

AI Regulation Is No Longer a Horizon Risk: It Is a Prudential, Conduct, Privacy and Resilience Issue

protiviti®
Global Business Consulting



Artificial intelligence (AI) is moving rapidly from experimentation to execution. For APRA-regulated entities (including banks, insurers, superannuation trustees) and material service providers that support them, AI is no longer confined to innovation labs. Increasingly, it is embedded in claims triage, fraud and scam detection, customer interaction, credit and operational decision support, and technology and cyber services such as software development, testing automation, and threat detection and monitoring.

APRA's letter to the industry in April 2026[1] makes the prudential message clear: AI offers productivity and customer benefits, but governance, risk management, assurance and operational resilience practices are not keeping pace with adoption. ASIC reached a similar conclusion in financial services and credit: some licensees are adopting AI faster than their governance arrangements are being updated, creating a potential governance gap.[2]

The practical implication is not that Australia has suddenly introduced an "AI Act" for financial services. It has not. The more important point is that existing obligations are already being applied to AI. APRA's principle-based prudential framework is technology-and vendor-agnostic. APRA expects regulated entities to manage AI through risk appetite, oversight, accountability and appropriate controls. ASIC has also emphasised that current licensee obligations, consumer protection laws and directors' duties are technology neutral and apply to AI use.

For boards and C-suite leaders, the question is therefore not whether a new AI law has arrived. Rather, it is whether existing governance, risk, privacy, cyber and resilience frameworks can evidence control over AI at the speed it is being adopted.

Although Australia does not yet have a dedicated AI Act, recent Government announcements regarding AI standards,[3] and consumer safety priorities[4] along with enacted privacy amendments should put both corporate and non-corporate entities on alert. A clear theme is emerging: The benefits of AI must be pursued in a way that safeguards human interests, identity and privacy. Ethical and responsible AI use, supported by sound risk

[1] APRA Letter to Industry on Artificial Intelligence.
[2] ASIC: Beware the gap: Governance arrangements in the face of AI innovation
[3] https://www.pm.gov.au/media/ai-australias-interests.
[4] AI Consumer safety priorities | Our ministers – Attorney-General's portfolio

management and governance, must be built in from the outset and not treated as an afterthought.

In August 2026[5], APRA and ASIC reinforced this message through joint industry roundtable findings, calling on financial-sector entities to move from awareness to action and strengthening the focus on tested response arrangements, critical dependencies, concentration risk and decision-making at AI-enabled threat speed.

Governing AI risk through existing prudential standards and frameworks

APRA's 2026 letter[6] observed that many boards remain in the process of developing the technical literacy required to provide effective challenge on AI risk, and that it expects boards to maintain sufficient AI understanding to set strategic direction, oversee an AI strategy aligned to risk appetite and tolerance, and receive monitoring and reporting that includes third-party dependencies and resilience triggers. This aligns with APRA's broader governance review,[7] which identifies AI, cyber risk and geopolitical disruption as emerging risks that test board direction-setting, risk appetite, challenge and oversight.

AI should therefore be anchored to existing prudential standards rather than run as a disconnected innovation workstream. CPS 230 requires entities to identify, assess and manage operational risks, maintain critical operations through severe disruption, and manage risks arising from service providers. CPS 234 requires entities to maintain information security capability commensurate with vulnerabilities and threats, implement controls, test and assure those controls, and manage information assets held by related parties or third parties. AI cuts across all of these areas — operational, technology, data and cyber risk, change management, supplier risk, continuity, and the ability to stay within tolerance.

The critical shift in approach and mindset is that AI risk is not "just another technology risk." AI systems can be (extremely) probabilistic, adaptive, embedded in broader platforms, and reliant on upstream foundation models or third/fourth parties. In

The critical shift is that AI risk is not "just another technology risk."

