PROTECT SERVICES. PRESERVE PUBLIC CONFIDENCE.

3 steps to prepare for rogue or agentic AI threats

If an AI agent tested your systems tonight, would your CIO or CISO have an accurate picture of their security risks tomorrow morning?

OpenAI acknowledged unauthorised access to Australian Government websites during internal model testing in June 2026. Its review found no evidence of medical-record access in the Medicare statistics incident. This supports preparing for unintended agent behaviour as well as deliberate AI-enabled attacks. [1]

1 Find and reduce exposure

Identify internet-facing services, APIs and legacy systems. Confirm owners, check access controls and exposed credentials, and prioritise fixes. Validate changes through authorised testing.

2 Limit access and watch for misuse

Give agency-operated agents only the access and tools they need. Require human approval for consequential actions. Monitor exposed services for unusual access and data extraction.

3 Practise your response

Rehearse an agent-enabled intrusion. Test containment, evidence preservation, supplier notification and executive escalation. Give each action an owner.

The DyDox value

DyDox can help teams analyse approved evidence, assess controls and prepare draft risk reports. This can reduce manual preparation and help reviewers identify gaps sooner. Reviewers validate findings; operational security teams remain responsible for detection and response.

Controlled deployment. DyDox can operate outside production on suitable controlled, sovereign or air-gapped GPU-backed infrastructure, without external frontier AI APIs or external AI token charges. Validate the configuration against agency requirements; infrastructure, licensing and support costs may apply.

Start with one exposed service: review its evidence, close priority gaps and rehearse escalation.

REFERENCES

[1] OpenAI: How we will do better for Australia (28 Sep; updated 4 Oct 2026)

[2] ASD: Using the cyber security framework (ISM, September 2026)

Sources checked 7 October 2026. Practical steps are editorial recommendations. DyDox capabilities reflect supplied product descriptions and require deployment validation.