

Governing the agent, not just the algorithm

By Brandon Voight, Fellow at Future Government Institute, and Director of Enterprise Information Management - OpenText Australia & New Zealand - submitted to AIIA Catalyst, October 2026

Australia's next AI governance challenge is already arriving. Organisations are moving beyond generative AI that drafts and summarises towards agentic systems that can retrieve information, invoke tools and take actions across enterprise environments. The opportunity is significant, but so is the change in risk. When AI can act, information governance and cyber security can no longer operate as separate disciplines.

The Australian Signals Directorate's September guidance on agentic AI harnesses is especially timely. It directs attention beyond the large language model to the software layer connecting that model with organisational data, tools and systems. This "harness" determines what an agent can see, what it can do, and which controls stand between a suggestion and an action. ASD highlights least-privilege access, secure design, monitoring, audit logging, human oversight for high-impact actions and clear accountability.

That is sound cyber security. It is also sound information governance.

An agent's authority should never exceed the authority of the person or process it supports. Yet many organisations still retain excessive, duplicated and poorly classified information, with permissions accumulated over years. Connecting an AI agent to that environment does not fix the weakness. It may accelerate it. The central governance question is therefore not simply, "Is the model safe?" It is, "What information can the agent access, what actions can it initiate, and what evidence will remain afterwards?"

Australian policy is converging around this need for evidence. The Commonwealth's Policy for the responsible use of AI in government requires accountable officials, internal AI use-case registers, staff training and impact assessment. The National Archives of Australia says AI-generated information is subject to the same Commonwealth records obligations as other government information, applying a risk-based approach under the Archives Act 1983. Inputs, prompts, source data and metadata may need to be retained where they provide context for a significant record or demonstrate how a decision was reached.

Privacy obligations are also becoming more explicit. From 10 December 2026, new Australian Privacy Principle 1 transparency requirements will require covered entities to disclose relevant uses of computer programs that make or inform decisions significantly affecting individuals' rights or interests. A privacy statement, however, is only the visible layer. Organisations must be capable of substantiating it through reliable records of data use, decision logic, approvals, human intervention and outcomes.

Leaders should now establish an agent governance register linking each use case to an accountable owner, approved data sources, identity and access controls, delegated actions, retention requirements, monitoring and an emergency stop mechanism. High-impact actions should require human authorisation, while logs must be protected as governance records rather than treated as disposable technical exhaust.

The organisations that scale agentic AI successfully will not be those with the fastest pilots. They will be those that can prove, continuously, that every agent acts within authority, uses trustworthy information and leaves an auditable trail. In the agentic era, governance is not paperwork around innovation. It is the control system that makes innovation safe to scale.