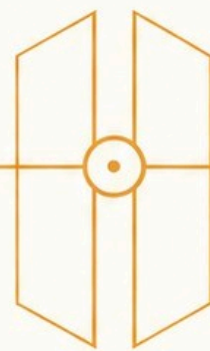


AI AGENT GOVERNANCE AUSTRALIA

THE REGULATORY
LANDSCAPE



COGNAONE.COM

SEPTEMBER 2026



■ REGULATORY REVIEW · SEPTEMBER 2026 · V3.2 AUSTRALIAN FOCUS

AI Agent Governance Australia: The 2026 Regulatory Landscape

Australian Focus: prudential standards, conduct regulators, cyber guidance and privacy, and what they already ask of AI agents

Daniel Mackle · CognaOne · cognaone.com

Australia has no AI Act, but agents are already regulated. This edition maps the Australian instruments that reach AI agents, from APRA's prudential standards and April 2026 AI letter to ASIC's licence obligations, the Five Eyes agentic AI guidance, privacy, ransomware reporting, critical infrastructure and scam prevention, and applies them to one worked example.

In Australia the question is not whether agents are regulated. It is which of several existing regimes an agent's actions fall under, and whether the organisation can show control under each of them.

CONTENTS

1. Australia regulates agents already, without an AI Act	8. Privacy: automated decisions and the data agents touch
2. How to read this edition	9. Scams, consumer law and government agencies
3. The Australian instrument map	10. Where the Australian obligations converge
4. The prudential core: CPS 230, CPS 234, CPS 220 and SPS 220	11. A worked example: a supplier payment agent
5. What APRA's April 2026 AI letter found	12. The Australian dates that matter
6. ASIC: licence obligations, enforcement and the May 2026 letter	13. Questions for the board and the accountable executive
7. Cyber security: the Five Eyes guidance written for agents	14. Conclusion
	About the author and CognaOne

References

1. Australia regulates agents already, without an AI Act

Australia has no AI Act. That is often read as a light-touch position. For an organisation deploying AI agents, it is closer to the opposite: the obligations that reach an agent are already in force, spread across prudential standards, licence conditions, privacy law, cyber security law and critical infrastructure rules, and regulators are now saying so in writing.

In the space of eight days in 2026, the Australian Prudential Regulation Authority (APRA) told every regulated entity that AI governance, risk management and assurance are not keeping pace with adoption¹, cyber agencies from Australia and its Five Eyes partners published joint guidance written for agentic AI^{2,3}, and the Australian Securities and Investments Commission (ASIC) told licensees that cyber resilience is a core licensing obligation, not an IT issue⁴. None of these created new law. Together they show how regulators and cyber agencies expect organisations to manage agent risk under the rules that already apply.

In Australia the question is not whether agents are regulated. It is which of several existing regimes an agent's actions fall under, and whether the organisation can show control under each of them.

The Government has also announced Australian Standards for AI, with legislation expected early next year. As announced, those standards focus on large data centres and AI training rather than on how organisations deploy agents⁵. For agent governance, the regimes described here are the ones that apply now.

This edition goes deeper than the Global Focus edition on the Australian instruments. It covers the APRA prudential standards and the April 2026 AI letter, ASIC's licence obligations and its May 2026 open letter, the Five Eyes agentic AI guidance, privacy, ransomware reporting, critical infrastructure, scam prevention and the rules that apply to Commonwealth agencies.

2. How to read this edition

The instruments carry different force. Prudential standards and statutes are binding on the entities within their scope. Regulator letters set supervisory expectations and signal enforcement priorities; they do not create new rules. Guidance from cyber agencies and the National AI Centre is voluntary, although supervisors increasingly point to it. We state the force of each instrument plainly.

Every factual statement about an instrument is referenced to the issuing body's own text wherever that text was accessible, and all sources were checked in September 2026. Where a primary source restricts automated access, we say so and identify the source used to confirm its content. Secondary sources are labelled as such in the references.

The mapping from obligations to engineering controls is our analysis and nothing here is legal advice. CogniaOne builds controls of the kind described in Section 10. Product descriptions are identified separately from regulatory findings.