[5] Insights from the APRA-ASIC Industry Roundtables | APRA
[6] APRA Letter to Industry on Artificial Intelligence (AI)
[7] Governance Review Consultation Paper

its 2026 letter, APRA found gaps in post-deployment and model behaviour monitoring, change management, and decommissioning of AI capabilities. The agency expects entities to establish consistent governance arrangements that include frameworks and reporting lines, lifecycle ownership, an inventory of AI tooling and use cases, human involvement in high-risk decisions, and staff training on AI use, misuse and limitations. This is an operating model issue, not only a technology control issue.

The cyber threat environment has changed — the basics must now work faster

AI is also changing the cyber risk equation. APRA has observed that AI increases attack pathways and may lead to more frequent cyber-attacks, including prompt injection, data leakage, insecure integrations, exploit injection, and manipulation or misuse of autonomous AI agents. ASIC's May 2026 open letter[8] to AFS licensees and market participants makes the same point: Frontier AI models are lowering barriers to sophisticated cyber activity, increasing the speed and scale of attacks, and placing existing controls under greater pressure.

The response, however, is not to chase every new AI security product in pursuit of the perfect defence; it is to prove that resilience fundamentals work end-to-end. ASIC expects entities to reassess cyber plans, protect critical assets, validate core controls, reduce attack surface, review access privileges, patch promptly, exercise incident response plans and manage third-party risk actively. APRA's expectations run in parallel, adding privileged access management, hardened configurations, automated vulnerability discovery, penetration testing, controls over agentic workflows, and security testing across AI-generated code and components.

None of this is novel. The question is whether control execution can keep pace with AI-enabled threat velocity. A patching cycle that was adequate last year may be too slow now. Identity controls designed for humans may not govern non-human agents, and change controls built for periodic releases will strain under AI-assisted development. CPS 234 already requires



[8] ASIC Open Letter to AFS Licensees and market participants

testing and assurance commensurate with threats. AI raises the bar on how fast those capabilities must operate.

Key takeaways

- Identify AI platforms, models, agents and integrations as technology assets and classify their criticality based on their use, data sensitivity, decision impact and dependency of critical operations.
- Protect model and data integrity against prompt injection, model manipulation, data poisoning, model theft and unauthorised access to training data.
- Reduce shadow AI risk by strengthening acceptable use standards, data protection controls, and monitoring for unauthorised use of public or unmanaged AI tools.
- Include AI applications, agents and third-party services in penetration testing, red teaming, resilience testing and incident/crisis simulations.
- Assume AI accelerates both attack and defence — refresh threat models for AI-enabled phishing, fraud, social engineering, malware and reconnaissance.

Privacy, data governance and transparency are now core AI controls

AI use cases typically rely on large volumes of customer, claims, transaction, hardship, health or behavioural data. From 10 December 2026[9], APP entities must include additional information in their privacy policies where they have arranged for a computer program to use personal information to make, or substantially and directly inform, a decision that could reasonably be expected to significantly affect an individual's rights or interests.

OAIC guidance for model development[10] also warns that just because data is publicly available does not mean it can be used legally to train or fine-tune generative AI, and that developers

Leaders need to know what data is used, whether it is personal or sensitive, where it flows, who can access it, whether it feeds training or interpretation, whether customers were told, and whether outputs can be explained and corrected.

[9] OAIC: Automated Decision-Making Transparency Obligation (APP 1) – Issues Paper
[10] OAIC: Guidance on privacy and developing and training generative AI models

must take particular care with sensitive information, which generally requires consent.

The control implication is straightforward: Leaders need to know what data is used, whether it is personal or sensitive, where it flows, who can access it, whether it feeds training or interpretation, whether customers were told, and whether outputs can be explained and corrected. Data source, minimisation, quality, lineage, retention and access are no longer back-office concerns; they are front-line AI governance controls.

Key takeaways

- Review and update privacy policies and procedures, notifications, disclosures, and compliance protocols (including those for third parties) in preparation for the transparency obligation to take effect in December 2026.
- Strengthen privacy training for technology and software development teams, with specific focus on OAIC guidance for model development and the responsible use of personal and sensitive information.
- Map AI data flows and ownership in a use-case inventory showing what is consumed, stored and processed, and where personal or sensitive information feeds training, fine-tuning or inference.
- Make privacy impact and AI risk assessments mandatory for high-risk use cases, covering Privacy Act compliance, fairness, transparency and explainability.
- Review data classification, retention, access, consent and minimisation — with a focus on exposure through public generative AI tools and AI-enabled providers.



