

Across industries

AI Agents and Digital Trust

Building trust in AI agents across the enterprise

AI agents are moving from recommendations to action, creating new opportunities to automate work, improve decisions and reshape how organisations operate. But as agents gain greater autonomy, a fundamental question emerges: can you verify what an agent is authorised to do? Recent reports from US banks illustrate the challenge. The reporting has not established confirmed customer losses, but it shows how agent access can outpace the controls needed to verify delegated authority.

The broader challenge for every organisation: **agent access can outpace the controls needed to verify who or what is acting, and with what authority?**

48%

of Australian organisations identify AI reliability and maturity as a leading barrier to greater AI-agent autonomy in security operations.

47%

point to accountability, explainability and liability concerns.

83%

of Australian organisations expect their cyber budget to increase, with AI a top investment priority.

Why this matters in Australia

Agents are a new kind of identity

APRA has observed that identity and access management has not yet adjusted to non-human actors such as AI agents. The same gap applies outside financial services: an authenticated user or application session does not establish which agent is acting, for whom, or within what limits.

AI is making fraud cheaper

Fabricated documents, synthetic identities and automated submissions are already testing Australian lenders and insurers. Any organisation that accepts applications, claims, invoices or account changes online faces the same pressure.

Expectations are rising

The Government's Guidance for AI Adoption calls for risk management that accounts for agentic AI and for telling people when they are interacting with AI. From 10 December 2026, privacy policies must explain where computer programs make or inform significant decisions.

Source: PwC 2027 Global Digital Trust Insights: Australian insights

What is changing, and what to do differently

1 **The risk changes as agents move from recommendations to autonomous action.**

Agents can now affect sensitive data, business processes and customer interactions. Unintended or excessive actions create risk, not only malicious ones. For example: changing a business record, initiating a purchase or accessing sensitive data.

2 **Security and governance must account for agent identity and delegated authority.**

A valid login does not, by itself, establish the agent's identity, its delegated authority or who is accountable across systems.

3 **Govern the agents you deploy. Protect against the agents you do not control.**

Know and govern the agents you deploy, decide your stance on external agents, set boundaries, enforce permissions, monitor activity and keep the ability to intervene.

Where agents meet your controls

Enterprise applications

Agents embedded in ERP, CRM and productivity platforms acting with inherited permissions.

Data access

Agents reaching sensitive or privileged data stores, including through low-code tools built by business users.

Autonomous transactions

Agents initiating purchases, changing business records or approving workflow steps.

Customer and supplier channels

External agents lodging applications, invoices, claims or account changes at scale.

From verified access to verified authority

Applies to agents built in-house, provided by third parties, embedded in software or acting across systems.

Govern the agents you deploy

Many of the agents in your environment are ones you build, buy or switch on inside platforms you already use. Treat each one as a new kind of identity, with the same discipline you apply to people and privileged systems.

Know what you run

Inventory agents you build, buy or receive embedded in vendor platforms, including those business users build with low-code tools.

Give each its own identity

A distinct, verifiable identity for each agent, not shared or inherited human credentials.

Scope its authority

Permissions fit the task, are time-bound, revocable and enforced outside the agent, with approval for high-impact actions.

Own and test it

An accountable owner for each agent, testing for unintended or out-of-scope actions, and risk-based human oversight.

Decide your stance on the agents you do not control

Baseline your exposure to external agents, legitimate or not, that interact with your customers, suppliers and systems, and agree a risk appetite: is this a practice you condone?

If you do not support it

Make it explicit in terms of use and policy. Put preventative and detective controls in place to identify non-human sessions, block or challenge them, and respond.

If you do support it

Set guardrails for acceptable use, such as permitted actions, rate limits and throttling. Make liability clear for users, agent providers and you.

Illustrative target-state control pattern

Authority established

A person or the organisation authorises the task.

Agent identified

Each agent has a verified, distinct identity.

Delegation scoped

Permissions fit the task, are time-bound and revocable.

Action enforced

Policy checks each action in applications, data and workflows.

Activity attributed

Actions are monitored and attributable; access can be suspended or revoked.