Who this edition is for. Find your lane before the detail, because the regimes do not overlap neatly:

- **APRA-regulated entities** (banks and other ADIs, insurers, superannuation trustees): CPS 230 and CPS 234 apply; the risk management standard is CPS 220 for banks and insurers and SPS 220 for superannuation trustees; the Financial Accountability Regime applies (Section 4 and 5).
- **AFS licensees and market participants**, whether or not APRA-regulated: section 912A licence obligations and ASIC's expectations (Section 6).
- **Critical infrastructure responsible entities**: the SOCI Act risk management program (Section 7).
- **Banks, telecommunications providers and digital platforms**: the Scams Prevention Framework (Section 9).
- **Non-corporate Commonwealth entities** and their suppliers: the government AI policy (Section 9).
- **Broadly applicable, subject to scope conditions**: the Privacy Act for APP entities (the organisations and agencies the Act covers), the Australian Consumer Law for dealings with consumers, ransomware payment reporting for reporting business entities that make a payment or become aware of one, and the Five Eyes guidance as a voluntary technical baseline.

3. The Australian instrument map

INSTRUMENT	FORCE AND SCOPE	WHAT IT MEANS FOR AGENTS
APRA Prudential Standard CPS 230 Operational Risk Management ⁶	Binding on APRA-regulated entities. Commenced 1 July 2025; amendments and the end of the pre-existing contract transition on 1 July 2026 ^{7,8}	An agent that supports a critical operation sits inside the operational risk framework, the tolerance levels and the business continuity plan. An AI provider that the entity relies on for a critical operation is a service provider risk. Material operational risk incidents must be notified to APRA within 72 hours ⁶ .
APRA Prudential Standard CPS 234 Information Security ⁹	Binding on all APRA-regulated entities; in force since 1 July 2019	Agents are information assets with privileges. The entity must keep an information security capability commensurate with threats, including for assets managed by third parties, and notify APRA of material incidents within 72 hours and of material unremediated control weaknesses within 10 business days ⁹ .
APRA Prudential Standard CPS 220 Risk Management ¹⁰	Binding on authorised deposit-taking institutions (ADIs), general, life and private health insurers; in force since 1 July 2019	Agent risk has to sit within a Board-approved risk management framework, risk appetite statement and risk management strategy ¹⁰ .
APRA Prudential Standard SPS 220 Risk Management ¹¹	Binding on all registrable superannuation entity (RSE) licensees; CPS 220 does not apply to them	The superannuation equivalent: a risk management framework covering all material risks to the trustee's business operations ¹¹ .