Resilience means trusted outcomes, not just system uptime

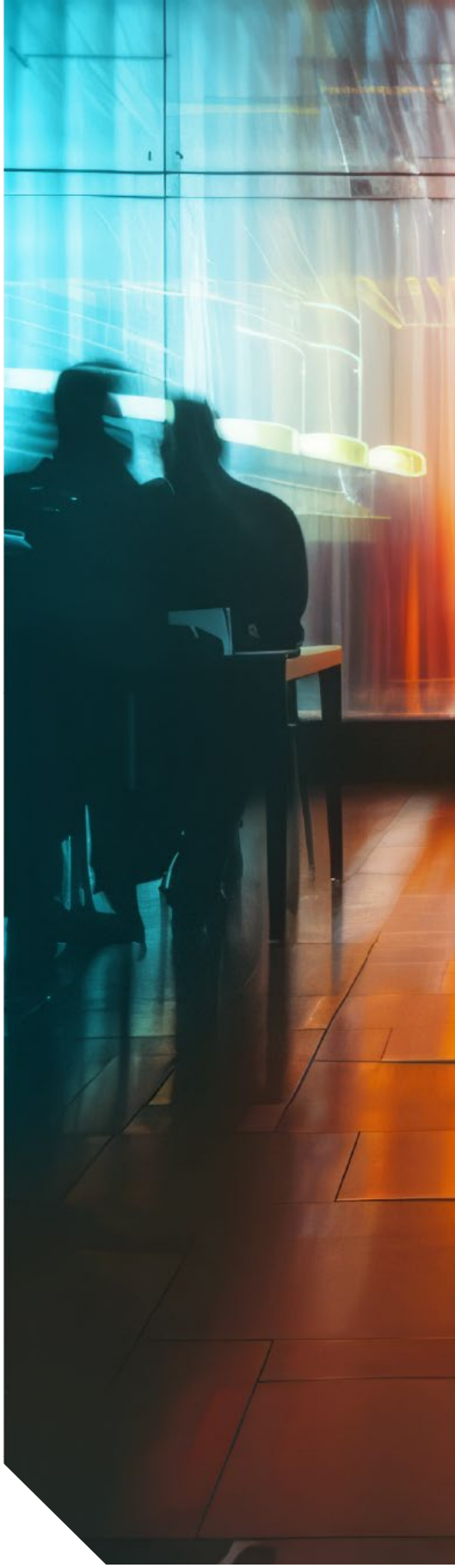
CPS 230, with the Financial Accountability Regime (FAR), makes boards, responsible persons and critical operation owners accountable for keeping critical customer-facing services within tolerance through severe but plausible disruption. AI complicates that picture. It can strengthen resilience through automation, predictive analytics, incident response and fraud detection, while creating new dependencies, concentration risk, and failure

modes such as inaccurate outputs, opaque decisions, model drift and reliance on providers outside direct control.

The key question is whether a critical operation can continue safely if the AI supporting it fails, becomes unavailable, errs or is compromised. And an AI system that is available but producing biased or non-compliant outputs may cause more harm than an outage. Monitoring must cover model performance, output quality, decision accuracy and explainability — not availability alone.

Key takeaways

- Where an AI system or service materially supports a critical operation, reflect that dependency in the entity's resource mapping, tolerance setting, scenario testing and recovery planning.
- Understand and manage AI concentration risk. Assess dependencies on foundation models, cloud providers and AI service providers, and develop contingency arrangements where practical.
- Test resilience beyond technology outages. Incorporate AI-specific disruption scenarios, including model failures, erroneous outputs, data poisoning, provider outages and AI-enabled cyber-attacks.
- Maintain meaningful human oversight. Critical decisions and customer outcomes should not become wholly dependent on AI systems without appropriate monitoring, escalation pathways and intervention mechanisms.



