provides a boundary for overseeing assessment work and data.	Validate local logging and reporting; monitor access, job activity, resource use and outcomes.
Overall performance	Optimised prompts and caching contributed to the reported 11½-to-7-hour result.	Benchmark output quality and throughput. Size GPU memory, storage and concurrency for the workload.
Token exhaustion	No external token quota, API rate limit or token budget is needed for supported local jobs.	Local context, memory and compute limits remain. Manage job size, queues and failures.
Denial-of-tokens attacks	There is no external token allocation to deplete; offline operation removes live remote API exposure.	Protect against local resource exhaustion through access controls, job limits and input checks.
Denial-of-wallet attacks	No external per-token charges can be driven up by abusive requests.	Restrict cloud provisioning and scaling; monitor infrastructure spend. Compute costs can still rise.

Data exposure is a related security consideration. Local processing avoids sending evidence to an external AI API. It does not remove risks from compromised devices, malicious evidence, removable media or mishandled outputs. Agency controls must cover the full information lifecycle.

KEEP EVIDENCE LOCAL. KEEP PEOPLE ACCOUNTABLE.

Controlled offline operation. A practical business case.

DyDox can run outside the production system being assessed on suitable GPU-backed end-user devices. After successful provisioning and validation of models, dependencies, licences and reference material, supported workflows can operate fully offline. The Azure hosting example is a separate deployment option; a disconnected device cannot rely on a remote Azure VM.

Potential uses include remote military bases, cruise ships, remote mining sites and locations where communications or location disclosure are restricted. Approved IT and OT evidence can be analysed locally to generate draft assessment reports. Required inspections, technical testing and specialist judgement remain part of the assessment.

1 Prepare and validate

Complete deployment, load approved evidence and test the workflow offline. Disable wireless and other communications interfaces and disconnect network cables. Validate performance on the intended device.

2 Assess within the approved boundary

Process evidence locally and retain appropriate local audit records. Human reviewers validate drafts. No external AI call is needed; wider location-disclosure risks still require operational controls.

3 Sanitise before reconnection

Export authorised outputs through an approved process. Remove containers and sanitise persistent volumes, reports, caches, logs and backups as required. Verify sanitisation before reconnecting.

Defence against malicious prompts and inputs

DyDox reports implementing security checks intended to improve resilience to prompt injection and input-poisoning attacks. These protections should be tested against representative malicious evidence and workflows. No quantified attack-tolerance result is asserted here; offline deployment and prompt checks do not make a system immune to attack. [2]

Co-create your monthly savings business case

Reach out to explore how much your organisation could save by moving repetitive document generation to DyDox while specialists review and validate the drafts. Compare monthly job volumes, current preparation effort, reviewer time, rework, hosting and licensing costs in a measured pilot. Quantify released staff capacity separately from cash savings.

REFERENCES & EVIDENCE

[1] [ASD ACSC: Managing AI token consumption for operational resilience](#), 23 October 2026. Source checked 25 October 2026. The guidance does not endorse DyDox.

[2] DyDox-supplied product, performance, security and hosting statements for this newsletter. Independently validated benchmarks, a test protocol and an Azure cost breakdown were not provided. Validate claims against the proposed deployment. No claim of zero footprint, immunity, guaranteed performance or absolute sovereignty is made.