INSTRUMENT	FORCE AND SCOPE	WHAT IT MEANS FOR AGENTS
Financial Accountability Regime (FAR) ¹²	Binding. Applies to ADIs from 15 March 2024 and to insurers and superannuation trustees from 15 March 2025; administered jointly by APRA and ASIC	Named accountable persons carry responsibility for the areas where agents operate. Delegating work to an agent does not delegate the accountability.
APRA letter to industry on artificial intelligence, 30 April 2026 ¹	Supervisory expectations for all APRA-regulated entities	Sets out observed weaknesses and minimum expectations on board literacy, AI inventories, human involvement in high-risk decisions, supplier concentration, identity management for AI agents, assurance and fallback for critical operations (Section 5).
Corporations Act 2001, section 912A general licence obligations, as applied in ASIC v FIIG Securities ¹³	Binding on Australian financial services (AFS) licensees	The Federal Court's first civil penalty for cyber security failures under the general licence obligations: controls must be demonstrably effective and proportionate to the business ^{4,13} . The case did not address AI agents; its application to them is our analysis.
ASIC open letter on cyber resilience and frontier AI, 8 May 2026 ⁴	Supervisory expectations for all licensees and market participants; must be tabled at the board and risk committees	Treats AI-accelerated cyber risk as a licensing matter and asks for specific steps, including access reviews, privilege reassessment and third-party risk management (Section 6).
ASIC Report 798, Beware the gap, 29 October 2024 ¹⁴	Review findings on 23 AFS and credit licensees	ASIC's baseline view of how licensees govern AI and manage consumer risk.
Privacy Act 1988, APP 1.7 to 1.9 ¹⁵ and OAIC guidance on commercially available AI products ¹⁶	Binding. Security and retention duties under APP 11 now; automated-decision transparency in privacy policies from 10 December 2026; guidance published 21 October 2024	Privacy obligations apply to personal information an agent takes in, holds and outputs. From December 2026, privacy policies must describe decisions made by a computer program, or where it does something substantially and directly related to making them, using personal information, that could significantly affect individuals (Section 8).
Careful adoption of agentic AI services, Five Eyes joint guidance, 1 May 2026 ^{2,3}	Voluntary guidance co-authored by the Australian Signals Directorate's Australian Cyber Security Centre (ASD's ACSC) and five partner agencies	The most specific control guidance for agents in any Australian source: distinct identities, least privilege, human approval, fail-safe defaults, logging (Section 7).
Cyber Security Act 2024, Part 3 ¹⁷	Binding on reporting business entities (above the turnover threshold, or responsible for a critical infrastructure asset). Payment reporting since 30 May 2025	If such an entity makes a ransomware or cyber extortion payment, or becomes aware one was made on its behalf, it must report to ASD within 72 hours, whether or not an agent was the entry point ¹⁷ .
Security of Critical Infrastructure Act 2018 (SOCI Act) ^{18,19}	Binding on responsible entities for critical infrastructure assets	Agents that touch a critical asset fall within the critical infrastructure risk management program, which covers cyber and information security, personnel, supply chain and physical and natural hazards, with an annual report.

INSTRUMENT	FORCE AND SCOPE	WHAT IT MEANS FOR AGENTS
Scams Prevention Framework ^{20,21}	Binding on banks, telecommunications providers and digital platforms, in stages: AFCA membership from 1 September 2026; most obligations from 31 March 2027; further obligations expected by the end of 2027 ²¹	Agents that initiate payments or communicate with customers sit inside a regulated entity's duty to prevent, detect, disrupt, report and respond to scams.
Australian Consumer Law, as reviewed by Treasury in October 2025 ²²	Binding on businesses dealing with consumers	Misleading or deceptive conduct and false representations are prohibited whatever technology produces them, and misleading conduct can be contravened without fault (Section 9).
Guidance for AI Adoption ^{23,24}	Voluntary. Published 21 October 2025	Six practices, from deciding who is accountable to maintaining human control.
Policy for the responsible use of AI in government, version 2.0 ^{25,26}	Mandatory for non-corporate Commonwealth entities; effective 15 December 2025, phased through 2026	Accountable officials, an internal register of AI use cases and impact assessments, the same building blocks any enterprise needs for agents.

Most binding instruments on this list were not written with AI agents in mind; the regulator letters and guidance show how they are expected to apply.

4. The prudential core: CPS 230, CPS 234, CPS 220 and SPS 220

For APRA-regulated entities, agent governance starts with prudential standards that were not written for AI and apply to it fully: CPS 230 and CPS 234 across all APRA-regulated industries, and a risk management standard (CPS 220 for banks and insurers, SPS 220 for superannuation trustees).

CPS 230: agents inside critical operations

CPS 230 requires an entity to manage its operational risks with effective controls, to continue delivering critical operations within tolerance levels through severe disruption with a credible business continuity plan, and to manage the risks that arise from service providers⁶. It commenced on 1 July 2025. The transition for pre-existing service-provider contracts ended at the earlier of renewal or 1 July 2026, when targeted amendments also took effect^{7,8}.

Three consequences follow for agents. First, an agent that performs or supports a step in a critical operation, such as payments, claims or member services, brings that step into the tolerance and continuity analysis: the entity has to know how the operation continues if the agent is unavailable or behaves wrongly. APRA's AI letter makes this explicit: where AI supports critical operations, credible fallback processes are required¹. Second, the model provider or agent platform behind that step is a service provider risk, and a material one if the operation depends on it. Third, an agent-driven incident that is likely to have a material financial impact or a material impact on a critical operation must be notified to APRA within 72 hours⁶.