Material service providers: AI will widen existing third-party risk gaps

The most critical AI challenge may sit in the supply chain. Regulators have observed dependencies on single AI providers, weak contingency planning, and limited contractual evidence for audit rights, model updates, incident notification and changes to data handling. CPS 230 already requires entities to maintain a service provider management policy, identify material service providers, maintain registers, conduct due diligence, manage material downstream and fourth-party dependency risk, maintain formal agreements and monitor material arrangements. APRA's AI letter extends the practical pressure: Entities need visibility

over the full AI supply chain, including third- and fourth-party dependencies, model behaviour, material changes, performance issues and concentration risk.

Material service providers should expect more demanding client due diligence and contractual expectations. This will likely include questions about AI use within service delivery, data handling, model governance, access controls, cyber and recovery testing, incident notification, auditability, and AI-generated outputs. A provider that cannot evidence how AI is governed may become a prudential risk to its client.

Key takeaways

- Gain visibility across the AI supply chain, including where subcontractors and fourth parties use AI in critical operations.
- Strengthen contracts for AI-specific risks: privacy, data ownership, training data use, model change, security and incident notification, audit rights, observability, resilience, and exit.
- Assess concentration and single points of failure and decide where contingencies or manual workarounds are needed.
- Hold providers to your own standards, with independent assurance over AI governance, security, privacy and resilience.

A provider that cannot evidence how AI is governed may become a prudential risk to its client.

Global developments are setting the benchmark

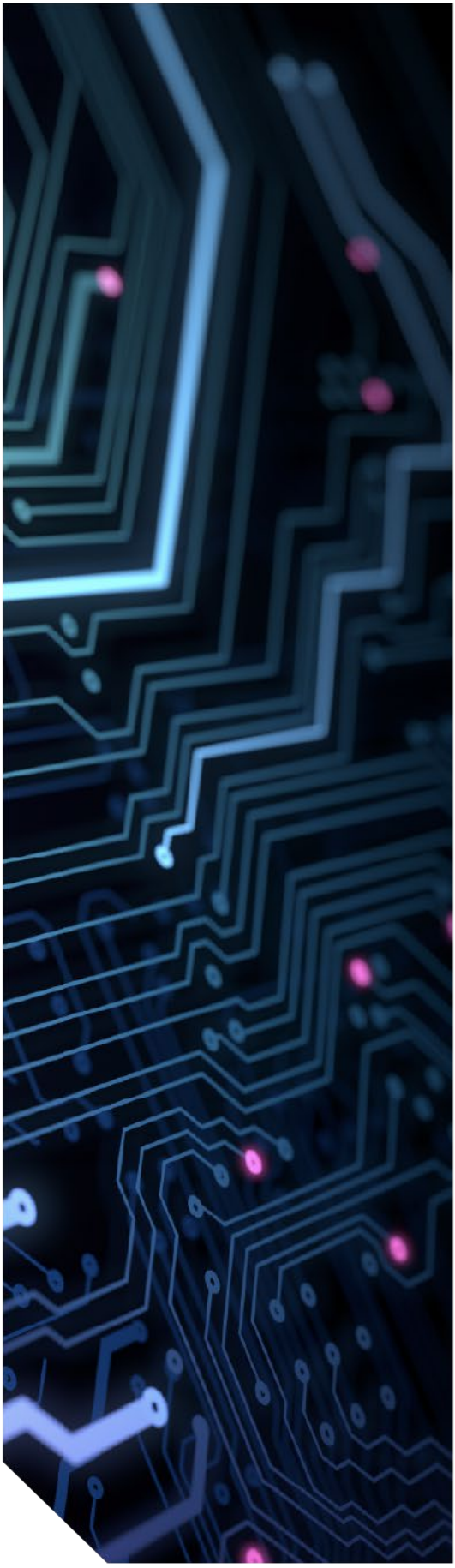
Australia's approach remains principles-based and regulator-led, but global practice is shaping expectations. The EU AI Act entered into force on 1 August 2024 with a risk-based framework covering prohibited practices, high-risk AI systems, specified transparency obligations, general-purpose AI models and lower-risk uses.

