

Australian financial services

AI Agents and Digital Trust

Building trust in AI agents across financial services

> Identity and access management capabilities have not yet adjusted to nonhuman actors such as AI agents.
>
> **APRA, April 2026**

For banks, lenders, insurers, superannuation trustees and payments providers

AI agents are moving beyond helping customers find information to taking actions on their behalf. For financial institutions, that creates a new trust question: can you identify the agent, verify the authority it has been given and confirm that each action reflects the customer's intent? Recent reports from US banks illustrate how third-party agents may access customer accounts using customer-provided credentials. The reporting has not established confirmed customer losses.

The question it raises applies equally here: **can you tell whether an action was taken by your customer or by an agent acting for them, and was that action authorised?**

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.

Why this matters in Australia

Regulators have set the bar

APRA expects MFA or equivalent controls for high-risk activities and privileged access, and enforceable preventative controls rather than reliance on policy or after-the-fact detection. An agent holding a customer's credentials can pass the login. MFA shows a person was present at login, not that a person is directing each action that follows.

It goes well beyond transactions

AI is making fraud cheaper to produce. AUSTRAC's Fintel Alliance identified potentially hundreds of millions of dollars in suspected fraudulent home loans across 10 major banks, supported by falsified documents. Insurers report AI-generated documents in claims. Agents can submit applications and claims at a scale and speed manual checks were not built for.

Obligations are arriving

Under the Scams Prevention Framework, banks must prevent, detect, disrupt, respond to and report scams, with most obligations commencing 31 March 2027. Consumer Data Right action initiation will create a sanctioned route for accredited parties to act for consumers, sharpening the line between permitted and unpermitted agent access.

Sources: PwC 2027 Global Digital Trust Insights: Australian insights; APRA letter to industry on artificial intelligence (30 April 2026)

Where agents meet your controls

Banking and lending

Your own agents in servicing, operations and lending, and customers' agents in digital banking, payees and transfers; AI-assisted loan applications built on fabricated documents.

Insurance

Your own claims and underwriting agents, and external agents lodging or progressing claims; synthetic images, invoices and identities supporting claims.

Superannuation

Agents in administration and member services, and external agents changing member details, requesting withdrawals, rollovers or investment switches.

Payments

Agent-initiated payments and merchant checkout; scam payments authorised through a compromised or manipulated agent.

What is changing, and what to do differently

1 **Legitimate agents can create risk the customer never intended.**

Permission to access an account does not authorise every action that follows. Exposure: financial loss, customer harm, fraud and disputes, unclear accountability.

2 **Verifying access is not the same as verifying delegated authority.**

You can recognise an authenticated customer without identifying the agent, confirming what was delegated, or knowing whether an action matches the customer's intent.

3 **Govern the agents you own. Verify and authorise the agents you do not.**

Know and govern the agents you deploy, decide your stance on external agents, verify agent identity, enforce delegated authority, monitor high-risk activity and keep the ability to stop or revoke.

From verified access to verified authority

Authorisation should be enforced independently of the agent. Monitoring supports it but does not replace it.

Govern the agents you deploy

Institutions are deploying their own agents across operations, customer service, claims, lending and member services, and buying software with agents embedded. APRA's observation that identity and access management has not yet adjusted to non-human actors applies to these agents as much as to anyone else's.

Know what you run

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

Give each its own identity

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

Scope its authority

Permissions fit the task, are time-bound, revocable and enforced outside the agent, with approval for actions that move money or change details.

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, and agree a risk appetite: is this a practice you condone?

If you do not support it

Make it explicit in customer terms 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, value limits and throttling. Make liability clear for the customer, the agent provider and you.

Illustrative target-state control pattern

Principal established

The customer consents to delegation, or an accountable owner authorises your own agent.

Agent identified

The agent, yours or a third party's, is independently identified as non-human.

Authority delegated

Scope, duration and conditions are set and revocable.

Action authorised

Policy and transaction risk are checked on each request, with step-up for high-risk actions.

Activity attributed

Actions are monitored and attributable; authority can be revoked.

Four control questions

1 **Identity and delegated authority**

Can you tell a human from an agent, independently identify the agent, whether yours or a third party's, and verify what the customer, member or accountable owner authorised?

2 **Policy and transaction enforcement**

Can activity be limited by action, account, value, duration and risk, with added approval for high-risk actions?

3 **Monitoring, intervention and accountability**

Can you detect deviations and fabricated submissions, revoke access and reconstruct what happened?

4 **Incident containment and management**

Do you have a clear process to follow if an agent causes harm? Is it clear how you will contain it, notify impacted customers, members and regulators, and meet your regulatory obligations?

Illustrative control pattern only. It is not a prescribed sequence, an industry standard or a reconstruction of any reported event



Preparing for trusted agent access

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

Immediate priorities

Understand agent exposure.

Inventory the agents you deploy, including those embedded in vendor platforms, and identify third-party agents reaching customer accounts, digital channels, APIs, claims and member services, including agents using shared credentials. Agree your stance and risk appetite for external agents.

Close critical authorisation gaps.

Test MFA and step-up for high-risk activities against agent use. Where authority cannot be established, restrict or add approval for actions that move money or change details. Assign owners.

Strengthen visibility and detection.

Can fraud, scams and security teams distinguish agent from customer activity, spot AI-generated documents, and trace, stop or reverse agent actions?

Longer-term capabilities

Trusted agent identity and delegated authority.

Identify agents, scope time-bound permissions, verify delegation and revoke on demand, ready for Consumer Data Right action initiation.

Governance tied to enforceable controls.

Policies, testing and accountable owners across the agent lifecycle, linked to transaction-level controls and your accountability map.

Continuous monitoring and intervention.

Risk-based monitoring that correlates identity, behaviour and transaction signals, aligned to scam and fraud response.

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, authorisation and fraud 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 we deploy, including those embedded in vendor platforms, each with a distinct identity and an accountable owner
- We have a written position on whether customers or members may use external agents with their credentials
- We can distinguish agent activity from human activity in digital channels
- MFA or step-up applies to high-risk actions even within an authenticated session
- We can detect AI-generated documents in applications and claims
- We can revoke an agent's access and reconstruct what it did
- We have a clear process to contain an agent incident and notify impacted customers, members and regulators
- Liability for agent-initiated actions is clear in our customer terms

References

1. AI Trust
2. Scaling AI risk management and governance
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. APRA, Critical Authentication Controls, 10 Jun 2025
8. APRA, Use of MFA, 26 May 2023
9. AUSTRAC, Operation Claw, 19 Aug 2026
10. Insurance Business, 7 Oct 2026 (Gallagher Bassett 2026 claims study)
11. ASIC, Scams Prevention Framework
12. Treasury Laws Amendment (Consumer Data Right) Bill

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, Data & AI Governance
jon.benson@au.pwc.com



Penny Dunn
Partner, Fraud & Scams
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. PWC201315340