CPS 234: agents as privileged information assets

CPS 234 requires an information security capability commensurate with vulnerabilities and threats, including for information assets managed by related parties and third parties, and places ultimate responsibility on the Board⁹. An agent holds credentials, reads data and calls systems; it is an information asset with privileges. Material information security incidents must be notified to APRA within 72 hours, and material control weaknesses the entity does not expect to remediate in time within 10 business days⁹.

APRA's AI letter supplies the observation that makes CPS 234 urgent for agents: identity and access management capabilities have not yet adjusted to non-human actors such as AI agents¹. An agent running under a shared service account, or under a person's credentials, is exactly the control weakness CPS 234 is designed to surface.

CPS 220, SPS 220 and the FAR: appetite and accountability

CPS 220 requires ADIs and insurers to maintain a Board-approved risk management framework, risk appetite statement and risk management strategy covering all material risks¹⁰. SPS 220 imposes the equivalent risk management framework requirement on superannuation trustees¹¹. If agents can take consequential actions, the risk appetite needs to say which actions, at what value or impact, and under what conditions. The Financial Accountability Regime, which applies to ADIs from 15 March 2024 and to insurers and superannuation trustees from 15 March 2025, assigns responsibility to named accountable persons¹². In our reading, when an agent acts inside an accountable person's area, the accountability does not move to the agent or its vendor.

5. What APRA's April 2026 AI letter found

APRA's letter of 30 April 2026 is addressed to all regulated entities and draws on a targeted supervisory engagement¹. It does not create new rules. It tells entities what their existing obligations mean for AI, and it is specific.

On boards, APRA observed strong interest in AI's benefits but many boards still developing the technical literacy to challenge AI risks, and an overreliance on vendor presentations without sufficient examination of risks such as unpredictable model behaviour and the impact on critical operations¹.

On governance, APRA's expectations include ownership and accountability across the AI lifecycle, from design through deployment, monitoring and decommissioning; an inventory of AI tooling and AI use cases; human involvement for high-risk decisions; and staff training on AI use, misuse and limitations¹.

On suppliers, APRA found some entities heavily dependent on a single provider for multiple AI use cases, few with tested exit and substitution strategies for critical AI providers, and contracts that often lagged practice, with limited provisions on audit rights, model updates and deviations, incident notification or changes to data handling¹.

On security, APRA found that identity and access management has not adjusted to AI agents, that the volume and speed of AI-assisted software development is straining change and release management controls, and that staff use of enterprise AI tools outside approved frameworks often relied on policy direction or detective, after-the-fact measures rather than enforceable preventative controls. Its expectations include controls over agentic and autonomous workflows and security testing of AI-generated code¹.

On assurance, APRA found internal audit and risk functions lacking the specialist skills and tools to assess AI, particularly where agentic behaviour, automated decision making or AI-assisted code generation were involved, so assurance lagged deployment. It expects recognised control frameworks and control libraries, change control for AI implementations, integrated assurance across cyber, data, model, resilience, privacy and conduct risk, and second-line and internal audit functions able to independently assess probabilistic models and agentic workflows¹.

APRA says it is finalising a forward plan for supervising AI risk, including prudential reviews, thematic activities and AI supplier engagement¹.

Read as a control specification, the letter asks for four things an agent programme rarely has on day one: an inventory, an identity for each agent, preventative controls rather than detective ones, and assurance that can test an agentic workflow.

6. ASIC: licence obligations, enforcement and the May 2026 letter

For AFS licensees, the anchor is section 912A of the Corporations Act 2001. In early 2026 the Federal Court ordered FIIG Securities to pay a \$2.5 million penalty and \$500,000 towards ASIC's costs, and to undertake a compliance programme with an independent expert, over cyber security failures. ASIC describes it as the first time the Court has imposed civil penalties for cyber security failures under the general AFS licensee obligations¹³. ASIC draws the lesson that cyber risk management controls must be demonstrably effective and proportionate to the size, nature and complexity of the business⁴. The case concerned cyber security under the general licence obligations; it did not decide what controls an AI agent must have. What follows for agents is our analysis.