Standards are converging on the same themes. These include, but are not limited to, NIST's voluntary, sector-agnostic AI Risk Management Framework; ISO/IEC 42001, which sets requirements for an AI management system; and the OECD AI Principles, updated in 2024, promoting transparency, robustness, security and accountability.

Australia's Voluntary AI Safety Standard sets out 10 guardrails spanning accountability, risk management, data governance, testing, human oversight, disclosure, contestability, supply-chain transparency, recordkeeping and stakeholder engagement. While voluntary and not themselves legally binding, these materials provide a useful reference point for demonstrating responsible AI governance.

Calls to action: Four key actions for the next six months

1. **Establish a board-visible AI inventory and risk tiering.** Capture sanctioned and unsanctioned tools, AI embedded in platforms, AI-enabled providers and live use cases, tiered by customer impact, decision autonomy, data sensitivity and critical-operation dependency.
2. **Anchor AI to existing frameworks rather than a decorative standalone policy.** Map AI controls to CPS 230, CPS 234, CPS 220/SPS 220, CPS 510, privacy, conduct and supplier obligations with named owners, control standards, escalation triggers and board reporting.
3. **Uplift AI-specific third-party and data controls.** Refresh contracts and due diligence for AI-enabled services, and link them to privacy impact assessments, data minimisation and the December 2026 ADM transparency obligation.
4. **Make AI assurance continuous.** Move from point-in-time assurance to continuous validation, monitoring for drift, bias, accuracy, vulnerabilities, model changes and misuse — with the second line and internal audit technically capable of assessing probabilistic AI models and agentic workflows.



Conclusion

AI governance cannot be owned by technology alone. Legal, compliance, privacy, conduct, product, risk, cyber, data, procurement and the business all have a role. For APRA-regulated entities and their material service providers, the path forward is not to wait for a standalone Australian AI Act. It is to make existing frameworks work for AI — visibly, continuously and with evidence. Boards should be able to answer four questions:

- *What AI systems are we using?*
- *What outcomes do they influence?*
- *What could go wrong?*
- *How do we know the controls are working?*

If those cannot be answered with confidence, the organisation is already behind the risk curve.

About the author



Hirun Tantirigama
Managing Director

Hirun Tantirigama is a Managing Director and Technology Consulting Lead at Protiviti Australia with more than 18 years of experience in technology risk, regulatory compliance, cyber resilience and transformation programs.

As Asia-Pacific lead for operational resilience and a member of Protiviti's global Operational Resilience team, Hirun advises boards and executive teams on governance, risk management and resilience in complex regulatory environments. His expertise spans prudential standards, cyber security, third-party risk and operational resilience, including the governance, privacy and risk implications of artificial intelligence. He has advised organisations across Australia, New Zealand, the United Kingdom and Asia-Pacific.

Protiviti (www.protiviti.com) is a global consulting firm that helps clients transform and protect their businesses, and respond to planned and unexpected events. Through a network of more than 90 offices in over 25 countries, Protiviti and its independent and locally owned member firms deliver deep expertise and tailored capabilities across technology, artificial intelligence, data, operations, finance, legal, compliance, HR, marketing, digital, risk, and internal audit—enabling organisations to accelerate innovation, navigate risks and safeguard what matters most.

Named to the Fortune 100 Best Companies to Work For® list since 2015, Protiviti Inc. has served more than 80 percent of Fortune 100 and nearly 80 percent of Fortune 500 companies. The firm also works with government agencies and smaller, growing companies, including those looking to go public. Protiviti Inc. is a wholly owned subsidiary of Robert Half (NYSE: RHI).

© 2026 Protiviti. 10/26
Protiviti is not licensed or registered as a public accounting firm and does not issue opinions on financial statements or offer attestation services.

protiviti®