Four control questions

1 **Identity and delegated authority**

Can we verify which agent is acting, for whom, and with what delegated authority?

2 **Policy and action-level enforcement**

Are permissions scoped to the task, time-bound and enforced outside the agent, with approval for high-risk actions?

3 **Monitoring, intervention and accountability**

Can we detect deviations, intervene or revoke, and reconstruct who or what acted, under which authority?

4 **Incident containment and management**

Do we have a clear process to follow if an agent causes harm? Is it clear how we will contain it, notify impacted parties and regulators, and meet our regulatory obligations?

Illustrative example: an AI agent initiates a purchase across several applications. Each step may be authorised without anyone verifying that the complete transaction sits within the agent's delegated authority.

Illustrative control pattern only. It is not a prescribed sequence or an industry standard.



Building the foundations for trusted agentic AI

Reduce near-term exposure first, then build the capabilities that make trusted agent access sustainable.

Immediate priorities

Understand agent exposure.
Inventory internal, embedded and third-party agents, including those business users build with low-code tools. Map the sensitive systems, data and workflows they reach. Agree your stance and risk appetite.

Close critical authorisation gaps.
Review identity, MFA, permission and access controls for agents. Where authority cannot be established, restrict or add approval for high-impact actions. Assign owners.

Strengthen visibility and detection.
Can security teams tell agent from human activity, and trace, stop or reverse an agent-executed transaction across systems?

Longer-term capabilities

Trusted agent identity and delegated authority.
Extend non-human identity management, least privilege and revocation to agents.

Governance tied to enforceable controls.
Policies, testing and accountable owners across the agent lifecycle, including third-party and embedded AI.

Continuous monitoring and intervention.
Maintain visibility and enforce controls as agents interact across applications, workflows and other agents.

Test agents for unintended actions.
Probe behaviour under ambiguous, conflicting, manipulated or out-of-scope instructions; set risk-based human oversight.

Three questions for your leadership team

1 **The agents you deploy**
Are we deploying and operating agents we can trust, with clear accountability?

2 **What agents are allowed to do**
Can we independently verify and enforce what agents may do?

3 **When something goes wrong**
Can we detect and respond to agent risk, knowing detection is not authorisation?



Next step: understand your agent exposure and control gaps

Identify the agents across your environment, those you deploy and those you do not control, agree your stance on external agents, assess identity and authorisation controls, and set priorities for secure adoption.

Quick check: where do you stand?

Tick what is true for your organisation today. Anything you cannot yet tick is a priority to act on.

- ☐ We have an inventory of the agents operating in our environment, including embedded and low-code agents
- ☐ We have a written position on external agents interacting with our systems
- ☐ Agents have distinct identities, not shared or inherited human credentials
- ☐ Agent permissions are scoped to the task, time-bound and revocable
- ☐ We can tell agent activity from human activity across applications
- ☐ We can reconstruct who or what acted, under which authority
- ☐ We have a clear process to contain an agent incident and notify impacted parties and regulators

References

1. AI Trust
2. How to build an agent-friendly world
3. 2027 Global Digital Trust Insights: Australia
4. Agentic AI controls across platforms
5. Axios, 22 Oct 2026 (anonymous sources)
6. APRA, Letter to Industry on AI, 30 Apr 2026
7. AUSTRAC, Operation Claw, 19 Aug 2026
8. DISR, Guidance for AI Adoption
9. OAIC, AI and ADM transparency, 30 Sep 2026
10. ASD, Implementing MFA

Contacts



Nicola Costello
Partner, AI Trust Leader
nicola.costello@au.pwc.com



Peter Malan
Partner, Cybersecurity and Privacy
peter.malan@au.pwc.com



Robert Di Pietro
Partner, Cybersecurity and Privacy
robert.di.pietro@au.pwc.com



Jon Benson
Partner, Cybersecurity and Privacy
jon.benson@au.pwc.com



Penny Dunn
Partner, Advisory
penny.dunn@au.pwc.com

© 2026 PricewaterhouseCoopers. All rights reserved. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.