ASIC's open letter of 8 May 2026 builds on that case. It frames frontier AI as materially increasing cyber risk, asks for a principles-based, model-agnostic response, and treats cyber resilience as a core licensing obligation⁴. The steps it asks for include reassessing cyber plans against the current threat environment, protecting critical assets, reviewing user access and reassessing privileges, patching promptly, maintaining incident response plans, actively managing third-party risks where services introduce concentration or systemic exposure, and using AI defensively where appropriate. Entities are required to table the letter at their ultimate board and risk governance committees⁴.

Neither the case nor the letter is about agents specifically. In our reading, both apply directly to them: an agent with broad privileges is an access-control question, an agent platform is a third-party dependency, and an agent that can be misdirected is part of the attack surface. ASIC's earlier

Report 798, which reviewed how 23 AFS and credit licensees use AI and govern the associated consumer risks, remains its baseline on AI governance in licensed firms¹⁴.

7. Cyber security: the Five Eyes guidance written for agents

On 1 May 2026, ASD's Australian Cyber Security Centre, the United States Cybersecurity and Infrastructure Security Agency and National Security Agency, and the national cyber centres of Canada, New Zealand and the United Kingdom published Careful adoption of agentic AI services^{2,3}. It is voluntary guidance. It is also the most specific statement any Australian government source has made about controlling agents.

The guidance groups agentic risk into five categories: privilege, design and configuration, behavioural, structural and accountability risks³. Its central recommendations translate almost directly into controls:

Scope. Align agentic AI risks with the organisation's existing security model, never grant agents broad or unrestricted access, especially to sensitive data or critical systems, and use agents only for low-risk and non-sensitive tasks².

Identity. Construct each agent as a distinct principal with a cryptographically anchored identity, authenticate agent-to-agent and agent-to-service calls, and keep a trusted registry that binds identities to authorised roles and is reconciled against the live set of agents³.

Privilege. Apply least privilege strictly, and prohibit agents from modifying their own privileges or initiating unapproved delegation without expiry and a recorded grant chain³.

Human control. Build human control points into the workflow, such as live monitoring and interruption, human approval for decision-making steps, and auditing and reversibility after execution, so that an agent approved for low-risk tasks cannot progress into higher-risk activity on its own³. The guidance calibrates approval to risk: multi-agent approval for moderate-stakes actions, with human approval added for high-stakes actions².

Safe defaults. Configure systems to fail safe by default, requiring agents to stop and escalate to human reviewers in uncertain scenarios, with containment, versioning and rollback to known-good behaviour³.

Graduated autonomy and records. Deploy in phases with progressively increasing access and autonomy, roll back when failures occur, restrict tools to an approved allow list, and log tool use in human-readable form³.

On accountability, the guidance observes that agentic architectures can obscure what caused a particular action, making accountability hard to trace³. Most of its recommendations aim at systems that do not depend on trusting the agent's judgement.

Two cyber statutes complete the picture. Under Part 3 of the Cyber Security Act 2024, a reporting business entity (broadly, a business above the turnover threshold, or an entity responsible for a critical infrastructure asset) that makes a ransomware or cyber extortion payment, or becomes aware one was made on its behalf, must report it to ASD within 72 hours; the obligation commenced on 30 May 2025, with an education-first period that ended on 31 December 2025¹⁷. Under the SOCI Act, responsible entities for critical infrastructure assets must maintain a risk management program covering cyber and information security, personnel, supply chain, and physical and natural hazards, and report on it annually; the Cyber and Infrastructure Security Centre (CISC) describes that report as a mechanism for boards to assess security risk management^{18,19,27}. An agent with access to a critical asset is a cyber, personnel-equivalent and supply chain question at the same time.

8. Privacy: automated decisions and the data agents touch

Two privacy developments matter for agents. The first is binding and dated. From 10 December 2026, new APP 1.7 to 1.9, inserted by the Privacy and Other Legislation Amendment Act 2024, require an entity's privacy policy to include certain information where it has arranged for a computer program, using personal information, to make a decision, or to do a thing substantially and directly related to making a decision, that could reasonably be expected to significantly affect an individual's rights or interests¹⁵. The OAIC states that the obligations apply to decisions made from that date, regardless of when the arrangement was made or when the personal information was collected¹⁵. An agent that approves, declines, prioritises or routes on the basis of personal information may bring an organisation within scope, and the organisation will need to know which agents do so. The OAIC has said it will publish detailed guidance on these obligations in 2026¹⁵.

The second applies now. APP 11 requires an entity to take reasonable steps to protect personal information it holds from misuse, interference and loss and from unauthorised access, modification or disclosure, and to destroy or de-identify it once it is no longer needed for a permitted purpose²⁸. For agents, that reaches memory stores, conversation logs, retrieval indexes and tool outputs: they hold personal information, need access controls like any other store, and need a retention rule.

The OAIC's guidance on commercially available AI products states that privacy obligations apply to personal information input into an AI system and to output data containing personal information, and recommends as best practice that organisations do not enter personal information, particularly sensitive information, into publicly available generative AI tools¹⁶. For agents, whose memory and retrieval can accumulate personal information across tasks, the input and output framing is the practical one.

When something does go wrong, the Notifiable Data Breaches scheme applies. An APP entity that suspects an eligible data breach, such as an agent exposing or disclosing personal information without authorisation, must take all reasonable steps to complete an assessment within 30 calendar days, and the OAIC treats that as a maximum. If the breach is likely to result in serious harm and

cannot be remediated, the entity must notify affected individuals and the OAIC as soon as practicable^{29,30}. Logs that show exactly what an agent accessed and sent are what make that assessment fast.

9. Scams, consumer law and government agencies

The Scams Prevention Framework requires businesses in designated sectors to take reasonable steps to prevent, detect, disrupt, respond to and report scams, with penalties of up to \$50 million²⁰. Banks, telecommunications providers and digital platforms are the first designated sectors. It applies in stages: covered entities have had to be members of the Australian Financial Complaints Authority (AFCA) since 1 September 2026, most obligations commence on 31 March 2027, and further obligations are expected by the end of 2027, with the codes and rules still being finalised²¹. The framework is not about AI. It matters for agents in two directions: agents deployed by banks to detect and disrupt scams, and customer-side or bank-side agents that initiate payments and could be manipulated into sending money where they should not.

For customer-facing agents, the Australian Consumer Law is the nearer constraint. Treasury's October 2025 review of AI and the Australian Consumer Law notes that the prohibitions on misleading or deceptive conduct and false or misleading representations apply regardless of the technology used, and that misleading conduct can be contravened without fault²². An agent that describes products, prices or eligibility inaccurately exposes the business even if no person wrote the words.

For Commonwealth agencies, the Policy for the responsible use of AI in government version 2.0 took effect on 15 December 2025 for non-corporate Commonwealth entities. Each in-scope use case must have a clearly assigned accountable officer and be recorded in an internal register, and AI impact assessments are mandated; the first new mandatory requirement began on 15 June 2026 and the remaining requirements take effect in December 2026^{25,26}. Suppliers to government should expect the same structure in their contracts.

10. Where the Australian obligations converge

The instruments come from different regulators and different decades. For an organisation running agents, they converge on the same small set of controls. The mapping below is our analysis.

CONTROL QUESTION	WHAT AUSTRALIAN SOURCES POINT TO	INSTRUMENTS
Which agents exist, what do they do, and who owns each?	An inventory of AI tooling and use cases with lifecycle ownership; a registry of agent identities bound to roles; accountable owners per use case	APRA AI letter, Five Eyes guidance, DTA policy, FAR ^{1,3,12,26}
Does each agent have its own identity and only the access it needs?	A distinct, authenticated identity per agent; least privilege; no self-modification of privileges; access reviews	Five Eyes guidance, CPS 234, APRA AI letter, ASIC letter ^{1,3,4,9}

CONTROL QUESTION	WHAT AUSTRALIAN SOURCES POINT TO	INSTRUMENTS
Is a consequential action checked before it executes?	Preventative rather than detective controls; human approval at defined high-impact checkpoints, with monitoring or other controls for lower-risk actions; fail-safe stop and escalate when uncertain	APRA AI letter, Five Eyes guidance, CPS 220 risk appetite ^{1,3,10}
Does the operation survive the agent failing?	Tolerance levels, fallback processes and tested exit from critical AI providers	CPS 230, APRA AI letter ^{1,6}
Can the organisation reconstruct and report what happened in time?	Attributable logs of agent actions and tool use; notification within 72 hours to APRA or ASD where thresholds are met	Five Eyes guidance, CPS 230, CPS 234, Cyber Security Act ^{3,6,9,17}
Are personal information and automated decisions handled lawfully?	Knowing which agents make or substantially support significant decisions; controlling what personal information agents take in and produce	Privacy Act APP 1.7 to 1.9, OAIC guidance ^{15,16}
Can assurance test an agentic workflow?	Second-line and internal audit capability to assess agentic workflows; change control for AI; continuous monitoring proportionate to criticality	APRA AI letter ¹

Product description. CognaOne is building controls of this kind: an agent and AI asset register, controls held with their regulatory provenance, including CPS 230, a deterministic action gate and decision records. The Global Focus edition sets out that design in Section 6 of that edition.

11. A worked example: a supplier payment agent

The instrument map is easier to use when it is applied to one workflow. Take an accounts-payable agent at an APRA-regulated bank. It reads supplier invoices, matches them to purchase orders and proposes payments. This is a higher-consequence design example: it goes beyond the Five Eyes guidance's cautious recommendation to use agents only for low-risk, non-sensitive tasks², and shows what the controls have to carry once an organisation chooses to go past that line. Each step is marked as required by an instrument, expected by a supervisor, recommended in guidance, or our design response.

STEP	WHAT HAPPENS	BASIS	TYPE
Scope	The agent is registered with its own identity and a named owner. It may propose payments to verified suppliers up to a set limit. It cannot change supplier bank details or approve its own payments.	Distinct agent identity and least privilege ³ ; inventory and ownership ¹ ; information security capability for information assets ⁹	Guidance; supervisory expectation; requirement (capability), with the specific design ours
Proposed action	An invoice for \$48,500 arrives. The agent proposes payment to the supplier's bank account on file.		
State check	The account on file was changed yesterday after an inbound email, and differs from the last verified details. The agent's own plan treats the change as routine.	Fail safe: stop and escalate to a human when uncertain ³ ; preventative rather than detective controls ¹	Guidance; supervisory expectation

STEP	WHAT HAPPENS	BASIS	TYPE
Decision	The action gate holds the payment because the target account is contested and the change came from an unverified channel. It does not matter how confident the agent is.	Risk appetite for payment actions set by the Board ¹⁰	Requirement (framework); the gate is our design
Approval	A person sees the exact proposed payment, the conflicting account details and their sources, verifies with the supplier by phone, and approves or declines.	Human approval at high-impact checkpoints ^{2,3}	Guidance
Record	A decision record links the request, agent identity, policy version, the state it relied on, the approver and the outcome.	Logging of agent actions and tool use ³ ; ability to notify within 72 hours if an incident follows ^{6,9}	Guidance; requirement (notification)
Personal information	The supplier is a sole trader, so the agent's memory and logs hold personal information. Access is restricted and a retention rule applies; if the agent were to expose that information, the data breach assessment clock would start.	APP 11 security and destruction duties ²⁸ ; Notifiable Data Breaches scheme ²⁹	Requirement
Fallback	If the agent platform is unavailable, the manual payment process runs within the operation's tolerance levels, and the provider dependency is recorded.	Critical operations and service providers ⁶ ; fallback where AI supports critical operations ¹	Requirement where payments are a critical operation; supervisory expectation

No single instrument requires this exact design. Together they make it hard to justify a design without these steps.

12. The Australian dates that matter

These are the dates that create or change an obligation, plus the supervisory statements that tell regulated entities how existing obligations apply.

Already in effect

DATE	INSTRUMENT	MILESTONE
15 Mar 2024	FAR	Applies to ADIs; insurers and superannuation trustees from 15 Mar 2025 ¹²
30 May 2025	Cyber Security Act 2024	Ransomware payment reporting commences ¹⁷
1 Jul 2025	APRA CPS 230	Commences ⁷
15 Dec 2025	DTA AI policy v2.0	Takes effect for non-corporate Commonwealth entities ²⁶
30 Apr 2026	APRA	Letter to industry on AI ¹

DATE	INSTRUMENT	MILESTONE
1 May 2026	Five Eyes	Careful adoption of agentic AI services ²
8 May 2026	ASIC	Open letter on cyber resilience and frontier AI ⁴
15 Jun 2026	DTA AI policy v2.0	First new mandatory requirement begins ²⁶
1 Jul 2026	APRA CPS 230	Amendments commence; pre-existing contract transition ends ^{7,8}
1 Sep 2026	Scams Prevention Framework	AFCA membership required for covered entities ²¹

Scheduled

DATE	INSTRUMENT	MILESTONE
10 Dec 2026	Privacy Act	APP 1.7 to 1.9 automated-decision transparency commences ¹⁵
Dec 2026	DTA AI policy v2.0	Remaining mandatory requirements take effect ²⁶
31 Mar 2027	Scams Prevention Framework	Most obligations commence ²¹

Expected, dates may move

SIGNALLED TIMING	INSTRUMENT	WHAT IS EXPECTED
During 2026	OAIC	Detailed guidance on APP 1.7 to 1.9 ¹⁵
Early 2027	Australian Government	Legislation for Australian Standards for AI, focused on large data centres and AI training ⁵
By end of 2027	Scams Prevention Framework	Further obligations ²¹
Not announced	APRA	Forward plan for AI supervision: prudential reviews, thematic activities, AI supplier engagement ¹

13. Questions for the board and the accountable executive

1. Which agents touch a critical operation under CPS 230, and what is the fallback if the agent or its provider fails?
2. Does every agent have its own identity in our access management, with least privilege, and can any agent change its own privileges?
3. Which agent actions are blocked or held before execution, rather than detected afterwards?
4. Which AI providers are we dependent on for more than one use case, and have we tested exit and substitution?
5. Can second-line risk and internal audit test an agentic workflow end to end today?

6. Which agents make or substantially support decisions that could significantly affect individuals, and will our privacy policy describe them from 10 December 2026?
 7. If an agent is compromised or misdirected, can we reconstruct what it did quickly enough to meet 72-hour notification obligations and to assess a possible data breach well inside 30 days?
 8. Which accountable person under the FAR owns each agent's area of operation?
- An organisation that cannot answer these should narrow its agents' action space before expanding their autonomy.

14. Conclusion

Australia's approach is spread across regulators, but it is consistent. APRA supplies the operational and information security framework and has now said where AI breaks it. ASIC supplies the licence obligation and a court decision that makes proportionate, demonstrably effective controls enforceable. The cyber agencies supply the technical control set, written specifically for agents. Privacy law adds a dated transparency obligation for automated decisions. None of it waits for an AI Act.

The obligations already exist. What is missing in most organisations is the evidence that each agent's actions sit inside them.

About the author and CognaOne

Daniel Mackle is the founder of CognaOne. CognaOne is building the active governance platform for AI in regulated financial services: an inventory of AI systems and the agents connected to them, controls held with their regulatory provenance, an action gate that allows, holds or declines proposed actions and returns the same decision on the same facts and policy, and decision records created at the moment of authorisation. This review was written to stand on its sources, not on the product, and every factual statement in it is referenced to the issuing body wherever that body's text was accessible.

AI use: research and drafting for this review were supported by AI tools. Multiple models were used for research and to cross-check facts, and every factual statement was verified by the author against the cited sources. The author is responsible for the content.

References